MALDI-TOF and Tandem MS for Clinical Microbiology

MALDI-TOF and Tandem MS for Clinical Microbiology

Edited by

Haroun N. Shah
Department of Natural Sciences
Middlesex University
London, UK

Saheer E. Gharbia
Genomic Research
Public Health England
London, UK

WILEY

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices
111 River Street, Hoboken, NJ 07030, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
Boschstr. 12, 69469 Weinheim, Germany

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Shah, Haroun N., editor. | Gharbia, Saheer, editor.
Title: MALDI-TOF and tandem MS for clinical microbiology / edited by Haroun N. Shah, Middlesex University, Middlesex, UK, Saheer E. Gharbia, Public Health England, Genomic Research Unit, London, UK.
Description: Chichester, West Sussex : John Wiley & Sons, Inc., 2017. | Includes bibliographical references and index.
Identifiers: LCCN 2016054266 (print) | LCCN 2016055285 (ebook) | ISBN 9781118960257 (cloth) | ISBN 9781118960240 (pdf) | ISBN 9781118960233 (epub)
Subjects: LCSH: Diagnostic microbiology. | Microbiology. | Matrix-assisted laser desorption-ionization. | Tandem mass spectrometry.
Classification: LCC QR67 .M35 2017 (print) | LCC QR67 (ebook) | DDC 616.9/041–dc23
LC record available at https://lccn.loc.gov/2016054266

Cover image: Cover Figure Courtesy of Tim Chambers, Proteomics Research, Public Health England, UK
Cover design: Wiley

Set in 10/12pt Warnock by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

Dedication

For Amaya, Iona, Laila, Louise, Camille and Haja Naima

Contents

List of Contributors

Mariann Ábrók
Institute of Clinical Microbiology
Faculty of Medicine
University of Szeged
Hungary

Nadia Ahmod
Proteomics Research
Public Health England
London
UK

David J. Allen
Virology
Public Health England
London
UK

Jean Armengaud
CEA-Marcoule, DSV/IBITEC-S/SPI/Li2D
Laboratory "Innovative Technologies for Detection and Diagnostics"
Bagnols-sur-Cèze
France

Maud Arsac
bioMérieux SA
Marcy-l'Etoile
France

Peter S. Borriello
Veterinary Medicines Directorate
Surrey
UK

Fredrik Boulund
Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg
Sweden

Patrick Broyer
bioMérieux SA
Marcy-l'Etoile
France

Irene Burckhardt
Division of Medical Microbiology and Hygiene, Department of Infectious Diseases
University Hospital Heidelberg
Germany

Valeria Calvaresi
Structural Mass Spectrometry and Proteomics Unit
Institut Pasteur
Paris, France

Timothy Chambers
Proteomics Research
Public Health England
London
UK

Julia Chamot-Rooke
Structural Mass Spectrometry and Proteomics Unit
Institut Pasteur
Paris
France

Marie-Hélène Charles
bioMérieux SA
Marcy-l'Etoile
France

Sonia Chatellier
bioMérieux SA
La Balme-les-Grottes
France

Caroline H. Chilton
Leeds Institute for Biomedical and Clinical Sciences
University of Leeds
UK

Song Lin Chua
Singapore Centre for Environmental Life Sciences Engineering (SCELSE)
and
Lee Kong Chian School of Medicine
Nanyang Technological University
Singapore

Arnaldo Colombo
Department of Medicine
Infectious Diseases Section
Federal University of São Paulo (UNIFESP)
SP
Brazil

Renata Culak
Proteomics Research
Public Health England
London
UK

Ronald R. Cutler
School of Biological and Chemical Sciences
Queen Mary University of London
UK

Katrien De Bruyne
Applied Maths NV
Sint-Martens Latem
Belgium

Willem M. de Vos
Faculty of Veterinary Medicine
University of Helsinki
Finland
and
Laboratory of Microbiology
Wageningen University
The Netherlands

Itaru Dekio
Medical Centre East
Tokyo Women's Medical University
Japan

Magalie Duchateau
Structural Mass Spectrometry and Proteomics Unit
Institut Pasteur
Paris
France

Mathieu Dupré
Structural Mass Spectrometry and Proteomics Unit
Institut Pasteur
Paris
France

Min Fang
Proteomics Research
Public Health England
London
UK

Mark A. Fisher
University of Utah School of Medicine
and
Bacteriology and Antimicrobials
ARUP Infectious Diseases Laboratory
Salt Lake City, UT, USA

Elaine Francisco
Department of Medicine
Infectious Diseases Section
Federal University of São Paulo (UNIFESP)
SP
Brazil

Joseph Gault
Structural Mass Spectrometry and Proteomics Unit
Institut Pasteur
Paris
France

Tom Gaulton
Proteomics Research
Public Health England
London
UK

Saheer E. Gharbia
Genomic Research Unit
Public Health England
London
UK

Victoria Girard
bioMérieux SA
La Balme-les-Grottes
France

Lucia Gonzales-Siles
Department of Infectious Diseases
Institute of Biomedicine
The Sahlgrenska Academy
University of Gothenburg
Sweden

Gilbert Greub
Institute of Microbiology, University Hospital Center, and University of Lausanne
Switzerland

Nicola Hennessy
Genomic Research Unit
Public Health England
London
UK

Jenny Ho
ThermoFisher Scientific
Hemel Hempstead
Hertfordshire
UK

Christiane Honisch
Illumina, Inc.
San Diego, CA
USA

Martin Hornshaw
ThermoFisher Scientific
Hemel Hempstead
Hertfordshire
UK

Jay Hyman
bioMérieux Inc
Durham, NC
USA

Axel Karger
Friedrich-Loeffler-Institut
Federal Research Institute for Animal Health
Greifswald – Insel Riems
Germany

Anders Karlsson
Nanoxis Consulting AB
Gothenburg
Sweden

Roger Karlsson
Department of Clinical Microbiology
Sahlgrenska University Hospital
and
Nanoxis Consulting AB
Gothenburg
Sweden

Jae-Seok Kim
Department of Laboratory Medicine
Hallym University College of Medicine
Seoul
Korea

Yangsun Kim
ASTA Inc
Gyeonggi-do
Korea

Carolin Kolmeder
Faculty of Veterinary Medicine
University of Helsinki
Finland

Markus Kostrzewa
Bruker Daltonik GmbH
Bremen
Germany

Annika Kotovuori
ThermoFisher Scientific
Vantaa
Finland

Erik Kristiansson
Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg
Sweden

Kaarina Lähteenmäki
Finnish Red Cross Blood Service
Helsinki
Finland

Diogo B. Lima
Structural Mass Spectrometry and
Proteomics Unit
Institut Pasteur
Paris
France

Nelson Lima
CEB-Centre of Biological Engineering
Micoteca da Universidade do Minho
University of Minho
Braga
Portugal

Åsa Lindgren
Department of Clinical Microbiology
Sahlgrenska University Hospital
and
Department of Infectious Diseases
Institute of Biomedicine, The Sahlgrenska
Academy, University of Gothenburg
Sweden

Pierre Mahé
bioMérieux SA
Marcy-l'Etoile
France

Thomas Maier
Bruker Daltonik GmbH
Bremen
Germany

Christian Malosse
Structural Mass Spectrometry and
Proteomics Unit
Institut Pasteur
Paris
France

Jaana Mättö
Finnish Red Cross Blood Service
Helsinki
Finland

Mariana Mazza
Department of Mycology
INEI/Instituto Nacional de Enfermedades
Infecciosas "Dr. Carlos G. Malbrán"-ANLIS.
Ciudad Autónoma de Buenos Aires
Argentina

Raju V. Misra
Genomic Research Unit
Public Health England
London
UK

Hermine V. Mkrtchyan
School of Biological and Chemical
Sciences
Queen Mary University of London
UK

Edward R.B. Moore
Department of Clinical Microbiology
Sahlgrenska University Hospital

and
Department of Infectious Diseases
Institute of Biomedicine
The Sahlgrenska Academy
University of Gothenburg
Sweden

Leonardo P. Munoz
Department of Natural Sciences
Middlesex University
London, UK

Elisabeth Nagy
Institute of Clinical Microbiology
Faculty of Medicine
University of Szeged
Hungary

Ali Olkun
Department of Natural Sciences
Middlesex University
London, UK

Onya Opota
Institute of Microbiology, University
Hospital Center, and University of
Lausanne
Switzerland

Ana Carolina B. Padovan
Department of Medicine
Infectious Diseases Section
Federal University of São Paulo
(UNIFESP)
SP
Brazil

Nadine Perrot
bioMérieux SA
La Balme-les-Grottes
France

Bruno Pot
Applied Maths NV
Sint-Martens Latem
Belgium

Ian R. Poxton
Department of Medical Microbiology
University of Edinburgh
UK

Guy Prod'hom
Institute of Microbiology, University
Hospital Center, and University of
Lausanne
Switzerland

Christopher Ring
Department of Natural Sciences
Middlesex University
London, UK

Ilja Ritamo
ThermoFisher Scientific
Vantaa
Finland

Graham Rose
Genomic Research Unit
Public Health England
London
UK

Cledir R. Santos
Department of Chemical Sciences
and Natural Resources
CIBAMA, BIOREN, University
of La Frontera
Temuco
Chile

Vlad Serafim
Department of Natural Sciences
Middlesex University
London
UK

Ajit J. Shah
Department of Natural Sciences
Middlesex University
London
UK

Haroun N. Shah
Proteomics Research
Public Health England
and
Department of Natural Sciences
Middlesex University
London
UK

Liselott Svensson-Stadler
Department of Clinical Microbiology
Sahlgrenska University Hospital
and
Department of Infectious Diseases
Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg
Sweden

Hiroto Tamura
Laboratory of Environmental Microbiology, Department of Environmental Bioscience
Meijo University
Nagoya
Japan

Edith Urbán
Institute of Clinical Microbiology
Faculty of Medicine
University of Szeged
Hungary

Leena Valmu
ThermoFisher Scientific
Vantaa
Finland

Alex van Belkum
bioMérieux SA
La Balme-les-Grottes
France

Arie Jan van Winkelhoff
Department of Medical Microbiology
University Medical Center Groningen
The Netherlands

A.C.M. Veloo
Department of Medical Microbiology
University Medical Center Groningen
The Netherlands

Jean-Baptiste Veyrieras
bioMérieux SA
Marcy-l'Etoile
France

Egor Vorontsov
Structural Mass Spectrometry and Proteomics Unit
Institut Pasteur
Paris
France

Katleen Vranckx
Applied Maths NV
Sint-Martens Latem
Belgium

Pirjo Wacklin
Salwe Ltd
Espoo
Finland

John Walsh
bioMérieux Inc
Durham, NC
USA

Zhen Xu
School of Biological and Chemical Sciences
Queen Mary University of London
UK

Liang Yang
Singapore Centre for Environmental Life Sciences Engineering (SCELSE)
and
School of Biological Sciences
Nanyang Technological University
Singapore

Stefan Zimmermann
Division of Medical Microbiology and Hygiene
Department of Infectious Diseases
University Hospital Heidelberg
Germany

Preface

A Brief Tour of the Technology and New Grounds for Innovation

Mass spectrometry is an extremely powerful analytical technique, capable of elucidating not only the molecular weight but also the chemical structure of target compounds. Recent advances have paved the way for deciphering complex and rich biomolecules spanning applications in diagnostics, biodiscovery and detection. In recent years, spectacular advances in mass spectrometry technology have improved all aspects of the instrumentation, making it more robust and easier to use, thus breaking new frontiers in life sciences research, pharmaceutical analysis and, more recently, clinical diagnostics.

Modern mass spectrometers deliver information-rich data with high reproducibility and sensitivity, which makes them suitable for identifying and quantifying low-level biomarkers with high confidence. The ability of the technique to "read" multiple targets in a single experimental sequence allows mass-spectrometry-based assays to target a multitude of markers in a single assay. Structural analysis, powerful statistical techniques and database-based bioinformatics are employed to increase the fidelity of results.

The main types of mass spectrometers utilized in clinical practice today are triple quadrupoles (QqQ), quadrupole-time-of-flight hybrids (Q-TOF) and matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometers. Triple quadrupoles and Q-TOFs deliver very high sensitivity and specificity of results by mostly using targeted structural analysis. MALDI-TOF mass spectrometers deliver easy and economical high-throughput analysis with highly reproducible pattern recognition or 'fingerprints'. Several other types of mass spectrometers including FT-ICR-type (e.g. Orbitrap®-based instruments) are used in state-of-the-art life sciences research and biomarker discovery and are already finding their way into clinical laboratories.

Triple quads and Q-TOFs normally utilize electrospray ionization (ESI), a technique that generates a fine spray from liquid samples, presented directly into an orifice to be transferred into the instrument's vacuum system. Droplets dry out and release ions with one or several charges. These instruments need a high vacuum to operate, and thus a vacuum and ion optics system is utilized to transfer ions into the high-vacuum part of the device through direct infusion or liquid phase separation (HPLC). In many cases, fast HPLC runs may be adequate; however, such treatment increases complexity and time per analysis, which is the main drawback of ESI-based assays. Emerging technologies have been introduced that may present solutions to this challenge, including desorption electrospray ionization (DESI), disposable ESI nozzles and paper spray.

In analytical research, ESI is the dominant MS largely because of its seamless coupling with liquid phase separation methods. This great advantage for comprehensive analysis of complex samples becomes a disadvantage for high-throughput clinical diagnostics, when front-end simplicity, robustness and speed of analysis are of critical importance. This is where MALDI-TOF has dominated the developments in microbial analysis and clinical diagnostics.

MALDI-TOF MS and its variants (e.g. surface-enhanced laser desorption and ionization—SELDI-TOF MS) typically utilize high-power–short-pulse-width laser pulses for desorption and ionization of the sample. MALDI can be coupled with various ion analyzers and ion processors (e.g. ion traps); however, being a pulsed ionization method, it is preferably coupled with time-of-flight (TOF) mass analyzers. The fundamental principle of TOF mass analysis is to link the mass of an ion (or more precisely, its mass-over-charge ratio, m/z) to its "flight" time to a detection point. In the simplest theoretical construct, all ions need to start their flight at the same time and place, whereas ion velocities should differ in relation to their mass. In that case, a simple formula/calibration function would transform time of flight to m/z. MALDI-TOF mass spectrometers accelerate all ions at the same kinetic energy level in an ion source; as different mass ions travel with different velocities, they are allowed to 'fly' in a field-free region where they separate in space; then the device records their arrival time to a set point in space, and it is translated into their mass by the system electronics and software.

Although the principle is simple, modern instruments employ several technologies and state-of-the-art electronics to improve performance. Devices are designed to correct or minimize several aberrations, by means of specialized ion optics, electrodynamic extraction pulses and highly sophisticated detection devices. Another complication is that the entire process should be carried out in high vacuum. This is to minimize random collisions with ambient gas, which reduce constant velocity throughout measurement. A rather complex vacuum load lock is normally employed to transfer sample into the vacuum. The need for high analysis speed requires several samples to be loaded in a single loading cycle, leading to large multi-spot sample plates and complicated sample manipulation mechanisms. Various systems solutions are emerging to tackle these engineering challenges and enhance speed and sensitivity.

MALDI-TOF suitability for high-throughput analysis makes it an ideal technology to develop diagnostic tests spanning a wide variety of targets, including blood disorders (e.g. sickle cell anaemia), lipid profiling and other metabolic and proteomics-based assays. An important target application is early cancer diagnostics, which has been attempted at the research level with protein biomarkers (SELDI-based methods) or by probing key proteomic post-translational modification, for example, semi-quantitative glycosylation profiling. These exciting developments maintain a keen focus in the field and become an incentive to make the technology even more powerful, robust, economical and easy to use. The advance of mass spectrometry in the clinical laboratory will help early diagnosis, treatment monitoring and personalized medicine, thus contributing to public health and well-being.

Currently, all commercial MALDI-based clinical platforms utilize one of the simplest forms of TOF, a linear TOF mass analyzer. Its success in clinical microbiology over the last decade or so has been described as transformative. In the United Kingdom, for example, most clinical microbiology laboratories currently either possess a

MALDI-TOF MS instrument or are actively seeking to acquire one. Recent refinement in software and bioinformatic tools are supporting further extraction of data from samples to derive not only species identification, but typing differences among strains of the same species (see, e.g. Chapters 5, 6, 11–13, 16 and 21). Additional specialized pre-sample processing methods are also facilitating the characterization of antibiotic resistance of isolates (e.g. Chapters 10 and 15).

This first phase of MALDI-TOF MS applications has been designated as a 'new era in clinical microbiology' because of its dramatic impact on diagnostics (see Chapters 1–12 and 21). The next phase is likely to move towards higher-specification instruments as well as MS/MS-based methods, where considerable progress has already been achieved. Therefore, the second half of this book (Chapters 13–20) focuses on these methods. Mass spectrometry and proteomics are also being propelled forward by the increasing availability of whole genome sequencing of microorganisms, hence providing a wealth of data for in silico prediction and comparative validation of proteins, sequences and identity. This approach of combined data analysis is termed *proteogenomics*, which in clinical microbiology is not only yielding fingerprint capabilities and improved antibiotic sensitivity assays, but will have the capacity to identify toxins, signal molecules, attachment target proteins, metabolic intermediates and a plethora of virulence determinants that reveal the pathogenic potential of a microorganism in real time. This will transform patient treatment and will serve as a powerful, rapid tool for both patient care and public health interventions, from pursuing transmission to developing vaccines and therapeutics.

Emmanuel Raptakis
Managing Director, Citylabs, Manchester, UK

Part I

MALDI-TOF Mass Spectrometry

1

A Paradigm Shift from Research to Front-Line Microbial Diagnostics in MALDI-TOF and LC-MS/MS: A Laboratory's Vision and Relentless Resolve to Help Develop and Implement This New Technology amidst Formidable Obstacles

Haroun N. Shah[1] *and Saheer E. Gharbia*[2]

[1] *Proteomics Research and* [2]*Genomic Research Unit, Public Health England, London, UK*
[1] *Current address: Department of Natural Sciences, Middlesex University, London, UK*

1.1 Introduction

The outlooks of a microbial systematics research laboratory and a clinical diagnostic laboratory are scientifically and logistically polarized, due almost entirely to the nature of their approach to generating knowledge and translating technological tools. Research laboratories are generally sustained through scientific grants, and their focus is driven by development and applications of novel and emerging technologies; their direction is more fluid and innovative. By contrast, the clinical diagnostic laboratory is more circumspect and notoriously resistant to change in methodology and mindful of the meticulous level of validation and accreditation needed to implement new workflows. Thus, over the years, whereas researchers exploited biochemical and chemotaxonomical advances, the traditional clinical diagnostic laboratory remained largely disinclined.

The arrival of comparative 16S rRNA sequence analysis heralded a new era of microbiology and allowed for the first time the classification of microorganisms along phylogenetic lines. This began in the 1970s with the introduction of rRNA cataloguing, then reverse transcriptase sequencing of the 5S, 16S and 23S rRNA subunits and, finally, with the arrival of PCR and sequencing of genes in the 1990s. The impact of comparative 16S rRNA sequence analysis enabled the biggest change witnessed in microbial systematics for over a century and led to the assembly of the largest database in the history of life sciences. Although at the higher taxonomic level, the topography of many of the lineages remained stable, substantial changes in classification and nomenclature were being recorded at the genus and species levels. The need to study intra-generic and intra-species diversity became essential. Clinical microbiology was at a crossroads; it could not ignore the increasing volume of literature, and began slowly to embrace these changes.

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

1.1.1 Personal Experience at the Interface of Systematics and Diagnostics

The Public Health Laboratory Service (PHLS) established its first dedicated Molecular Identification Services Unit (MISU) in late 1997 with a remit to find novel methods to identify atypical, rarely isolated and emerging taxa that clinical laboratories failed to identify. PHLS at that time comprised a network of some 50 specialist laboratories in England and Wales that served as a national referral centre for the characterization of human pathogens that could not be delineated by clinical laboratories. In spite of the specialist nature of PHLS laboratories, a significant proportion of samples remained *incertae sedis* and were left unreported. MISU's remit was to address this problem, and one of the authors (HNS), joined PHLS to spearhead reshaping diagnostics along a systematic and technology framework. Despite years of experience (from the early 1970s) in the field of microbial systematics, particularly with the *Bacteroidaceae* using biochemical, chemical and molecular methods, the implementation of 16S rRNA from a research laboratory to a daily service function was a daunting challenge and only became a routine procedure by establishing a workflow from sample to sequence and data in hours rather than in days, the timeline associated with multiple steps in amplification and DNA sequencing. This was facilitated by the new Applied and Functional Genomics Unit (headed by author SEG in 2002) and provided the momentum for the service to be accredited. Over the years, a large number of novel genomics and proteomic technologies were explored by both laboratories and served as a platform to introduce new technologies and their potential applications. Matrix-assisted laser desorption/ionization (MALDI)-time-of-flight mass spectrometry (MALDI-TOF MS) was the first of these and dates back to its inception in October 1997.

1.1.2 MALDI-TOF MS: The Early Years

The arrival of MALDI-TOF MS coincided with attempts to introduce 16S rRNA as a diagnostic method in PHLS. The almost simultaneously publication of work from several laboratories – for example, Claydon *et al.* (1996); Cain *et al.* (1994); Holland *et al.* (1996) and Krishnamurthy and Ross (1996) – highlighted its potential to create a pattern derived from bacterial cells that discriminated between fairly disparate species. However, it was given little attention by clinical microbiologists, largely because of the success, ease and by now, the availability of public databases of 16S rRNA sequences. This was in direct contrast to MALDI-TOF MS, for which there was only a general bench-top instrument, no standard protocol nor more than a few dozen MALDI-TOF MS spectra that were published mainly in MS journals.

The motivation to explore the application of this technology by the author stems from the need to find new methods to identify the unusual range of taxa that were received for identification for which even 16S rRNA results were equivocal. Having previously reported extensive work on characterization of lipids by MS (see, e.g. Shah and Collins, 1980; Shah and Collins, 1983; Shah and Gharbia, 2011), the potential use of MALDI-TOF MS, which had the capability to ionize molecules orders of magnitude greater in mass, was an intriguing challenge and was explored initially using species of the genus *Porphyromonas*. A key character of this group is their inability to ferment carbohydrates, the primary means of identifying bacterial species at that time. DNA-DNA reassociation and lipid analyses by MS were the only

reliable methods to delineate members of this genus, but this was undertaken by only a few specialist laboratories and was lengthy, tedious, required a large biomass (>50 mg of cells) and was limited in scope (see, e.g. Shah *et al.*, 1982). By contrast, when *Porphyromonas* species were subjected to analysis by MALDI-TOF MS in early 1998, they were readily delineated using just a few cells directly from an agar plate (see Figure 1.1; see Shah *et al.*, 2000; Shah *et al.*, 2002). This watershed moment not only provided proof of concept but gave such confidence in the potential of this technology that MISU would go on to painstaking pursue its development for nearly two decades (Shah, 2005).

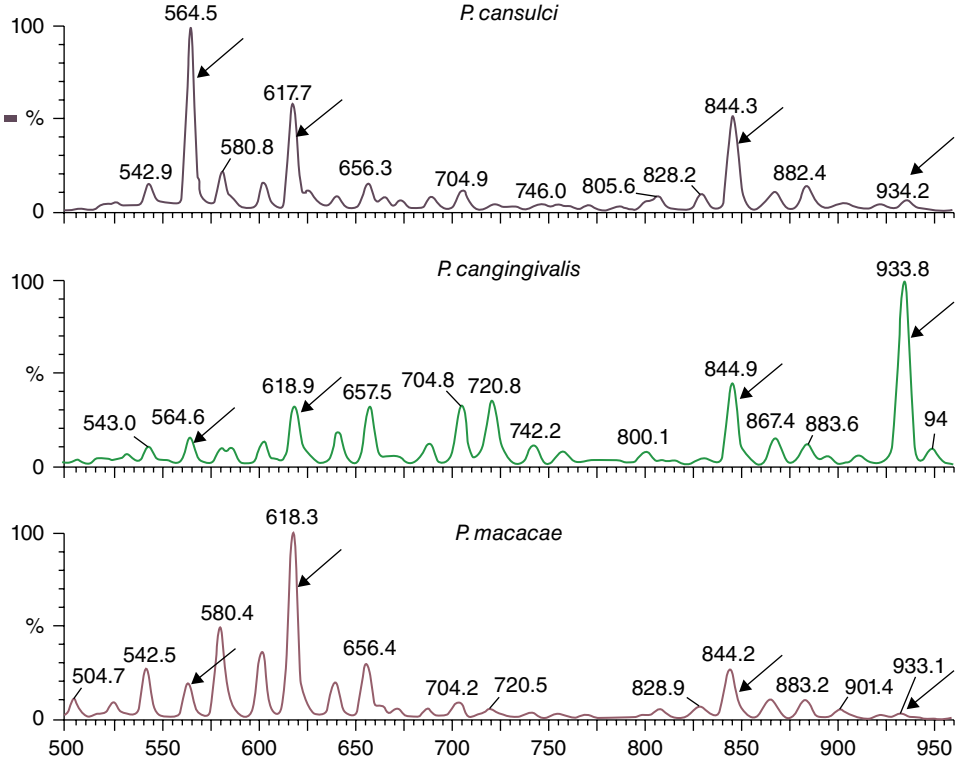


Figure 1.1 Examples of the distinctive MALDI-TOF-MS profiles of intact cells of *Porphyromonas* sp. containing both genus-specific (e.g. 618 and 844 Da mass ions) and also a significant number of species-specific mass ions (examples indicated by arrows). Members of the genus *Porphyromoas* now comprise 18 species, in addition to several others that have not yet been validated (Bergey's manual, 2011). However, with the exception of DNA/DNA reassociation, they could not be reliably delineated at the time. The three representative MALDI-TOF-MS spectra shown and those reported earlier (Shah *et al.*, 2002) revealed that each species could be unambiguously distinguished. It was this poorly characterized group of anaerobes that became one of the compelling forces for the development of this technique for microbial identification. Early meetings to demonstrate an appreciation of the technology were mostly presented at meetings on anaerobic taxa and helped rejuvenate interest in this area of microbiology again (see Chapter 5 and Figure 1.4).

1.1.3 The Formidable Challenge to Gain the Confidence of the Clinical Microbiologist in MALDI-TOF MS

At its commencement, many microbiologists pointed to the dismal failure of pyrolysis MS (PyMS, circa 1970–1990) as a clinical diagnostic tool and simply branded MALDI-TOF MS as a newer version of PyMS that generated a mass spectral profile of unidentified proteins of intact cells. However, a minority believed that MALDI-TOF MS was inherently superior, had more scope for development and took up the mantle to unlock its potential. Our laboratory was one of the forerunners, which embraced the drive to develop and implement MALDI-TOF MS for microbial identification.

The proposal to introduce MALDI-TOF MS for diagnosis in a molecular-based service for human infectious diseases seemed preposterous to many. Efforts to develop MALDI-TOF MS and proteomics, in particular, were met with fierce resistance both from within and outside PHLS. Consequently, no core funding was allocated for its development throughout the years; support came from scientific grants, industrial funding and collaboration with various MS companies. Initially, Kratos Analytical, Manchester, UK, which designed and built the first bench-top linear MALDI-TOF mass spectrometer, the Kompact Alpha^{TM} mass spectrometer, placed this instrument in MISU between 1998 and 1999 to help develop the methodology and explore the potential applications of the technique (Figure 1.2). To bring it to the attention of the wider scientific community, the first conference in the field entitled 'Intact Cell MALDI: A Novel Technique for the Rapid Identification of Microorganisms' was held at PHLS, London on 27 October 1998 and was attended by some 150 scientists (see Figure 1.3).

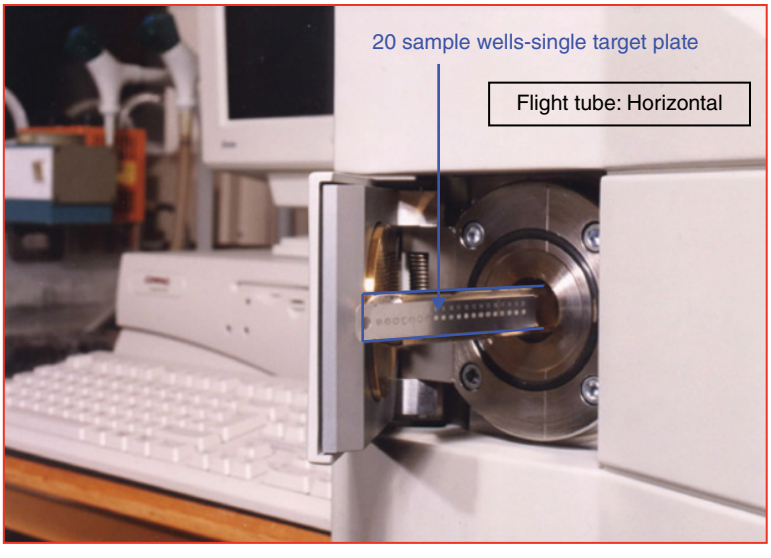
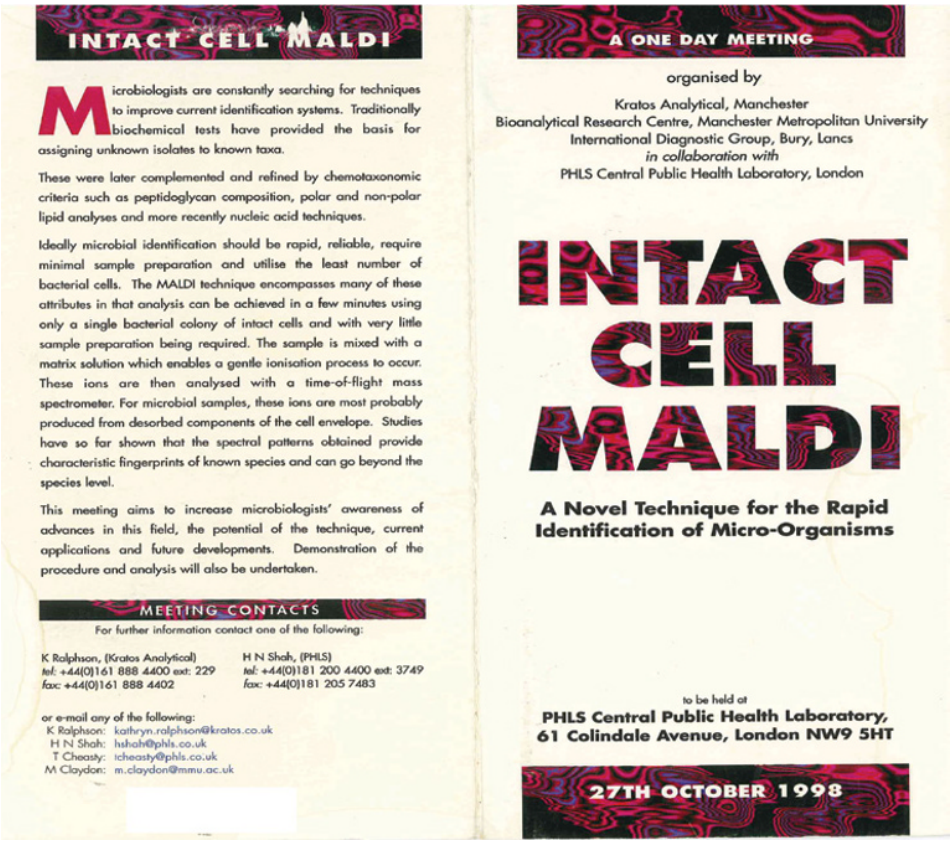


Figure 1.2 The first bench-top linear MALDI-TOF mass spectrometer, the Kratos Kompact Alpha (Kratos Analytical, Manchester, UK), that inspired the development of the technology for clinical microbiology. The target plate, with a capacity for 20 samples, is shown and was used for generating the spectra shown in Figure 1.1. The instrument revealed the potential of the technology, but it was manual and unsuitable for a microbial diagnostic laboratory.



INTACT CELL MALDI

Microbiologists are constantly searching for techniques to improve current identification systems. Traditionally biochemical tests have provided the basis for assigning unknown isolates to known taxa.

These were later complemented and refined by chemotaxonomic criteria such as peptidoglycan composition, polar and non-polar lipid analyses and more recently nucleic acid techniques.

Ideally microbial identification should be rapid, reliable, require minimal sample preparation and utilise the least number of bacterial cells. The MALDI technique encompasses many of these attributes in that analysis can be achieved in a few minutes using only a single bacterial colony of intact cells and with very little sample preparation being required. The sample is mixed with a matrix solution which enables a gentle ionisation process to occur. These ions are then analysed with a time-of-flight mass spectrometer. For microbial samples, these ions are most probably produced from desorbed components of the cell envelope. Studies have so far shown that the spectral patterns obtained provide characteristic fingerprints of known species and can go beyond the species level.

This meeting aims to increase microbiologists' awareness of advances in this field, the potential of the technique, current applications and future developments. Demonstration of the procedure and analysis will also be undertaken.

MEETING CONTACTS

For further information contact one of the following:

K Ralphson, (Kratos Analytical)
tel: +44(0)161 888 4400 ext: 229
fax: +44(0)161 888 4402

H N Shah, (PHLS)
tel: +44(0)181 200 4400 ext: 3749
fax: +44(0)181 205 7483

or e-mail any of the following:
K Ralphson: kathryn.ralphson@kratos.co.uk
H N Shah: hshah@phls.co.uk
T Cheasty: tcheasty@phls.co.uk
M Claydon: m.claydon@mmu.ac.uk

A ONE DAY MEETING

organised by

Kratos Analytical, Manchester
Bioanalytical Research Centre, Manchester Metropolitan University
International Diagnostic Group, Bury, Lancs
in collaboration with
PHLS Central Public Health Laboratory, London

INTACT CELL MALDI

A Novel Technique for the Rapid Identification of Micro-Organisms

to be held at
PHLS Central Public Health Laboratory, 61 Colindale Avenue, London NW9 5HT

27TH OCTOBER 1998

Figure 1.3 The conference leaflet of the first meeting held on 27 October 1998 to explore the potential application of MALDI-TOF mass spectrometry for microbial identification. These meeting were held annually to showcase developments in the field. However, it still took a decade of research and development for this technology to gain widespread acceptance in the clinical microbiology laboratory.

The authors' presentation titled 'A Review of the Current Methods of Bacterial Identification: MALDI-TOF MS in the Characterisation of the Obligate Anaerobes *Fusobacterium* and *Porphyromonas*' outlined the paucity of reliable characters for delineating many taxa within the *Bacteroidaceae* and presented early data using MALDI-TOF MS as proof of principle of the technique.

Although there were intermittent reports in the scientific community on the analysis of specific taxa, there was neither a generally accepted method nor a coherent plan to establish a universal approach to using MALDI-TOF MS as a diagnostic tool. Furthermore, major concerns were raised at the meeting about introducing a new technique to analyze human pathogens involving use of intact, viable cells. We carried out thorough investigations on the safety aspects of the procedures and produced a detailed application note for Kratos Analytical in 1999. To verify this, tough resistant spore-formers, *Bacillus stearothermophilus* (NCTC 10003) and *Bacillus subtilis* (NCTC 10073), which are industry standard controls for heat and

chemical sterilization, were processed through the instrument at various concentrations and samples collected at potential leakage points within the instrument and vacuum pumps. No growth was observed in any sample, and this, together with other successful safety tests, led to the Health and Safety Executive granting approval for the use of MALDI-TOF MS in a clinical laboratory and permission for engineers to carry out repairs in situ when necessary. MISU organized several hands-on workshops for service engineers from MS companies to alert them to the hazards of undertaking fieldwork.

However, in spite of careful assessment of the risk, one major accident occurred in the laboratory in 2000 that nearly terminated all future work at PHLS. The standard method used for cleaning target plates at the time was sonicating for half-hour intervals in methanol followed by washing in 33% (w/v) aqueous nitric acid, and final rinsing in distilled water (e.g. Evason *et al.*, 2001). If the methanol and nitric acid are mixed and kept in a sealed container, the reaction is explosive. This error occurred, and a huge accident resulted in the laboratory being closed for several months and the project nearly being terminated.

However apart from safety issues, the mere separation of a few disparate species by differences in MALDI-TOF MS spectra was insufficient to introduce a new technology for clinical diagnostics. Instrument design, robust sample preparation among the varied bacterial chemotypes, reproducible mass spectra on a single instrument as well as other instruments and development of a database were among the major challenges faced at the time. From these initial stages, we set out to explore the various forms of MALDI-TOF MS and later on, tandem MS/MS proteomics for the characterization of human pathogens. Annual international conferences were organized to showcase developments at each stage together with horizon scanning for future work (see Table 1.1).

1.2 Overcoming the Variable Parameters of MALDI-TOF MS Analysis: Publication of the First Database in 2004

Between October 1997 and December 1999, we began to explore the diagnostic potential of MALDI-TOF MS with Manchester Metropolitan University using the Kratos Kompact Alpha^{TM} mass spectrometer. Samples were taken from broth and plate cultures with cells from various growth phases. We analyzed intact and broken cells preparations, fractions from gradient centrifugation steps, and so on. Several matrix solutions, at a range of concentrations, were used to derive a reproducible MALDI-TOF MS spectrum. These included 5-chloro-2-mercaptobenzothiazole (CMBT), α-cyano-4-hydroxycinnamic acid (α-Cyano), sinapinic acid and ferulic acid. Several matrices were known at the time to improve signal reproducibility (e.g. Gusev *et al.*, 1995), and it soon became common practice to use CMBT for gram-positive and α-Cyano for gram-negative bacteria. We tested instrumental parameters, some of which were inflexible; for example, the laser was set at 337 nm, and the pulse width was fixed at 3 ns. However, other parameters such as the voltage could be altered, and this was rigorously tested and eventually set at 20 kV accelerating potential. Initially, data was collected over a wide range of m/z values, but the highest density of mass ions was

Table 1.1 International conferences organized by the Molecular Identification Services Unit (MISU) and Applied and Functional Genomics Unit (AFGU) to showcase work achieved and future directions in proteomics using MALDI-TOF MS and tandem MS/MS and genomics. (MISU and AFGU were amalgamated in September 2009 in the new Department of Bioanalysis and Horizon Technologies.)

Date	Title
27 October 1998	**1st**: Intact Cell MALDI – A Novel Technique for the Rapid Identification of Microorganisms
14–15 June 1999	**2nd**: The Impact of the Environment on Human Infections through Molecular and Mass Spectrometric Analyses
17–18 April 2000	**3rd**: Microbial Characterisation, Diversity and Function through Genome and Proteome Analysis
25–27 June 2001	**4th**: Decoding the Microbe Using Advanced Tool of Genomics, Transcriptomics and Proteomics
1–2 July 2002	**5th**: Applications of Biomics (Genomics, Proteomics and Bioinformatics) in the Research and Diagnostic Laboratory
Workshops: 23–24 June 2004	(Workshops in genomics, proteomics and bioinformatics)
16–17 June 2003	**6th**: Disease Biomarkers and Polymorphisms in Microbes
21–22 June 2004.	**7th**: Meeting the Challenges of Infectious Diseases through advances in Developing Technologies
Workshops: 23–24 June 2004	(Workshops in genomics, proteomics and bioinformatics).
13–14 June 2005	**8th**: Sequenced-Based Approaches to Diagnosis of Infectious Disease Agents
25–26 September 2006	**9th**: Development and Application of High Throughput Systems in Diagnostic Microbiology
18–19 September 2007	**10th**: The Changing Landscape of Diagnostic Microbiology; from Decades of Traditional Methods to Applied Genomics and Proteomics
26–27 June 2008	**11th** Use of New Technologies to Further Understand the Biology of Transient and Host-Derived Human Pathogens
25–26 June 2009	**12th**: Target Molecules and Biomarkers in the Characterisation of Microbes in Disease and the Environment
24–25 June 2010	**13th**: Microbial Infections: Novel Approaches to Looking at Old Problems
23–24 June 2011	**14th**: Exploration of Novel Technologies for Biomarker Discovery and Point of Care Diagnostics
21–22 June 2012	**15th**: Unlocking Genomic and Proteomic Signatures for Functional Characterisation of Human Pathogens
Supplementary meeting: 4–5 April 2012	Microbial Diagnostic Applications of Mass Spectrometry (Jointly between PHE and the University of Minho, Portugal)
27–28 June 2013	**16th**: Microbial Subtyping in Disease and the Environment; the Pivotal Role of Reference Collections
26–27 June 2014	**17th**: The Power of the Genome and Proteome in Public Health Interventions (Joint meeting between PHE and The Royal College of Pathologists)
25–26 June 2015	**18th**: Applications of High Throughput Genomics and Proteomics in Infection
23–24 June 2016	**19th**: Next Generation Genomics and Proteomics, Advances in Microscale Analysis (Joint meeting between PHE, Animal and Plant Health Agency and Middlesex University, London)

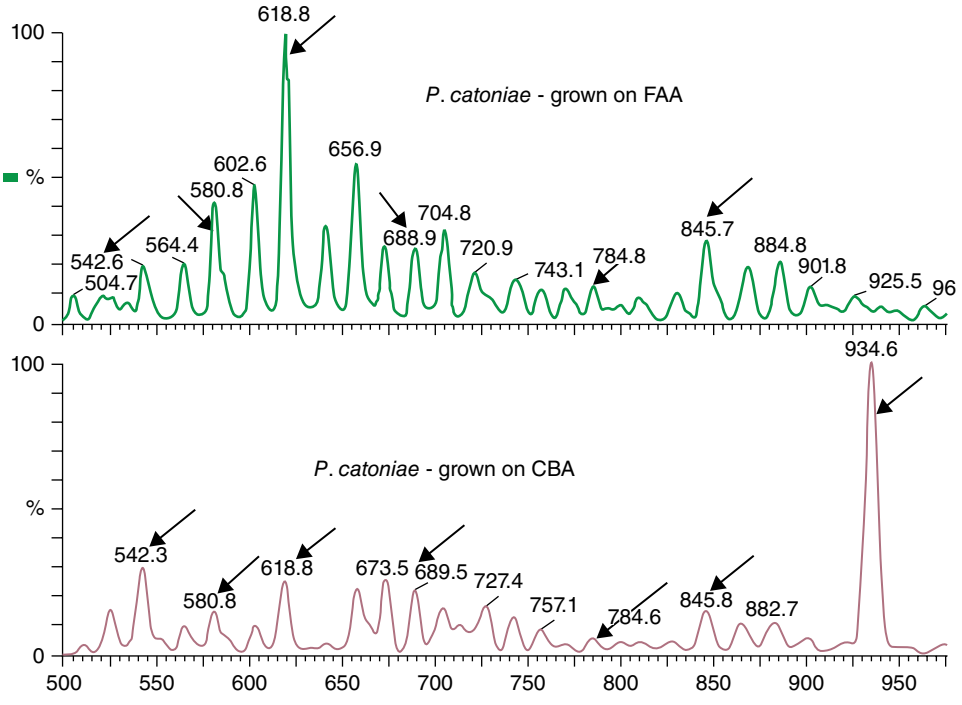


Figure 1.4 Changes in the MALDI-TOF-MS profile of the same strain of *Porphyromonas catoniae* (NCTC 12856) grown of Fastidious Anaerobic Agar (FAA) and Columbia Blood Agar (CBA). Many of the significant mass ions, e.g. 542, 580, 618, 689, 784 and 845 Da, are retained. However, significant mass ions such as 935 Da are present only in cells grown on CBA.

found to lie within the m/z range of 500 to 10,000 kDa. Thus, most samples were analyzed within this range. This was subsequently extended to 20,000 kDa. Once these constraints were set, various parameters that affect the stability of the mass spectrum were meticulously investigated. Media types and suppliers were tested (see Figure 1.4; Shah *et al.*, 2000). The effect of spores on ionization, impact of temperature, pH, growth phase and cell density were all investigated and used to derive a standard protocol. Data analysis was achieved by using a modified Jaccard coefficient and UPGMA to analyze and interpret interrelationships between strains. At each stage, the progress made was incorporated into an annual report presented at a series of international conferences (Table 1.1).

Following considerable preliminary work using this instrument, recommendations for improvement were made in 1999. The base of the instrument was quite large because of the need to accommodate the horizontally positioned flight tube. A suggestion to reposition the flight tube vertically would shrink its width and so enable the instrument to fit into the normally cramped space of a clinical laboratory. Other recommendations included the need to automate the many operational steps and redesign the fragile 20-well target plate (Figure 1.2), two of which were used as standards for lock mass corrections while those at the extreme ends of the target plate yielded unreliable results.

Figure 1.5 The first dedicated linear MALDI-TOF mass spectrometer for microbial identification used for creating the first MALDI-TOF MS database in 2004 (Keys *et al.*, 2004). The instrument, manufactured by Micromass (Manchester, UK), remedied many of the shortcomings of the Kratos Konpact Alpha and had the capacity to analyze 96 samples automatically.

A new instrument with the designed features described above was built by Micromass (Manchester, UK) in late 1999 and became the first upright bench-top linear and reflectron MALDI-TOF MS instruments (see Figure 1.5). The linear TOF instrument designated M@LDI was delivered to our laboratory on 12 April 2000, on a long-term loan to develop its applications.

This was two working days before the third annual conference on 17–18 April 2000 (see Table 1.1) and just prior to the Congress of the Confederation of Anaerobic Societies, Manchester, 10–12 July 2000 (Figure 1.6A). Work on poorly characterized anaerobic species was presented by author HNS in 2001 at the symposium 'An Anaerobe Odyssey' in UCLA, Los Angeles, CA, to demonstrate the optimism that was already developing (Figure 1.6B). PHLS underscored this momentous event in its 2001 yearbook, during which these early results were highlighted. This instrument, reconfigured with a vertical flight tube, utilized a solid stainless steel 96-well target plate which comprised rows of 12 sample wells with a central position between each of four wells for lock mass correction. Author HNS incorporated it into the clinical services of the Identification Service Unit, which published its first promotional flyer in 2000, a decision that emanated from its confidence in the use of MALDI-TOF MS for microbial identification (Figure 1.7).

(A)

ANAEROBE 2000; the first Congress of the Confederation of Anaerobic Societies: 10 – 12 July, 2000 in Manchester, England

Society for Anaerobic Microbiology (UK); Anaerobe Society of the Americas, Japanese Association for Anaerobic Infection Research

Micromass Satellite Symposium
New Techniques in Mass Spectrometry
Tuesday, July 11, 2000

Micromass UK Ltd. will offer a seminar on "New Techniques in Mass Spectrometry" during the Lunch Period on Tuesday, July 11, 2000. The session will be presented by Prof. Haroun N. Shah and will address how mass spectrometry techniques (MALDI-TOF MS) enable bacteria to be identified with unprecedented selectivity and speed. This workshop is intended both to illustrate the potential of these emerging typing strategies and stimulate debate on their practicability.

(B)

2001: An Anaerobe Odyssey
August 12, 2001
UCLA Faculty Center, Los Angeles, CA

The **Anaerobe Society of the Americas** and the **Infectious Disease Asssociation of California** present ***2001: An Anaerobe Odyssey*** at the UCLA Faculty Center in Los Angeles, CA. The symposium celebrates the 80th birthday of **Dr. Sydney M. Finegold,** noted anaerobe researcher and Founding President of the Anaerobe Society.

Program:

- **Ellen Baron**: *Speculations on the Microbiology Laboratory of the Future*
- **Diane Citron**: *Changes in the Genus Fusobacterium*
- **Sydney Finegold**: *The Possible Role of Intestinal Flora in Autism*
- **Ellie Goldstein**: *Therapy of Anaerobic Infections*
- **David Hecht**: *Evolution of Anaerobic Susceptibility Testing*
- **Hannele Jousimies-Somer**: *The Taxonomic Evolution of Gram-negative Anaerobes*
- **Carl Nord**: *The Status of Research on Anaerobes in Europe*
- **Haroun N. Shah**: *MALDI-TOF Mass Spectrometry and Proteomics: A New Era in Anaerobic Microbiology*
- **Dennis Stevens**: *Clostridial Toxins*
- **Kazue Ueno**: *The Status of Research on Anaerobes in Japan*
- **Hannah Wexler:** *Outer Membrane Proteins of Gram-negative Anaerobes*

Figure 1.6 Early meetings on MALDI-TOF MS were focused on poorly defined anaerobic species to emphasize, even at that time, the resolution of the method. Many of these species are non-fermentative so that methods involving the use of API or various biochemical tests, which were the primary methods then, were negative. New species were proposed mainly on the basis on DNA-DNA reassociation, which could not be applied in a clinical laboratory. MALDI-TOF MS when introduced contributed significantly to a resurgence of interest in anaerobic microbiology because of its capacity to resolve such complex taxonomic problems. (A) As early as the year 2000, a specific symposium was held to demonstrate the high resolving power of MALDI-TOF MS in delineating poorly defined anaerobic species in the United Kingdom and (B) in 2001, in the United States.

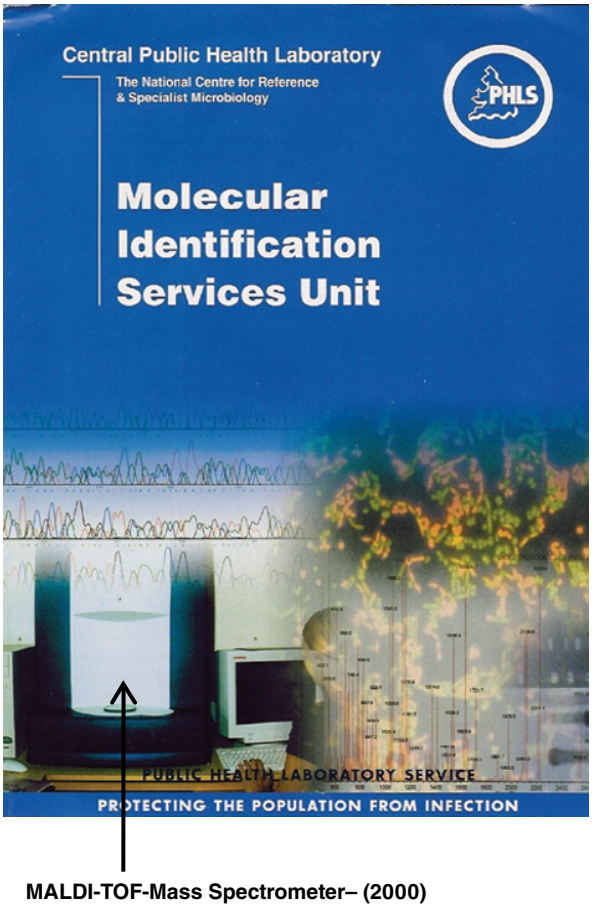


Figure 1.7 The first promotional service flyer for the Molecular Identification Services Unit (MISU, PHLS), which was printed and advertised from 2000. It signalled a strong declaration of MISU's vision to elevate MALDI-TOF MS as its principal method of microbial identification by placing a M@LDI-TOF mass spectrometer (Micromass, Manchester, UK) in the central position on the flyer.

With the basic parameters for a microbial database established, our major goal over the following five years was to develop a database of mass spectral profiles. Funding for this came through a collaborative multicentre venture with Manchester Metropolitan University and Micromass (Floats Road, Manchester) for a five-year intensive programme. The first year of the project was used to meticulously optimize protocols and interrogate the software and search engine. The National Collection of Type Cultures (NCTC) is part of the PHLS, and its vast range of type and reference strains were used as the resource for creating the database. For development of the database, NCTC's strains were sent out independently and blindly to each of the three collaborating laboratories for analyses. Twenty strains per week were analyzed over 50 weeks annually with each strain being analyzed 12 times at each site. Thus, during the first year, 36,000 spectra were collected and analyzed. Operation of the mass spectrometer was

performed using a specially designed software designated MassLynxTM. Automated calibration of the TOF tube, followed by automated acquisition of the bacterial spectra, was then performed using the real-time data selection (RTDS) function in the MassLynxTM software. Spectral profiles were collected in the mass range 500–10,000 kDa, acquiring 10 shots per spectrum at a laser firing rate of 20 Hz. Fifteen spectra per sample well and 10 spectra per lock mass well were collected using the RTDS option to optimize the collection of quality data. For database inclusion, the spectral reproducibility between the 12 replicates per sample was tested using a root mean square (RMS) calculation to identify and reject outliers at a value greater than 3.0. The RMS is the normalized deviation of the median of test spectra from the spectral average and therefore was used to compare each replicate spectrum in turn to the composite spectra of the remaining replicates. All verified spectra were combined to produce a composite spectral entry for each bacterium included in the database. The system was challenged with unknown and clinical isolates. Database searching was based on an estimation of the probability that the mass spectral peaks in the test spectrum are comparable with the database spectrum. A list of the top matches was provided together with RMS values. A high probability and low RMS value indicated a good match. The correlation between results at each laboratory was so high that after the first year, the compilation of the database was done independently and added to a composite database. The results were reported in 2004 when the database comprised 3500 spectra (see Keys *et al.*, 2004).

While this work was in progress, our laboratory, MISU, began using the instrument in tandem with 16S rRNA sequencing as part of its service function, and numerous unusual isolates that would hitherto be left unreported were now being identified (Figure 1.7). For example, the receipt of an unusual isolate from a patient who developed a severe wound infection after a visit to the Dead Sea was erroneously identified on primary culture as *S. aureus* because of its striking morphological resemblance to this species. However, its MALDI-TOF MS spectrum was significantly different, triggering further analysis, and it was subsequently identified as *Exiguobacterium aurantiacum* using 16S rRNA (Mohanty and Mukherji, 2008). The inclusion then of *E. aurantiacum* into the database enabled rapid identification of this species and prompted clinicians to send in samples from patients with similar symptoms. Within six months, 18 patients with bacteraemia were shown to have *E. aurantiacum.* A subsequent study using MALDI-TOF MS rapidly identified strains of this alkaliphilic, halotolerant bacteria from six patients with bacteraemia, three of whom had myeloma (Pitt *et al.*, 2007). Some of most difficult species received for identification by MISU belonged to the Acinetobacter, Kingella and Moraxella complex. These were extensively studied and delineated very early in this study and reported at the American Society for Mass Spectrometry 48th meeting in 2000.

Work continued on microbial identification using MALDI-TOF MS, with the database now containing nearly all the NCTC's type and reference strains. The next phase of the work involved the trial of the instrument in a hospital laboratory. Through The Royal London Hospital (University of London), it was possible to gain access to primary routine cultures over several weeks. Hospital staff were trained in sample preparation and analysis, and over 600 samples were analyzed. Because of the problems associated with MRSA and *C. difficile* infections at the hospital, priority was given to presumptive

isolates of these species. Although it was not possible to separate MRSA from sensitive strains, all samples were correctly identified to the species level (Rajakaruna *et al.*, 2009). However, the work suffered a major setback by the failure to confidently identify isolates of *C. difficile*. This was remedied through a collaboration which began in 2007 with a small company, AnagnosTec GmbH (Potsdam OT Golm, Germany), and led to a change of matrix solutions and formic acid extraction of samples prior to MALDI-TOF MS analysis. Meetings organized by AnagnosTec were held annually for four years in Potsdam Golm and comprised a small group of about 20–30 participants from various parts of Europe who had a vested interest in implementing MALDI-TOF MS as a diagnostic tool (Figure 1.8). Most of the motivation to establish MALDI-TOF MS in clinical laboratories was concentrated in Europe, as the United State's FDA approval seemed a distant goal. This is reflected in the authorship of the book *Mass Spectrometry for Microbial Proteomics* (Wiley, 2010).

Initially, the instrument was set up to profile the surface molecules of cells, the rationale being that differences between virulent and avirulent strains, where the pathogenic potential is due to surface-associated molecules, could be mapped and used for detection of pathogenic variants. For some species such as *Peptostreptococcus micros*, this was highly successful, where resolution of the two pathotypes ('smooth' and 'rough' variants) were readily distinguishable through characteristic mass ions (see Rajendram, 2003). However, to obtain such mass spectra, it was necessary to use rigorously standardized parameters that may alter the morphology of cells grown on agar plates. In our experience, a more fruitful way to approach this is the use of MALDI-TOF MS with ProteinChip Arrays (designated surface-enhanced laser desorption/ionization TOF MS, or SELDI-TOF MS).

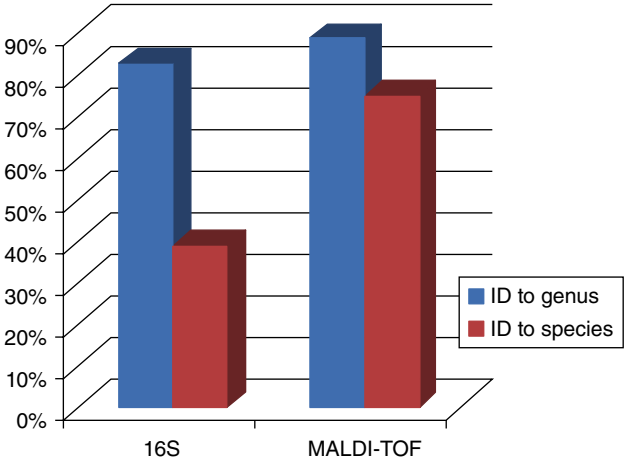


Figure 1.8 Identification of clinical isolates received by MISU using 16S rRNA and MALDI-TOF MS (Micromass, UK) over a 10-year period. MISU receives atypical, rarely isolated and emerging pathogens, and for such unusual isolates, the results shows that MALDI-TOF MS was significantly more useful in identifying these isolates to the species level.

1.3 SELDI-TOF MS: A Powerful but Largely Unrecognized Microbiological MALDI-TOF MS Platform

Microorganisms constantly undergo microevolution, and the expression of these changes is reflected in proteins; consequently, we envisaged that these biomolecules would be a rich source of biomarkers. Because proteins dictate virtually all the biological functions of a cell, we proposed to systematically catalogue cellular proteins of key pathogens and investigate how they change in response to disease and environmental factors including exposure to antibiotics. Initially, some investigators attempted to deduce components of the bacterial cell envelope using linear MALDI-TOF MS (see, e.g. Claydon *et al.*, 1996). We were acutely aware at the commencement of this programme that this instrument would only provide diagnostic profiles for microbial identification. To unravel the microbial proteome, various forms of higher-resolution MS would be necessary and, at that time, several gel-based methods were also indispensible prior to MS analysis.

Through a collaborative programme with Ciphergen Biosystems, we acquired the ProteinChip® System, designated SELDI-TOF MS. The ProteinChip technology has the capacity to rapidly perform separation, detection and analysis of proteins (at the femtomole level) directly from biological samples. Key components of the technology are ProteinChip arrays that comprise minute chromatographic wells that contained either a chemical (anionic, cationic, metal ion, hydrophobic, hydrophilic, etc.) or biochemical (antibody, receptor, DNA, etc.) surfaces to capture specific classes of proteins from a crude sample (see Figure 1.9). In practice, a few microlitres of a cell extract may be dispensed onto the ProteinChip surfaces, a quick wash is performed to remove unbound proteins and interfering substances and the sample allowed to air-dry for a few minutes. The matrix solution (sinapinic acid) is then added and samples analyzed by MALDI-TOF MS using the Ciphergen Biosystems instrument. The resulting molecular ions of the proteins, which remained bound to the ProteinChip surface, are deduced in minutes from its mass spectrum (see Figure 1.9). Because of the chemistry of the ProteinChip, the system has the capacity to capture a range of biomolecules. Our first applications of the technology for characterization of microbes were presented at the fourth of the annual conferences (Table 1.1) together with other potential applications. For example, antibodies may be covalently immobilized onto the ProteinChip Array surface by an initial incubation and washing, after which antigens may be specifically captured and analyzed directly to determine their intact masses. The potential to replace many conventional ELISA systems therefore existed. Furthermore, for epitope mapping, the bound proteins may be enzymatically digested by on-chip incubation with endoproteases. Following a wash to remove the unbound fragments, SELDI analysis may be used to identify the retained fragments and so enable characterization of the epitope. Through various grant-funded projects, MISU characterized a number of pathogenic determinants such as the botulinum toxins to find alternatives to animal experiments, determined pathotypes of *Peptostreptococcus micros* (Rajendram 2003), adhesins of *Enterococcus faecalis* (Reynaud *et al.*, 2007), etc. (see review by Shah *et al.*, 2010).

The potential of SELDI-TOF MS has been recognized by only a few microbiologists and consequently was never fully exploited. As early as 2001, the SELDI-TOF MS technology was interfaced with the Applied Biosystems QSTAR Hybrid LC/MS/MS to

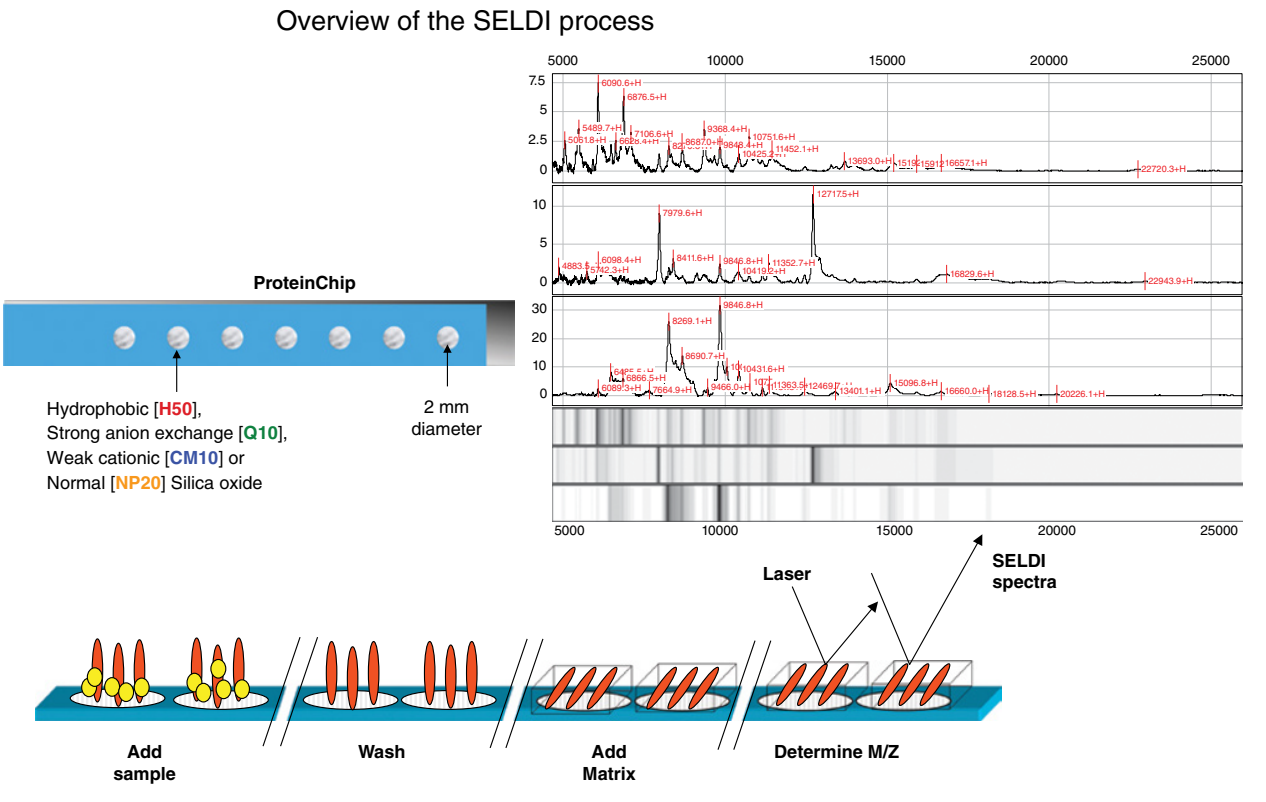


Figure 1.9 Some of the most popular types of ProteinChip arrays (H50, Q10, CM10 and NP20) used for analysis of microbial cells extracts and an overview of the process to obtain a mass spectral profile. From left to right, the sample is added to the ProteinChip array, the wells washed and air-dried for a few minutes, the matrix (Sinapinic acid) added, followed by mass spectral analysis to yield the spectra shown.

identify biomarkers with an upper limit of <4 kDa. This significantly expanded the scope of SELDI to include characterization of proteins, protein interactions and structural analysis. Had this developed further, SELDI-TOF would have had a greater presence in microbiology. Ciphergen Biosystems was acquired by Bio-Rad in 2006 and with their long-standing interest in characterizing microbes using SDS-PAGE methods, this again should have stimulated interest among microbiologists. The company phased out the early instruments, and customers were left with mass spectrometers that could not be serviced; interest rapidly declined. In July 2009, a similar attempt was made to combine Bio-Rad's ProteinChip (SELDI)-based technology with Bruker's ultrafleXtreme MALDI-TOF/TOF mass spectrometer, but as before, this had little impact among microbiologists and interest rapidly declined. As a typing tool, we believe that systems such as SELDI-TOF MS offer considerable advantage over traditional MALDI-TOF MS (see review, Shah *et al.*, 2005) and used this approach recently to delineate the heterogeneity of *Propionibacterium acnes.* For the first time, mass ions derived from both MALDI- and SELDI-TOF MS were used as new criteria to propose two new subspecies (see Chapter 5; Dekio *et al.*, 2015) and supports the confidence MALDI-TOF MS inspires today.

1.4 MALDI-TOF MS as a Platform for DNA Sequencing

Although MALDI-TOF MS is widely used for proteomic and peptide mass fingerprinting, very few laboratories worldwide have applied this powerful technology for microbial genomics. A barrier to its direct analysis stems from the instability of DNA when subjected to MALDI-TOF analysis because, during desorption, base protonation causes rapid destabilization of the *N*-glycosidic bond, causing base loss and fragmentation at many positions along the DNA (Figure 1.10). This is in contrast to RNA, whose 2′-hydroxyl group enables greater stability. In practice, primers tagged with a T7 promoter enable transcription of DNA amplicons to the more 'stable' RNA molecules. These are cleaved at nucleotide bases, and their molecular weights readily identified by MS and compared with the simulated spectra of reference sequences. Not only is the method rapid and very accurate but base modifications such as methylation and acetylation are readily detected. Sequenom GmbH was the first company to exploit this phenomenon and initially focused on single-nucleotide polymorphisms for which viruses were common targets (see Chapter 8). As early as 2002, the authors (HNS and SEG) began to investigate the potential of this technology for microbial genotyping, and early data were reported by Dirk van den Boom at the fifth annual meeting (Table 1.1). Data obtained was used to support a successful joint scientific grant application in 2005 to expand the depth of the technology for microbial typing (Honisch *et al.*, 2007, 2010). Among the applications developed using the Sequenom MassARRAY® System was the transfer of the traditional serotyping system of Kauffmann and White for typing of Salmonella, which because of identifiable mutations in the genes encoding the 'O' and 'H' antigens, was readily transferable (Bishop *et al.*, 2010).

The long awaited first visit of Franz Hillenkamp, the pioneer and inventor of the technique he called matrix-assisted laser desorption/ionization TOF-MS, took place at the Gordon Museum, Guy's Campus, University of London, on 13 December 2007, through an invitation of the London Biological Mass Spectrometry Discussion

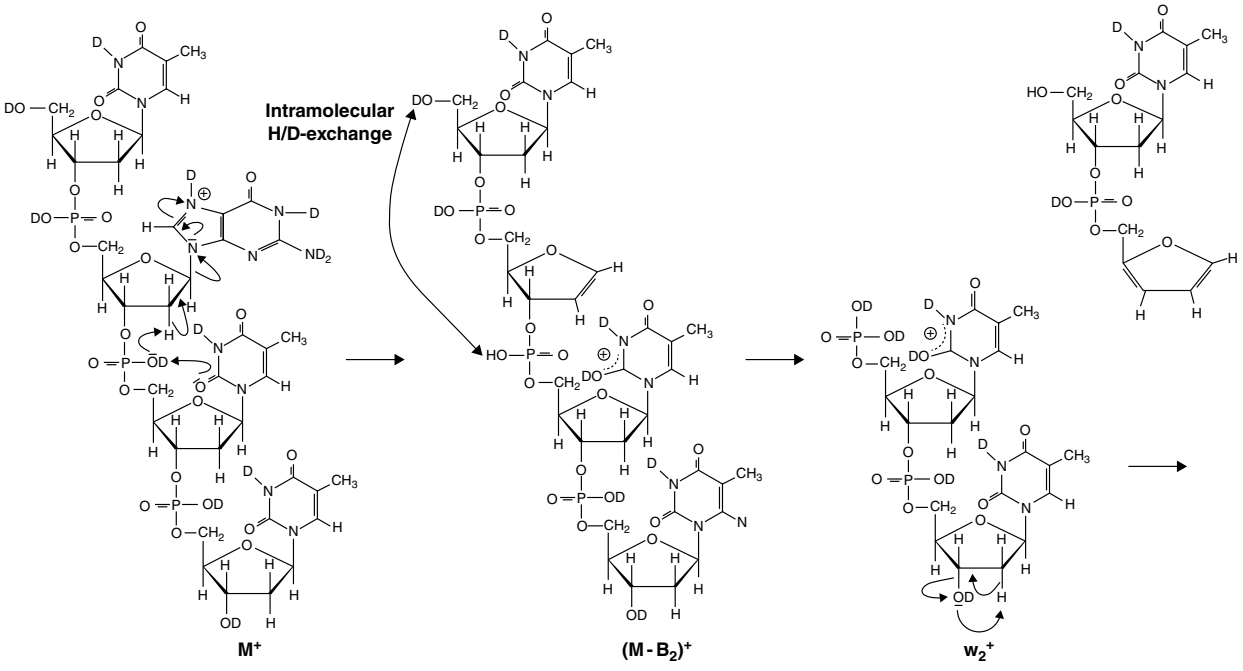


Figure 1.10 Proposed mechanism for the degradation of DNA in a MALDI-TOF mass spectrometer. This was taken from Franz Hillenkamp's notebook during his visit to MISU and AFGU in 2008. See text for details.

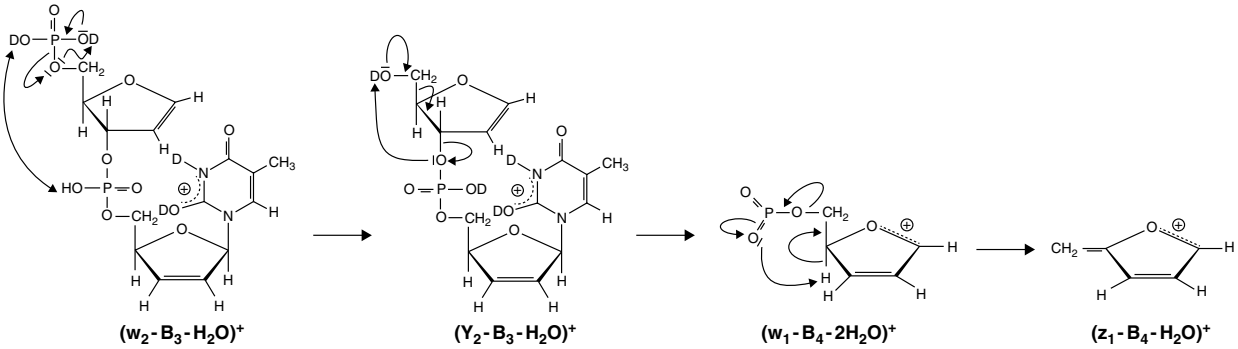


Figure 1.10 (Continued)

Group (LBMSDG). In a room packed to capacity with enthusiasts, Hillenkamp delivered a personal, riveting and moving lecture on the history of the technology; generously acknowledging his competitors and expressing his pride in seeing the applications of the technology today 'in areas he had not foreseen'. He generously complimented the work of our laboratory, and in the midst of his presentation invited author HNS to present some of the biological applications of the technology that were in use at PHE. He then resumed his presentation and speculated on future applications.

This visit led directly to his second trip to London the following year where he was invited to deliver the Plenary Lecture 'MALDI-TOF MS: Development of an Analytical Tool for Biological Sciences' at our 11th annual meeting (Table 1.1). He spent four days in the laboratory working with staff and PhD students and left a lasting impression on all who met him as a humble, gifted scientist who had a lifelong passion for science. The scientific community strongly believed he should have shared the Noble Prize in 2002 for Chemistry with Koichi Tanaka for the development of MALDI-TOF MS. He dismissed this as inconsequential and basked in the practical applications of his pioneering work. He took up an advisory post with Sequenom (which moved from Germany to San Diego, CA) and co-authored a chapter titled 'DNA Resequencing by MALDI-TOF MS and Its Application to Traditional Microbiological Problems' (Honisch *et al.*, 2010).

1.5 Insights into the Proteome of Major Pathogens 2005–2009: Field Testing of MALDI-TOF MS

The award of a major five-year scientific grant in 2005 to the authors for integrating genomics and proteomics for biomarker discovery (entitled 'Detection of virulence and species biomarkers of deliberate release pathogens using an integrated genomic-proteomic high resolution platform' (S.E. Gharbia and H. N. Shah for about £2 million) served as a springboard for the development of proteomics in PHLS (by then known as the Health Protection Agency [HPA]). Prior to this, our proteomics research was undertaken by PhD students. This major programme enabled the appointment of several dedicated staff and the purchase of a plethora of proteomic equipment including, a MALDI Reflectron, Sequenom's MassARRAY® System, various gel-based equipment, imaging systems for DIGE, robotic stations for trypsin digestion, etc. However, it was the acquisition of the new and novel Thermo Electron Corporation's nano-LC LTQ Orbitrap mass spectrometer system which propelled the organisation into advanced proteomics and had a dramatic impact in the field. At the eighth annual conference in 2005, the need for a London-based biological MS group took root and was eventually established by Anthony Sullivan as the LBMSDG, which up to now holds quarterly meetings in central London and attracts some of the pioneers in the field.

Alexander Makarov invented the Orbitrap, which subsequently led to a whole series of Thermo Fisher Scientific instruments that are used for both bottom-up and top-down proteomics. Michaela Scigelova, who worked closely with Makarov (Scigelova and Makarov, 2006), expounded the virtues of the Orbitrap in her presentation "Advances in the analysis of biomolecules using high resolution hybrid mass spectrometers" at our ninth annual conference in 2006. Although Makarov had presented

proof-of-principle results of the Orbitrap analyzer at the ASMS conference as early as 1999, commercial introduction of this analyzer by Thermo Fisher Scientific, as part of the hybrid LTQ Orbitrap instrument, did not materialize until 2005. With the timely success of this grant, we were therefore the recipient of one of the first instruments, and work began almost immediately. Proteome analysis of serovars Typhimurium and Pullorum of *Salmonella enterica* subspecies (Encheva *et al.*, 2005, 2007) and *Neisseria gonnorrhoeae* were reported (Schmid *et al.*, 2005). A basic proteome reference map of *Streptococcus pneumonia* was published in the journal *Proteomics* (Encheva *et al.*, 2006) while the Sequenom Mass Cleave technology was by then firmly established as a microbial typing tool in our laboratory (Honisch *et al.*, 2007). The bridging of genomics and proteomics was now established (see, e.g. Al-Shahib *et al.*, 2010; Misra *et al.*, 2012).

1.6 2010–2011: The Triumph of MALDI-TOF MS and Emerging Interest in Tandem MS for Clinical Microbiology

During the first decade of the millennium, ideas that were being explored at the turn of the century began to take root. There was a period of considerable development both in hardware, software and bioinformatics tools. The latter was already an integral part of genomics and was by then beginning to underpin developments in proteomics. Whole genome sequencing was becoming a reality for the clinical research laboratory, and with the vast amount of microbial sequence data accumulating, protein identification by MS/MS analysis, which is dependent on sequence data, accelerated significantly. The burgeoning field of proteogenomics was now inseparable and began expanding rapidly. The organization restructured its laboratories accordingly to facilitate this interaction. Thus, in 2010, Proteomics and Genomics Services and Research were amalgamated into one large specialty designated Department of Bioanalysis and Horizon Technologies, bringing genomics and proteomics laboratories in juxtaposition. Even though developments were gradual, a decade of work seemed to have reached a new milestone in 2010. New bioinformatics approaches were being used to characterize microbial biomarkers (see, e.g. Shah *et al.*, 2011). Bottom-up analysis of MS/MS data was permitting in-depth analysis of the microbial proteome, and new insights into their structure, metabolism and pathogenicity were being reported (see reviews in Shah and Gharbia, 2010). A notable example was the elucidation of the outer membrane proteome of *Salmonella enterica* serovar Typhimurium utilizing a lipid-based protein immobilization technique which was undertaken in collaboration with Roger Karlsson's laboratory in Sweden (Chooneea *et al.*, 2010). This work has now advanced as a means of proteotyping microbial species, which significantly expands the use of proteomics for microbial characterizations (see Chapter 16).

Bacterial pathogens considered category A biothreat agents such as *Bacillus anthracis, Francisella tularensis, Clostridium botulinium* and *Yersinia pestis* or category B agents such as *Burkholderia pseudomallei* and *Burkholderia mallei* require work in Biosafety level 3 cabinets (Rotz *et al.*, 2002). Because of the need to undertake all preparative work at such a high containment level, work on such pathogens was restricted to few laboratories and focused mainly on genomics. With the arrival of

nano-LC systems coupled to the LTQ Orbitrap, it was possible to work with microlitre volumes of cell extracts and undertake analysis of such pathogens. The number of strains and species could now be scaled up for MS/MS analysis. Consequently, we were able to develop a pipeline for biomarker discovery using MS and bioinformatics to included category A pathogens (Al-Shahib *et al.*, 2010). Among species where differentiation from closely allied taxa were contentious by genomic methods, unique strain-specific peptides were found that delineated each clearly (see Chapter 13) and was reported at the 20th meeting of the European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) in Vienna in 2010.

ECCMID, because of its strong focus on microbial diagnostics has, over the years, become a vehicle for reporting developments in the application of MALDI-TOF MS for microbial identification, typing and antimicrobial resistance detection. In previous years, lecture theatres would be sparsely occupied during presentations on MALDI-TOF MS. At ECCMID in Glasgow, 10–13 May 2003, the author's presentation on MALDI-TOF MS in microbial identification drew 15 people in a room with a capacity for 150. However, at ECCMID 2010, the largest lecture halls could not accommodate all who wished to attend, signalling a significant turning point in the acceptance of this technology. There was a dedicated symposium of invited speakers titled 'MALDI-TOF in Clinical Microbiology' (11 April 2010) in which our presentation titled 'MALDI-TOF MS of Surface-Associated and Stable Intracellular Proteins for Identification and Resistance Profiling of Human Pathogens' covered the work of the last decade. Many groups of species in which we previously encountered difficulties in obtaining confident identification, such as *Clostridium difficile, Bacillus* species or mycobacteria, were now being resolved and reported with improved confidence scores (e.g. Figures 1.11 and 1.12). The three other speakers of this session echoed this confidence to a large audience. A Poster Session on 12 April titled simply 'MALDI-TOF' reinforced this: there were 27 poster presentations on MALDI-TOF MS. The following day ended with an oral session titled 'What Can We Expect from MALDI-TOF?' in which there were 10 presentations. This immense endorsement of MALDI-TOF MS at one meeting led to a feeling of reflective optimism following a decade of persistent work and coincided with the publication the book *Mass Spectrometry for Microbial Proteomics* (Shah and Gharbia, 2010). Displayed at the publisher's, Wiley's, desk at ECCMID 2010, copies were sold out to the new wave of transformed microbiologists.

Perhaps the biggest barrier to acceptance of this radical change to MALDI-TOF MS was for microbiologists to leave behind the deeply rooted role that morphological and biochemical tests played in bacterial identification since its inception in the late nineteen century. The first chapter of the book, titled 'Changing Concepts in the Characterisation of Microbes and the Influence of Mass Spectrometry' (Shah *et al.*, 2010), charts the course of determinative microbiology and makes the case for this radical change to MALDI-TOF MS. Fundamental changes had already taken place in microbiology with the accessibility of DNA sequencing as described in the following chapter, 'Microbial Phylogeny and Evolution Based on Protein Sequences – The Change from Targeted Genes to Proteins' (Gupta, 2010), but clinical microbiologists still alluded to this as the domain of the research laboratory. Thus, nearly all the presentations at ECCMID 2010 used biochemical tests as a comparator to access the performance of MALDI-TOF MS. It is interesting that bioMérieux, one of the global leaders of biochemical tests, revealed their acclamation of MALDI-TOF MS at ECCMID 2010 by the

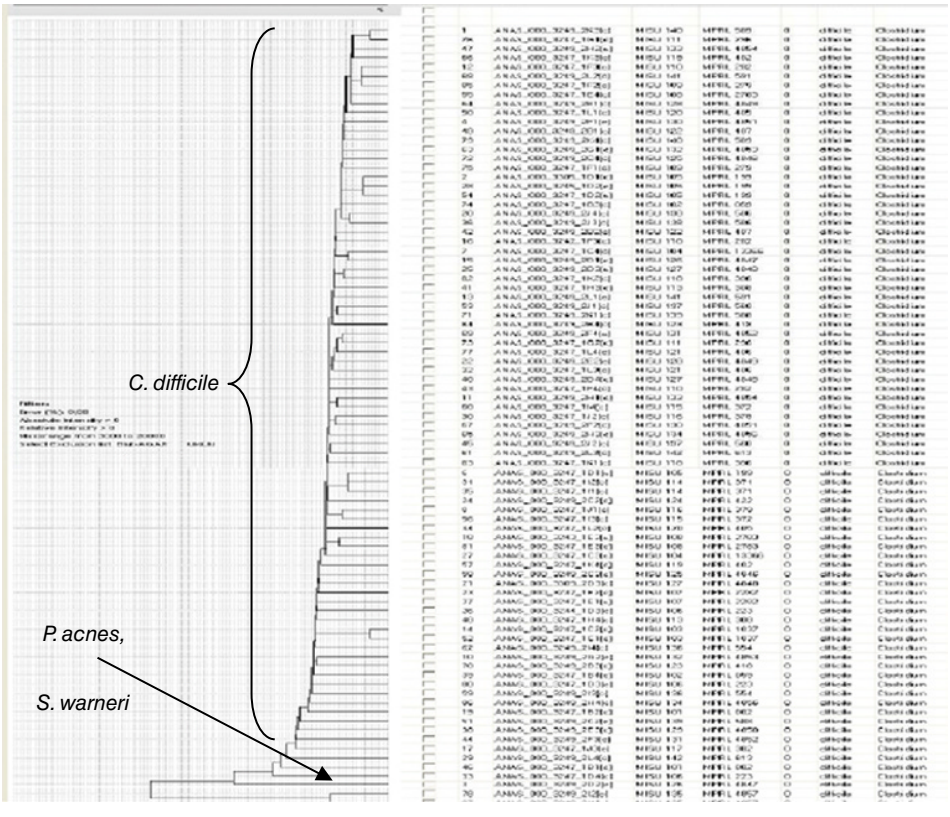


Figure 1.11 Identification of *Clostridium difficile*. Initially this posed a major problem for accurate and reproducible identification. The dendrogram shows some100 clinical isolates of *C. difficile* clustering in a single phenon and distant from other species such as *Propionibacterium acnes* and *Staphylococcus warneri*, with which it initially formed a common cluster.

acquisition of AnagnosTec in a sweeping public display in which they substituted all AnagnosTec marketing literature with bioMérieux's promotional material during the Trade Exhibition. bioMérieux states on their web site: 'Right from the moment it launched, API® completely revolutionized the field of bacteriology. API® brings together high quality and ease of use with standardized, miniaturized strips of biochemical tests to use with comprehensive identification databases. With API®, bacterial and fungal identification is simple, rapid and reliable' (http://www.biomerieux-diagnostics.com/apir-id-strip-range). To many, this change seemed extraordinary because although a few bioMérieux's staff visited and worked at PHE for short intervals on MALDI-TOF MS, the company had no experience in MS, while Shimadzu, which manufactured the instruments, remained anonymous.

Thermo Fisher Scientific, which has an enormous global portfolio in MS, also had a strong presence at ECCMID 2010 but from their microbiological products perspective. They too were strongly influenced at ECCMID 2010, and a meeting was set up with the authors on the first evening to explore ways in which they might enter microbial MS. We already had five years experience with Thermo Fisher's LTQ Orbitrap and presented

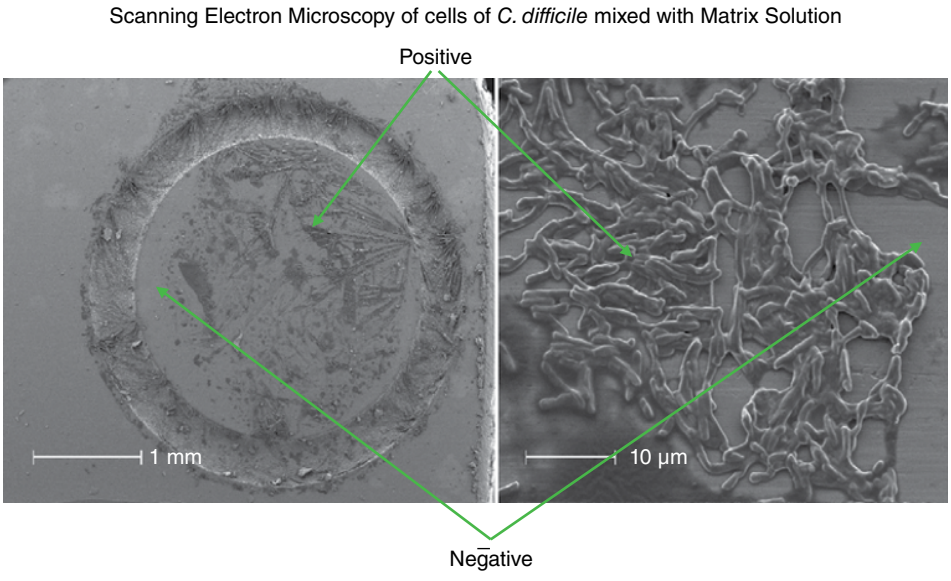


Figure 1.12 Scanning electron micrographs of *Clostridium difficile* cells mixed with the matrix solution 2,5- dihydroxy benzoic acid in acetonitrile: ethanol: water (1:1:1) with 0.3% TFA. The electron micrographs show the clumps of cells which give a MALDI-TOF MS spectrum (indicated by positive) if the laser strikes, whereas in areas where there are no cells, no spectra are obtained (negative).

a significant amount of microbial proteomics at this meeting using this instrument. The outcome of this meeting was to either develop the bottom-up approach using nano-LC MS/MS further or explore the use of a new MALDI-TOF mass spectrometer.

The euphoric end to ECCMID 2010 was followed by several hundred e-mails and communications for more information immediately following the meeting. This optimism was channelled into our 13th annual conference in June 2010 (Table 1.1) in which both MALDI-TOF MS and LC-MS/MS work on human pathogens was reported. Prominent among these was a presentation titled 'MALDI-TOF MS Detection of Low Abundance and Low Molecular Weight Proteins Using Nanoparticles' which was based on the preliminary work we had embarked upon with the National Physical Laboratory; the hypothesis being to incorporate nanoparticles to capture low-abundance proteins that normally evade mass spectral analysis.

1.7 Preparations for MALDI-TOF MS Analysis on a Grand Scale: The Looming London 2012 Olympics

Following the summer of 2011 in the which a major outbreak of food-borne illness was caused by *E. coli* (see below), attention began to focus on the summer of 2012 in which the London Olympics and Paralympic Games were due to take place. Mass gatherings both in winter and summer pose a health threat, the most common being viral respiratory tract infections, including influenza, measles and multi-resistant TB, STDs and low-dose enteric infections that show person-to-person spread. Through the consumption

and sharing of food and drink, the list of bacterial and fungal pathogens can be greatly extended. The HPA (later titled PHE) had a central role in infectious disease surveillance for these events, setting up a suite of robust and multisource surveillance systems (see, e.g. Severi *et al.*, 2012).

At the laboratory level, cases were being made for simple, rapid, cost-effective, accurate, high-throughput methods. Up to this point, all MS work in PHE from 1998 onwards developed through successful research grants by the authors. For implementation to service function, it was necessary for the organization to purchase its own equipment, and between 2005 and 2010, request was made annually by the authors but declined. With the London Olympics looming, we made a fresh case again for a MALDI-TOF MS for its service function and was this time successful. An instrument was quickly installed, and we began reporting data on patients' samples by October 2011, three days after its arrival. The immediate success of this led to instruments being purchased for other PHE laboratories in Cambridge, Southampton, Birmingham and Bristol. With such a network developing, the authors established a PHE User Group in January 2012 which held meetings and symposia. With the University of Minho, Portugal, we immediately organized a two-day conference titled 'Microbial Diagnostic Applications of Mass Spectrometry' on 4–5 April 2012 at PHE, London. Just prior to this, ECCMID held its 22nd meeting between 31 March and 3 April 2012 in London, within the vicinity of the London 2012 main Olympics venue. Bruker Daltonik sponsored a session titled 'Microbial Identification for the 21st Century – and Beyond' which was chaired by Markus Kostrzewa. A presentation by Matthew Ellington (HPA, Cambridge, Addenbrookes, UK) titled 'HPA MS Implementation Group: Current Results and Future Plans' highlighted our hope and aspirations for MALDI-TOF MS within PHE. Soon after this meeting, PHE purchased a second instrument for our site in London to meet the demands of many laboratories that were now processing clinical samples daily using the microflex Biotyper. Over the next few months, this would expand to about 20 laboratories within PHE- and NHS-affiliated laboratories and now represents the largest global network of MALDI-TOF MS instruments in a single organization.

1.8 Investigating the Detection and Pathogenic Potential of *E. coli* O104:H4 during Outbreak of 2011

On 1 May 2011 an *E. coli* outbreak began in Germany with a trickle of patients presenting bloody diarrhoea (Askar *et al.*, 2011). By the end of the month, the number of reported cases surged to 1240, including cases reported in eight other European countries (see Shah and Gharbia, 2012). The whole outbreak resulted in more than 4000 reported cases and 50 deaths. A number of patients suffered from haemolytic-uraemic syndrome (HUS), a devastating and rare disease characterized by disintegration of red blood cells, acute kidney failure, and impaired ability to clot blood. The outbreak strain was positively identified on 25 May 2011. Workers at the University of Münster and the Robert Koch Institute identified the strain through serotyping and PCR assays. Multilocus sequencing was used to confirm that the outbreak was caused by a single clone, HUSEC041, and that it had the rare serotype of O104:H4. This serotype is normally associated with enteroaggregative *E. coli* (EAEC) that are known to cause

persistent diarrhoea, but not haemorrhaging or HUS (Frank *et al.*, 2011). The strain lacked features characteristic of O157:H7, such as an enterocyte effacement pathogenicity island and an intimin-positive gene, but produced aggregative factors typical of EAEC. However, the strain did exhibit high resistance to third-generation cephalosporins, trimethoprim/sulfamethoxazole and tetracycline, which is typical of O157:H7. The strain also possessed the rarer and more potent Shiga toxin 2 gene.

DNA sequencing would provide the initial blueprint for understanding the pathogen's novel set of characteristics. In June, two independent groups completed DNA sequencing of the outbreak isolate's 5.2 Mb genome and two large plasmids using short-read DNA sequencers. Both groups released the sequencing data to the scientific community, which rapidly performed bioinformatics to explain the strain's pathogenicity and evolutionary origin. It was also suggested that the strain may harbour genes unique from those in other strains.

By early June, the first cases were being reported in the United Kingdom, and samples were sent to our laboratories for whole genome sequence analysis. Cultures were grown overnight in 5 × 100 LB broth and nutrient agar plates in a Class 3 cabinet. All extracts were subcultured into enriched broths and retested for loss of viability before any further work was allowed. Author SEG had just acquired a Roche Junior and began 454 sequencing on 7 June 2011. By 9 June, the group began to assemble the scaffolds, which was successfully completed the following day and the full sequence deposited in real time in the National Center for Biotechnology Information (NCBI) library for crowdsourcing bioinformatics analysis. This work, which was acknowledged in *Nature Biotechnology* (see editorial, 2011), provided the blueprint for a unique opportunity to explore whether it was possible to undertake proteome analysis of the cell extracts in real time using nano-LC-MS/MS.

Proteomic analysis was performed using an established workflow (Figure 1.13) on five *E. coli* strains of serotype O104: three clinical isolates from patients affected by the German outbreak and two other isolates that were previously characterized as serotype O104, but have EAEC, EHEC/STEC genetic composition, respectively. The genomes of the three German outbreak isolates were sequenced to confirm they were from the same strain. Strains were cultured on LB broth and agar as described above for DNA extracts and then harvested prior to employing two parallel approaches for reducing complexity of the mixture for MS analysis. In the first approach, lysates were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis, and 1 cm gel slices were digested with trypsin. Peptides were analyzed using nano-LC-MS/MS. In the second approach, the entire cell lysate was digested directly with trypsin in solution and injected onto two LC-MS/MS systems (LTQ Orbitrap, DBHT and LTQ Orbitrap Velos (Thermo Fisher, Hamel Hampstead), each with a front-end Ultimate 3000 Dionex nano/capillary liquid chromatography system, Thermo Fisher Scientific) that provided ultra-high-resolution and accurate masses for differentiating closely related peptides.

The recorded peptide MS/MS spectra were matched to both protein and in silico genome-translated databases to identify expressed proteins. The peptides were then fed into a bioinformatics pipeline (Al-Shahib *et al.*, 2010) to acquire unique signatures at the genus and species levels. An extensive list of identified peptides was then searched, using Blast and Scaffold, for virulence determinants, *E. coli* virulence factors and putative EHEC/STEC and EAEC-specific virulence markers. The peptide lists identified proteins

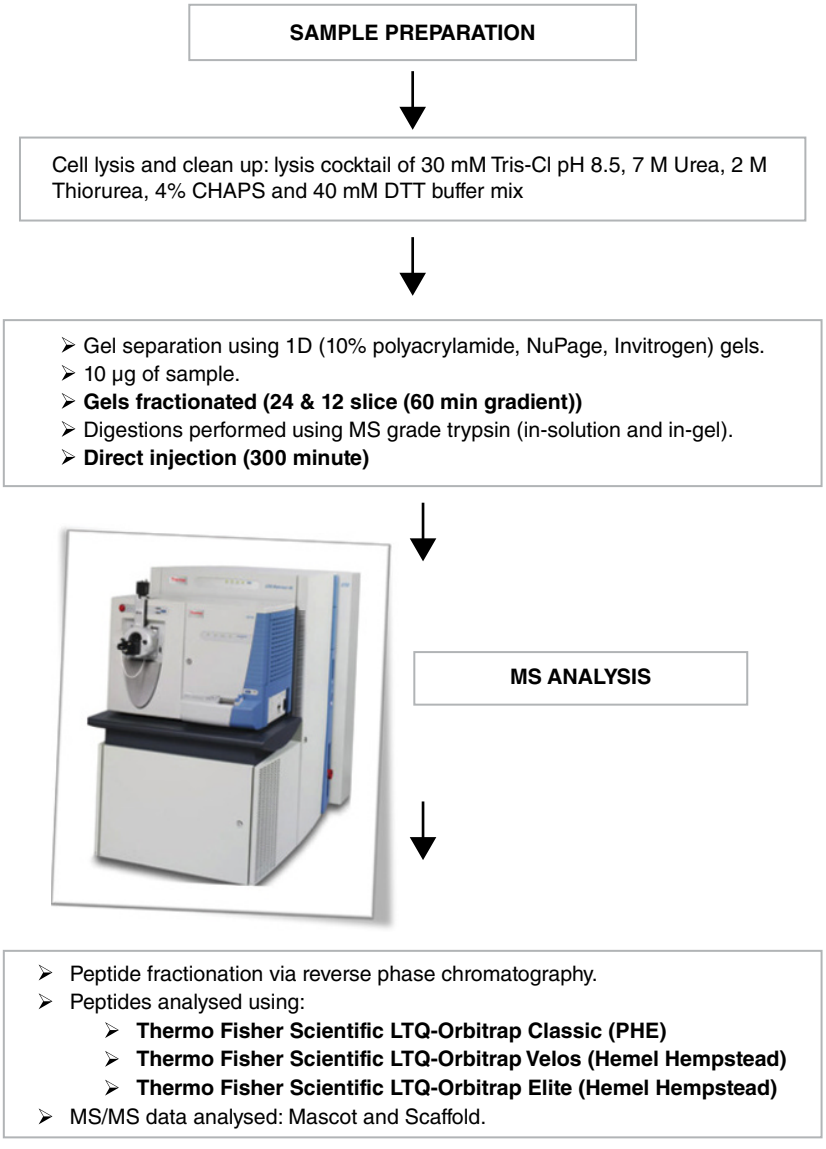


Figure 1.13 Bottom-up workflow used to deduce strain-specific peptides and virulence determinants of *E. coli* O104:H4 during the outbreak of 2011. The classical 1-D SDS-PAGE followed by in-gel trypsin digestion or directly loaded onto LC-MS/MS were also used. Three different Orbitrap mass spectrometers, with varying resolutions, were used to obtain comprehensive profiles of the proteome of strains.

that covered a significant percentage of the predicted open reading frames of the sequenced outbreak strain genome, indicating the sensitivity and reliability of the nano-LC-MS/MS method in yielding protein profiles using selective or enriched culture preparation. Peptides resulting from high-abundance proteins were then analyzed for markers and signatures that uniquely identified genus, species or virulence characteristics.

The mosaic outbreak strain's virulence signatures were compared to both EAEC and EHEC protein signatures, all obtained using the same proteome approach.

Data for isolate *E. coli* O104:H4 strain 280 was the first genome to be assembled into near complete topology. Approximately 2500 proteins from the outbreak isolates were identified. A collection of 68 peptide signatures were unique to the outbreak *E. coli* isolates and not shared by the parent source, separating the outbreak strain from other closely related Enterobacteriaceae. Species-level peptide signatures were also detected, including those for the AggR transcription factor, haemolysin protein, Aaf fimbriae protein and Iha adhesion protein. In all, 3031 peptides were identified as unique to the outbreak strains when compared against control isolates.

The technique detected features that were expected on the basis of prior laboratory tests and genomic data, including the production of Shiga toxin, Pic serine protease (autotransporter toxin) and tellurium resistance. The list of peptides was then filtered to exclude physiological and regulatory proteins. Search of the simplified list for *E. coli* pathotype virulence determinants and virulence factors resulted in a definitive list of expressed virulence determinants of the outbreak strain. The results supported the view that the background genome came from an EAEC progenitor that acquired plasmids and prophages, and exchanged chromosomal loci, leading to the emergence of an aggressive strain with a distinctive profile. All strains shared 89% of the expressed proteins. The two large plasmids encoded 31 proteins. Peptide signatures for adhesion and multidrug resistance (including β-lactamase, CTX-M extended spectrum β-lactamase and metallo-β-lactamase enzymes) were observed.

These experimental results demonstrated that a proteomic approach, based on nano-LC-MS/MS and comparison against a database of known pathogenic markers, accelerates the identification and characterization of the sources of *E. coli*-related illnesses and diseases. This study revealed for the first time that nano-LC-MS/MS was able to identify a significant number of pathogenic markers with no requirement for enrichment, selective media or antibiotic incorporation that can otherwise delay analysis. The protein signatures detected provide definitive characterization at the genus, species and often strain level, as well as detection of expressed pathogenic determinants and antibiotic resistance mechanisms. This mass-spectrometry-based approach enables clinical laboratories investigating outbreak strains to design screening and verification tests directly and in an unbiased manner, rather than performing multiple, potentially futile detection approaches while the outbreak is under way. The level of resolution achieved in this study could not be done using a linear MALDI-TOF MS and paves the way for MS/MS-based analysis for simultaneous microbial strain typing and pathotyping during outbreak investigations.

1.8.1 The Transition from MALDI-TOF MS to High-Resolution LC-MS/MS: Merits of Bottom-Up and Top-Down Proteomics for Microbial Characterization

By 2010, our laboratory had worked with all available forms of MALDI-TOF MS and now had five years experience with nano-LC-electrospray-MS/MS (see overview in Figure 1.14). Between 2011 and 2015, we investigated high-resolution MS to rapidly analyze minute volumes of complex cell extracts for strain identification and typing of expression markers. Supported by Thermo Fisher Scientific, we used three different models of the Q-Exactive instruments during the *E. coli* 0104:H4 outbreak of 2011 that

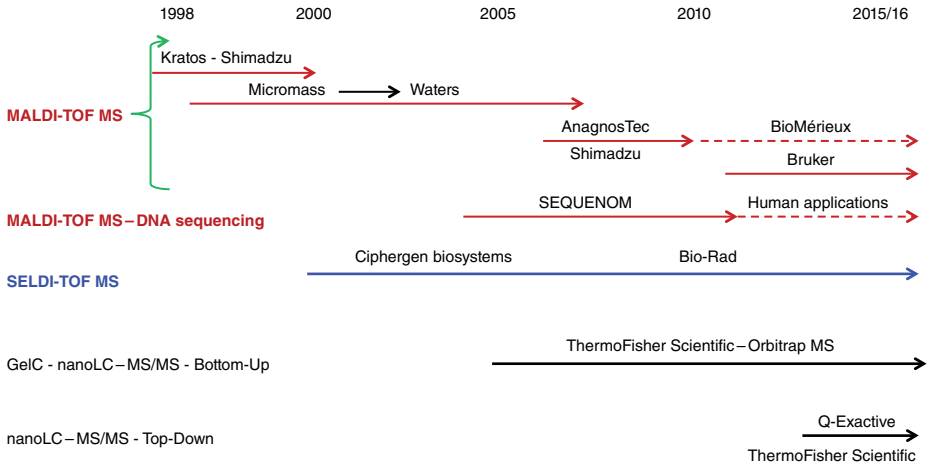


Figure 1.14 Phasing in of MS-based-methods in MISU-AFGU into clinical microbiology from 1998. These methods were used to characterize clinical isolates by our laboratories. The time lines and providers of the technology are shown in the figures. In 2010 bioMérieux acquired AnagnosTec, while Sequenom's interest shifted away from microbiology to solely human applications.

provided proof of principle that LC-MS/MS could be used to type and reveal the pathogenic potential of strains simultaneously (Shah and Gharbia, 2012).

Unlike MALDI-TOF MS, scrupulous attention needs to be given to bacterial protein extraction prior to MS/MS analysis. The challenge for the clinical microbiologists is to devise a universal protocol that encompasses the most fragile gram-negative cells to the most robust structures such as spores and the complex cell envelopes of mycobacteria. Furthermore, although effective reagents are needed, lysis methods need to avoid chemicals/detergents that cause interference with chromatography or MS. This was investigated, and the lysis mixtures that best met these criteria were designated as follows:

DIGE	7 M Urea, 2 M thiourea, 4% CHAPS, 30 mM Tris and 70 mM DTT
CHAPS	4% CHAPS, 30 mM Tris and 70 mM DTT
ACN/TFA	Acetonitrile : Water : Trifluoroacetic acid (33%:66.9%:0.1%)
ACN/FA	Acetonitrile : Water : Formic acid (50%:45%:5%)

ACN/TFA and ACN/FA were introduced to mitigate the potential problems associated with urea and were being used for top-down proteomics. As an example, two clinically relevant species representing opposite ends of the spectrum, *Clostridium difficile* and *Escherichia coli*, were tested. *E. coli* is a gram-negative, non-spore-forming, facultative anaerobe, whereas *C. difficile* is a gram-positive spore-forming anaerobe; they represent many of the physiological and phenotypic features that are likely to be encountered in the clinical laboratory and would provide a challenging panel for mass-spectrometry-based workflows.

A standard shotgun protocol for *C. difficile* and *E. coli* was, briefly, to harvest cells grown for 24 h on Columbia blood agar plates and suspend the cells in 300 µl of one of

the lysis solutions. Samples in triplicate were then lysed in one of two ways: (1) either incubating the bacteria in the lysis solution for 30 min at room temperature before centrifugation and protein extraction or (2) with mechanical lysis using the FastPrep system (MP Biomedicals). The FastPrep samples contained 100–150 µl of glass beads added to the sample and were beaten at 4 m/s for three rounds of 20 s. With lysis complete, the samples were centrifuged at 21,000 × *g* for 30 min at 4 °C, and the supernatant stored at −20 °C until LC-MS/MS analysis. If lysis is efficient, samples were separated on a 12% SDS-PAGE gel for a brief period (≈15 min) to exclude impurities and salts. Gel slices of 1 cm were excised, destained and digested overnight with trypsin. Chromatographic separation of the complex mixture of peptides derived from tryptic digest was achieved on C_{18} column over a 60 min gradient and immediately analyzed with the LTQ Orbitrap mass spectrometer. Protein identification was performed by searching databases with search engines such as Mascot, Sequest or Phenyx (IS 2011). The resultant datasets were processed using a pipeline that we established as an in-house marker discovery workflow (Al-Shahib *et al.*, 2010). For *C. difficile* there were 2383 different peptide markers specific for *C. difficile* and conserved between the three biological replicates. For *E. coli* there were 104 different peptide specific markers conserved between the three replicates. The disparity between the number of markers is due in part to the difficulties in species resolution. *E. coli* is closely related to species of the genus *Shigella*; hence, many peptides are shared, whereas *C. difficile* is phylogenetically more distinct and therefore has a greater number of unique peptide biomarkers.

Despite the marked differences between these species, the results demonstrate unequivocally that among bacterial species several hundred unique peptides are present that may be used to characterize bacterial isolates at the species and strain levels. Furthermore, markers of antibiotic resistance are evident that correlate with known antimicrobial profiles of strains (unpublished work; see abstract ECCMID, 2010). A drawback of the method is that the workflow is still too complex and lengthy for routine processing of clinical samples. However, methodologies are rapidly changing, and trypsin digestion, for example, which hitherto required overnight incubation is now done in minutes in most laboratories (Hustoft *et al.*, 2012). With processing rapidly decreasing, bottom-up proteomics will soon be achievable in a few hours. It is expected that the entire workflow will be automated, which will bring this approach into the realms of the clinical laboratory.

An alternative approach would be to utilize top-down proteomics, which is infinitely faster, capable of quantitating unique proteoforms including post-translational modifications (PTM) and avoids the need for enzymic digestion of samples (Roth *et al.* 2008). In the top-down protocol, proteoforms are delivered to the mass spectrometer intact and then sequenced by fragmentation inside the instrument, thereby retaining their critical linkage information (see overview of the method compared to the earlier bottom-up approach in Figure 1.15). This is technically more demanding in that the intact proteins are more difficult to fractionate and fragment than peptides, and more challenging to separate by liquid chromatography. Given the need to distinguish proteins varying by only small chemical differences, high-end, high-resolution instruments are required to resolve such large molecules when they are so similar in size. Dedicated software (e.g. ProSightPC 2.0) is used for analysis. A range of instruments are now amenable to top-down analysis. Prior to 2010, Fourier-transform ion-cyclotron resonance (FT-ICR) mass spectrometers were the main instruments used for top-down analysis.

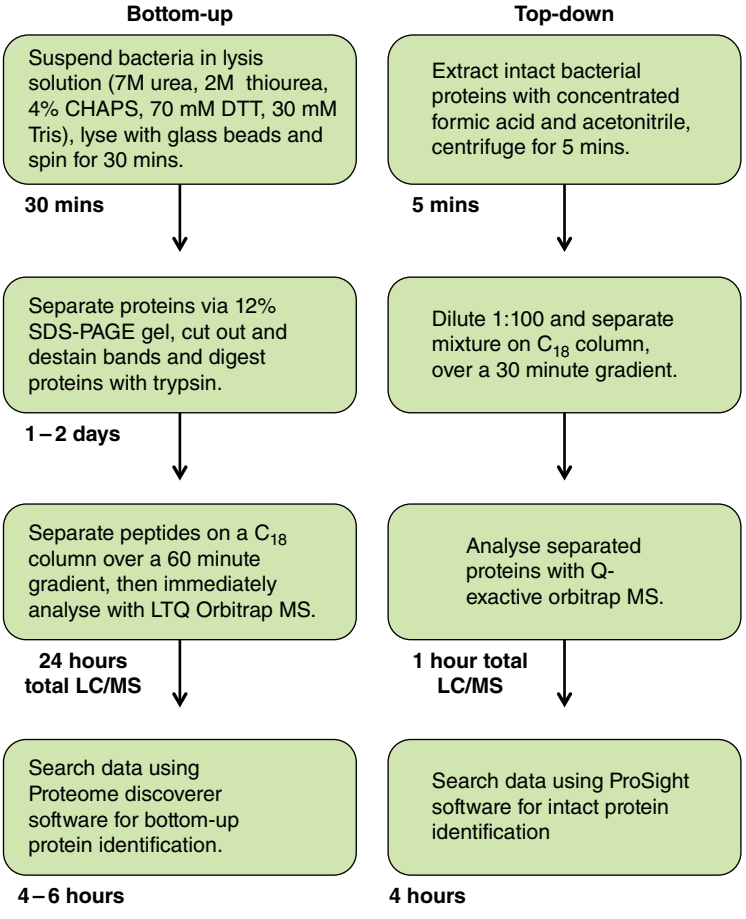


Figure 1.15 A comparative overview of bottom-up versus top-down proteomics. Although the former has been used extensively for comprehensive analysis of the proteome of several species, it is currently too cumbersome for a clinical laboratory. It seems likely that top-down based approaches will supersede MALDI-TOF MS as the next-generation approach to microbial identification and typing.

Today, more affordable quadrupole-time-of-flight (qTOF) instruments, such as the Waters SYNAPT G2-Si, the Bruker maXis ll or the new range of Orbitrap instruments (Thermo Fisher Scientific) have made the technology more accessible (Shah *et al.*, 2015).

Unlike the human proteome bacterial extracts are not subjected to the same level of complexity, and hence pre-separation methods such as those employed by Ahif *et al.* (2013) are superfluous. However, the multiplicity of the bacterial proteome is known to be significantly enhanced through PTMs (see review, Mijakovic *et al.*, 2014). These include methylation, phosphorylation, acetylation, glycosylation, pupylation, sirtuin acetylation, lipidation, carboxylation, bacillithiolation, etc. This rapidly expanding catalogue, made possible by advanced detection methods in MS, will provide a plethora of new proteoforms. Because many of these have regulatory functions, the importance of their detection in microbial pathogenicity cannot be overstated. A significant advantage of top-down analysis is its ability to detect these PTMs, sequence variants

and degradation products, markedly increasing the potential number of species/strain protein signatures.

In terms of elucidating molecular mechanisms of microbial pathogenicity, the ability to now deduce protein variations that arise from alternative splicing, allelic variation, or PTMs opens a new chapter in this field. One elegant example by Julia Chamot-Rooke's group used a combination of bottom-up and top-down MS to characterize a PTM on the major pilin (PilE) of *Neisseria meningitidis* Type IV pili. It was then shown that this modification (glycerophosphorylation), which is induced in vivo after several hours of host cell contact, is a prerequisite for the dissemination of the bacterium, a crucial step that precedes invasive infection (Chamot-Rooke *et al.*, 2011; see Chapter 18). Such applications are the real beneficiaries of these new technologies and will have a profound impact on the next phase of elucidating the diverse range of pathogenic mechanisms exhibited by microorganisms.

1.9 Conclusions

In the vast landscape of proteomics, the greatest success achieved to date in the clinical laboratory is the rapid, accurate, low-cost and simple method of identification of microbial pathogens using a linear MALDI-TOF mass spectrometer. A statement of our confidence in the method was evident since 2000 through various publications (Shah *et al.*, 2010). Having analyzed tens of thousands of clinical samples in parallel with 16S rRNA sequencing, chemotaxonomic and biochemical tests, we believe that MALDI-TOF MS has significantly surpassed all as the method of choice today, and the implementation of a network of over 20 instruments in the Health Services laboratories at present is a testament to this confidence.

There is currently no comparable achievement in clinical biomarker application for complex human and animal diseases, but many believe that top-down proteomics will achieve equivalent success in the near future (see review; Gregorich and Ying, 2014). Because the proteome of microorganisms are orders of magnitude smaller than eukaryotes, success in clinical microbiology may be more forthcoming using top-down MS. Although top-down proteomic analysis can be operated at relatively high speed in order to successfully map the bacterial proteome, the rate of identification of new proteins after initial rounds will need to be significantly enhanced for application in the clinical laboratory. Direct infusion is possible but may not be practical for automation; instead, an integrated single or dual LC with tandem MS that is completely 'hands-free' is likely to be the way forward.

There is little doubt that highly automated sophisticated instruments will soon be available. For MS/MS to reach a similar level of success as MALDI-TOF MS, the quality and reproducibility of cell extracts from the same strain grown in different media, various time intervals and analyzed by different models of the same instrument are critical parameters that need validation in the transition to top-down proteomics from the research to the clinical laboratory. Unlike MALDI-TOF MS, protein concentrations need to be predetermined and standardized and up to the present time, available methods for quantifying protein concentrations do not give similar results. For example, newer, simple methods using NanoDrop, which performs efficiently for DNA, does not give comparable results with the standard 40-year-old Bradford assay (1976). A simple, rapid method needs to be incorporated as a baseline for strain comparison

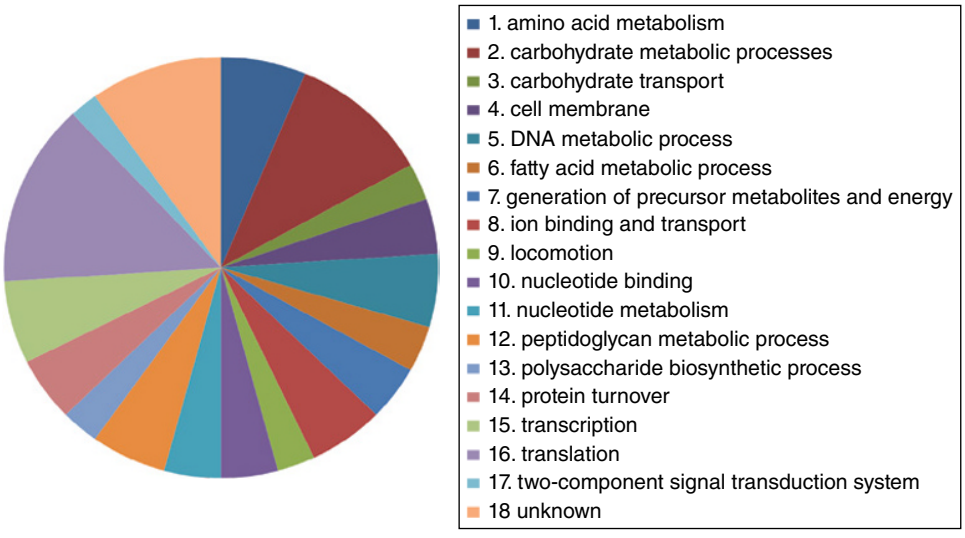


Figure 1.16 Distribution of cellular proteins of a cell extract to assess the efficacy of an extraction method. The figure shows the protein distribution of a standard *E. coli* strain using mechanical breakage and centrifugation steps. Methods such as sonication released mainly ribosomal proteins and yields profiles similar to MALDI-TOF MS but more complex. A method that captures a broad cross-section of protein is, in our view, far more beneficial for a future database.

because the key to success is the quality of the sample and the standardization of cell preparations.

The success of MALDI-TOF MS has been underpinned by the ability to retain the basic platform on which to assemble the database. Thus, although current commercial systems are capable of higher resolution and protein identification in reflectron mode, microbial identification is performed in the linear mode. In our experience, analysis of the proteome of cells cultured on agar plates is more reproducible than broth cultures but in assessing the pathogenic potential of a strain such as *C. difficile* for toxin expression, it may be necessary to culture in both broth and agar plates. The distribution of cellular proteins is markedly influenced by the type of extraction method. Some methods involving harsh detergents can yield very high levels of the more stable ribosomal proteins (>70%), which can mask the cellular proteins (unpublished). We believe that comprehensive representation of the proteome (e.g. in the form of a simple pie chart; Figure 1.16) enables visualization of the sample and is necessary prior to assembly of an MS/MS database. Current work in several laboratories is now enabling this baseline to be established, and it is expected in the near future that MS/MS-based methods will become a reality for strain typing, antibiotic sensitivity profiling and pathotyping in real time.

References

Ahif, D. R., Thomas, P. M. and Kelleher, N. L. (2013). Developing top down proteomics to maximize proteome and sequence coverage from cells and tissues. *Curr. Opin. Chem. Biol.* 17: 787–794.

Al-Shahib, A., Misra, R., Ahmod, N., Fang, M., Shah, H. N. and Gharbia, S. E. (2010). Coherent pipeline for biomarker discovery using mass spectrometry and bioinformatics. *BMC Bioinformatics.* 11: 437.

Askar, M., Faber, M. S., Frank, C., Bernard, H., Gilsdorf, A., Fruth, A., Prager, R., Höhle, M., Suess, T., Wadl, M., Krause, G., Stark, K. and Werber, D. (2011) Update on the ongoing outbreak of haemolytic uraemic syndrome due to Shiga toxin-producing *Escherichia coli* (STEC) serotype O104, Germany, May 2011. *Euro. Surveill.* 16: 19883.

Bishop, C., Arnold, C. and Gharbia, S. E. (2010). Transfer of a traditional serotyping system (Kauffmann-While) onto a MALDI-TOF MS Platform for the rapid tying of *Salmonella* isolates. In H. N. Shah and S. E. Gharbia (Eds.), *Mass Spectrometry for Microbial Proteomics* (pp. 463–496). Wiley, Chichester.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248–254.

Cain, T. C., Lubman, D. M. and Weber, W. J. Jr. (1994). Differentiation of bacteria using protein profiles from matrix assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Comm. Mass Spectro.* 8: 1026–1030.

Chamot-Rooke, J., Mikaty, G., Malosse, C., Soyer, M. Dumont, A., Gault, J., Imhaus, A. F., Martin, P., Trellet, M., Clary, G., Chafey, P., Camoin, L., Nilges, M., Nassif, X. and Duménil, G. (2011) Posttranslational modification of pili upon cell contact triggers N. meningitidis dissemination. *Science.* 331: 778–782.

Chooneea, D., Karlsson, R., Encheva, V., Appleton, H. and Shah, H. N. (2010). Elucidation of the outer membrane proteome of *Salmonella enterica* serovar Typhimurium utilising a lipid-based protein immobilization technique. *BMC Microbiol.* 10: 44.

Claydon, M. A., Davey, S. N., Edwards-Jones, V. and Gordon, D. B. (1996). The rapid identification of intact microorganisms using mass spectrometry. *Nature Biotech.* 14: 1584–1586.

Dekio, I., Culak, R., Misra, R., Gaulton, T., Fang, M., Sakamoto, M., Oshima, K., Hattori, M., Klenk, H.-P., Rajendram, D., Gharbia, S. E. and Shah, H. N. (2015). Dissecting the taxonomic heterogeneity within *Propionibacterium acnes*: Proposal for *Propionibacterium acnes* subsp. *acnes* subsp. nov. and *Propionibacterium acnes* subsp. *elongatum* subsp. nov. *Int. J. Syst. Evol. Microbiol.* 65: 4776–4787.

EDITORIAL (2011). Whole-genome sequencing and crowd sourced analyses proved a powerful adjunct to traditional typing in the recent *Escherichia coli* outbreak. *Nature Biotech.* 29: 769.

Encheva,V., Wait, R., Gharbia, S. E., Begum, S. and Shah, H. N. (2005). Proteome analysis of serovars Typhimurium and Pullorum of *Salmonella enterica* subspecies. *BMC Microbiol.* 5: 42

Encheva, V., Wait, R., Gharbia, S. E., Begum, S. and Shah, H. N. (2006). Comparison of extraction procedures for proteome analysis of *Streptococcus pneumoniae* and a basic reference map. *Proteomics.* 6: 3306–3317.

Encheva, V., Wait, R., Begum, S., Gharbia, S. E. and Shah, H. N. (2007). Protein expression diversity amongst serovars of *Salmonella enterica*. *Microbiology.* 153: 4183–4193.

Evason, D. J., Claydon, M. A. and Gordon, D. B. (2001). Exploring the limits of bacterial identification by intact cell-mass spectrometry. *J. Amer. Soc. Mass Spectro.* 12: 49–54.

Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., an der Heiden, M., Bernard, H., Fruth, A., Prager, R., Spode, A., Wadl, M., Zoufaly, A., Jordan, S., Kemper, M. J., Follin, P.,

Müller, L., King, L. A., Rosner, B., Buchholz, U., Stark, K. and Krause, G. (2011). Epidemic profile of Shiga-toxin–producing *Escherichia coli* O104:H4 outbreak in Germany. *N. Engl. J. Med.* 365: 1771–1780.

Gregorich, Z. R. and Ying, G. (2014). Top-down proteomics in health and disease: Challenges and opportunities. *Proteomics.* 14: 1195–1210.

Gupta, R. S. (2010). Microbial phylogeny and evolution based on protein sequences – the change from targeted genes to proteins. In H. N. Shah and S. E. Gharbia (Eds.), *Mass Spectrometry for Microbial Proteomics* (pp. 35–53). Wiley, Chichester.

Gusev, A. I., Wilkinson, W. R., Proctor, A. and Hercules, D. M. (1995). Improvement of signal reproducibility and matrix/co-matrix effects in MALDI analysis. *Anal. Chem.* 67: 1034–1041.

Holland, R. D., Wilkes, J. G., Rafii, F., Lay, J. O. (1996). Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ ionization with time-of-flight mass spectrometry. *Rapid Comm. Mass Spectro.* 10: 1227–1232.

Honisch, C., Chen, Y. and Hillenkamp, F. (2010). DNA resequencing by MALDI-TOF MS and its application to traditional microbiological problems. In H. N. Shah and S. E. Gharbia (Eds.), *Mass Spectrometry for Microbial Proteomics* (pp. 443–462). Wiley, Chichester.

Honisch, C., Chen, Y., Mortimer, C., Arnold, C., Schmidt, O., van den Boom, D., Cantor, C. R., Shah, H. N. and Gharbia, S. E. (2007). Automated comparative sequence analysis by base-specific cleavage and mass spectrometry for nucleic acid-based microbial typing. *Proc. Natl. Acad. Sci. U S A.* 104: 10649–10654.

Hustoft, H. K., Malerod, H., Wilson, S. R., Reubsaet, L., Lundanes, E. and Greibrokk, T. (2012). A critical review of trypsin digestion for LC-MS based proteomics. In Dr. Hon-Chiu Leung (Ed.), *Integrative Proteomics*. ISBN: 978-953-51-0070-6, InTech.

Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M. and Shah, H. N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect. Genet. Evol.* 4: 221–242.

Krishnamurthy, T. and Ross, P.L. (1996). Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Comm. Mass Spectro.* 10: 883–888.

Mijakovic, I., Grangeasse, C. and Stülke, J. (Eds.). In: Regulatory potential of post-translational modifications in bacteria. Frontiers Media SA. ISBN: 978-2-88919-610-4. Product Name: *Frontiers Research Topic Ebook*. From *1st International Conference on Post-Translational Modifications in Bacteria* (September 9–10, 2014, Göttingen, Germany).

Misra, R. V., Ahmod, M. Z., Parker, P., Fang, M., Shah, H. N. and Gharbia, S. E. (2012). Developing an integrated proteo-genomic approach for the characterisation of biomarkers for the identification of *Bacillus anthracis*. *J. Microbiol. Meth.* 88: 237–247.

Mohanty, G. and Mukherji, S. (2008). Biodegradation rate of diesel range *n*-alkanes by bacterial cultures *Exiguobacterium aurantiacum* and *Burkholderia cepacia*. *Intern. Biodeter. Biodegrad.* 61: 240–250.

Pitt, T. L., Malnick, H., Shah, J., Chattaway, M. A., Keys, C. J., Cooke, F. J. and Shah, H. N. (2007). Characterisation of *Exiguobacterium aurantiacum* isolates from blood cultures of six patients. *Clin. Microbiol. Infect.*13: 946–948.

Rajakaruna, L., Hallas, G., Dare, D., Sutton, H., Encheva, V., Culak, R., Innes, I., Ball, G., Sefton, A. M., Eydmann, M., Kearns, A. M. and Shah, H. N. (2009). High throughput identification of clinical isolates of *Staphylococcus aureus* using MALDI-TOF-MS of intact cells. *Infect. Genet. Evol.* 9: 507–513.

Rajendram, D. (2003). PhD Thesis: 'Dissecting the Diversity of the Genus *Peptostreptococcus* using Genomics and Proteomic Analyses'. (University of East London). Supervisors: H. N. Shah and S. E. Gharbia.

Reynaud af Geijersstam, A., Culak, R., Molenaar, L., Chattaway, M., Røslie, E., Peciuliene, V., Haapasalo, M. and Shah, H. N. (2007). Comparative analysis of virulence determinants and mass spectral profiles of Finnish and Lithuanian endodontic *Enterococcus faecalis* isolates. *Oral Microbiol. Immunol.* 22: 87–94.

Roth, M. J., Parks, B. A., Ferguson, J. T., Boyne, M. T. and Kelleher, N. L. (2008). Proteotyping: Population proteomics of human leukocytes using top down mass spectrometry. *Anal Chem.* 80: 2857–2866.

Rotz, L. D., Khan, A. S., Lillibridge, S. R., Ostroff, S. M. and Hughes, J. M. (2002). Public health assessment of potential biological terrorism agents. *Emerg. Infect. Dis.* 2: 225–230.

Schmid, O., Ball, G., Lancashire, L., Culak, R. and Shah, H. N. (2005). New approaches to identification of bacterial pathogens by surface enhanced laser desorption/ionisation time of flight mass spectrometry in concert with artificial neural networks, with special reference to *Neisseria gonorrhoeae*. *J. Med. Microbiol.* 54: 1205–1211.

Scigelova, M. and Makarov, A. (2006) Orbitrap mass analyzer – overview and applications in proteomics. *Proteomics.* 6 Suppl 2 (S2): 16–21.

Severi, E., Heinsbroek, E., Watson, C. and Catchpole, M. (2012). HPA Olympics Surveillance Work Group. Infectious disease surveillance for the London 2012 Olympic and Paralympic Games. *Euro Surveill.* Aug 2; 17(31): 20232.

Shah, H. N. (2005). MALDI-TOF-Mass Spectrometry: Hypothesis to proof of concept for diagnostic microbiology. *Clin. Lab. Intern.* 29: 35–38.

Shah, H. N., Chilton, C., Rajakaruna, L., Gaulton, T., Hallas, G., Atanassov, H., Khoder, G., Rakowska, P. D., Eleonora Cerasoli, E. and Gharbia, S. E. (2010). Changing concepts in the characterisation of microbes and the influence of mass spectrometry. In H. N. Shah and S. E. Gharbia (Eds.), *Mass Spectrometry for Microbial Proteomics* (pp. 3–34).. Wiley, Chichester.

Shah, H. N. and Collins, M. D. (1980). Fatty acid and isoprenoid quinone composition in the classification of *Bacteroides melaninogenicus* and related taxa. *J. Appl. Bacteriol.* 48: 75–84.

Shah, H. N. and Collins, M. D. (1983). Genus *Bacteroides* a chemotaxonomical perspective. *J. Appl. Bacteriol.* 55: 403–416.

Shah, H. N., Encheva, V., Schmid, O., Nasir, P. et al. (2005). Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS): A potentially powerful tool for rapid characterisation of microorganisms. In M. J. Miller (Ed.), *Encyclopedia of Rapid Microbiological Methods* (Vol. 3, pp. 57–96). DHI Publishing, IL, USA.

Shah, H. N. and Gharbia, S. E. (2010). *Mass Spectrometry for Microbial Proteomics*, H. N. Shah and S. E. Gharbia (Eds.). Wiley, Chichester.

Shah, H. N. and Gharbia, S. E. (2011). A century of systematics of the genus *Bacteroides:* From a single genus up to the 1980s to an explosion of assemblages and the dawn of MALDI-TOF-Mass Spectrometry. *The Bulletin of BISMiS.* 2: 87–106.

Shah, H. N. and Gharbia, S. E. (2012). Using nano-LC-MS/MS to investigate the toxicity of outbreak *E. coli* 0104:h4 strain. *Culture*. 33, ISSN 0965-0989.

Shah, H. N., Keys, C., Gharbia, S. E., Ralphson, K., Trundle, F., Brookhouse, I. and Claydon, M. (2000). The application of MALDI-TOF Mass Spectrometry to profile the surface of intact bacterial cells. *Micro. Ecol. Health. Dis.* 12: 241–246.

Shah, H. N., Keys, C. J., Schmid, O. and Gharbia, S. E. (2002). Matrix-assisted laser desorption/ionisation time of flight mass spectrometry and proteomics; a new era in anaerobic microbiology. *Clin. Infect. Dis.* 35: 58–64.

Shah, H. N., Rajakaruna, L., Ball, G., Misra, R., Al-Shahib, A., Fang, M. and Gharbia, S.E. (2011). Tracing the transition of methicillin resistance in subpopulations of *Staphylococcus aureus* using SELDI-TOF Mass Spectrometry and Artificial Neutral Network Analysis. *System. Appl. Microbiol.* 34: 81–86.

Shah, H. N., van Steenbergen, T. J. M., Hardie, J. M. and de Graaff, J. (1982). DNA base composition, DNA-DNA reassociation and isoelectric focusing of proteins of strains designated *Bacteroides oralis. FEMS Microbiol. Letts.* 13: 125–130.

Shah, H. N., Stephenson, J., Cardasis, H., Neil, J., Yip, P., Ravela, S., Ritamo, I., Damsbo, M., Grist, R., Freeke, J., Stielow, B., Gvozdyak, O., Dukik, K., de Hoog, S., Damoc, E., Valmu, L., Cherkassky, A., Fang, M., Gaulton, T., Misra, R. and Gharbia, S. E. (2015). A Global Diagnostic Approach for Microbial Identification: Accurate characterization of difficult to differentiate pathogens – using Top-down proteomics. *Clinical Microbiology and Infectious Diseases (ECCMID) – 25th European Congress*. Copenhagen, Denmark-April 25–28, 2015.

2

Criteria for Development of MALDI-TOF Mass Spectral Database

Markus Kostrzewa[1] *and Thomas Maier*[2]

[1] *Bruker Daltonik GmbH, Bremen, Germany*
[2] *Bruker Daltonik GmbH, Fahrenheitstrasse 4, 28359, Bremen, Germany*

2.1 Introduction

The database of a MALDI-TOF mass-spectrometry-based microorganism identification system is its key component and clearly the most important element for secure identification and differentiation of organisms. The quality as well as the quantity of database entries largely determine the usefulness of a system in practice, in clinical microbiology but also in other areas. Manufacturers of commercially available systems are offering libraries which cover a broad range of microorganisms or are restricted to a more specific area, that is, dedicated to particular groups of microorganisms such as filamentous fungi or mycobacteria. On the other hand, reference libraries created by users themselves help fill gaps in the systems of manufacturers or enable the coverage of species or systematic groups which are of special interest to them. Databases are further closely related to the preparation method which has been applied to the generation of reference entries or identification rules/biomarkers. Generally, the same method as that used for database generation gives the best identification results for the isolates investigated. Extensive and thoroughly quality-controlled databases have made MALDI-TOF MS the new laboratory standard for microorganism identification.

2.2 Commercially Available Databases

Manufacturers of MALDI-TOF-MS-based identification systems are delivering libraries which are specific for their instrument, preparation technique and identification algorithm. Currently, three main commercial systems from two market-dominating manufacturers are available. Each of these systems has its own database(s) as well as its dedicated strategies for library creation and identification algorithms.

SARAMIS (Spectral ARchive and Microbial Identification System) is working with the so-called SuperSpectra as the main reference entries for identification [1].

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

SuperSpectra are peak lists that are generated from individual peak lists of different strains which belong to a common systematic category, for example, a microbial species. For their generation, mass spectra are acquired from each strain from different cultivation media, and the respective single peak lists are created using the mass spectrometer's processing software. The individual peak lists are compared, and stable, representative peaks are chosen in a manual or semiautomatic manner to build a specific compromise peak list, the SuperSpectrum. Peaks can be weighted according their importance for identification in this process. The match against the database comprising the SuperSpectra of the contained species (or other systematic entities) results in an identification of up to 99.9% probability. If no sufficient match is obtained against the SuperSpectra, a match against all single spectra can help give a hint about the identity of the unknown organism. Originally, SARAMIS was developed by a small German company (AnagnosTec) as an MS-manufacturer-independent software and database which could be combined with any MALDI-TOF MS system. In 2010, the French microbiology company bioMérieux acquired the microbial database for bacterial identification, the related intellectual property and know-how from AnagnosTec. Currently, SARAMIS is the research-use-only MALDI-TOF solution of bioMérieux (Vitek MS RUO), combined with a mass spectrometer manufactured by Shimadzu (Japan).

For diagnostic purposes, bioMérieux, as a part of the Vitek MS IVD, has developed a new dedicated software and database system which is not directly related to the SuperSpectra database or even approach of SARAMIS. The Advanced Spectra Classifier (ASC) system [2] uses a kind of biomarker approach to characterize microbial species. Mass spectra are acquired from multiple strains representing different geographical locations and cultivation conditions. These spectra are divided into bins, where each bin is considered a possible marker. In a huge bioinformatics comparison approach for each bin/biomarker, the significance for a given species is determined: is it apparent in this species only, is it found in this and other species or does it not occur in the respective species but only in others? The identification of an unknown isolate is done on the basis of the biomarker network and given with a probability value (up to 99.9%). It has to be mentioned that this calculation has to be done in a 'closed system', that is, for a given number of species which are set in a 'biomarker network' for differentiation. An extension of a database based on the ASC biomarker network can change relationships inside the network (e.g. a formerly unique biomarker can occur also in a newly added species). But, as the ASC approach is only used in the diagnostic MALDI-TOF MS system of bioMérieux, any recalculation is done (and has to be done) by the manufacturer. The Vitek MS IVD is CE marked, and a version has obtained FDA clearance in 2012. For more information, the reader may visit the homepage of the manufacturer (http://www.biomerieux.com/).

The third identification system, the MALDI Biotyper, comes from the second main manufacturer of MALDI-TOF MS for microbial identification, Bruker (USA). The MALDI Biotyper database and software algorithm are not related to any of the systems described above, but again use their own strategy to identify and differentiate microorganisms based on their MALDI-TOF spectra. Here, independent references for single strains of a given species are created and stored in the database. The reference for a single strain, a main spectrum (MSP), is based on multiple measurements, that is, technical replicates of spectra of the respective strain. Peak lists derived from the individual spectra are used by a proprietary algorithm of the MALDI Biotyper software to generate

the MSP. The MSP contains the 70 most prominent peaks of the spectra collection representing the strain, including information on average peak intensity, peak position and peak occurrence frequency. Identification of an unknown strain is performed by matching the peak list derived from its profile mass spectrum against each reference entry of the database. A score is calculated by multiplication of three factors (each factor can reach a maximum of 1), that is, matches of the unknown spectrum against the MSP, matches of the MSP against the unknown spectrum and similarity of the intensity profile of matching peaks. The logarithm of these calculated scores (multiplied by 1000) gives the log(score), maximum 3 ($\log(1 \times 1 \times 1 \times 1000) = 3$). All log(scores) are ranked from the highest to the lowest. The match with the highest log(score) results in a species-level identification if a threshold is exceeded (log(score) ≥2 for high confidence, log(score) ≥1.7 for low confidence).

The MALDI Biotyper is offered in two diagnostic versions, (1) the IVD MALDI Biotyper for the European market and other countries where IVD-CE labelling is applicable and (2) the MALDI Biotyper CA in the United States. As diagnostic systems, both do not allow database extension or library addition by the user and are restricted to their validated workflow. In contrast, the MALDI Biotyper RUO, a research-use-only system, allows extension of the libraries available from the manufacturer as well as the creation of custom user-defined libraries. The RUO system uses the same identification engine as the MALDI Biotyper IVD, but the parameters can be modified by the user for special purposes.

2.3 Establishment of User-Defined Databases

A particular strength of MALDI-TOF technology is that the user can build his or her own databases. Such databases can consist of special groups of organisms of particular importance as described for filamentous fungi [3–13], mycobacteria [14–19], and anaerobes [20–23], to more rare taxonomic groups, for example, *Prototheca* [24], or even to higher evolved or "exotic" systematic groups such as, for example, insects [25–31], crustaceans [32], and others. Also, databases for more dedicated application areas, for example, for beer spoiling organisms in breweries [33–40], can easily be established. Such user-generated databases might be deployed stand-alone for identification purposes, or they might be used to supplement a database supplied by the manufacturer to improve the sensitivity and/or specificity of identification results.

At this point, it has to be mentioned that the authors of this chapter are specialists in the MALDI Biotyper system. Therefore, this chapter is focused on the criteria which have been found to be important for the creation of high-quality databases for this particular system. Nevertheless, most of the following statements and descriptions can be generalized and also will be important for any MALDI-TOF-MS-based reference database system which is being deployed for microorganism identification. The basis for MALDI-TOF MS reference databases always is inclusion of well-controlled organisms and thorough inspection of mass spectrometry conditions at certain checkpoints.

To guarantee a high quality of reference entries, some important rules have to be followed in all parts of the database creation process, that is, reference identification of included strains, sample preparation, MALDI-TOF measurement, reference spectra calculation, library entry quality control, and so on.

For MALDI reference library generation, four important phases should be considered. For each phase of library creation, control measures as well as standard operation procedures (SOPs) should be established to ensure their reliability.

Pre-analytical phase
- Correct taxonomical species identification

Sample preparation
- Microorganism cultivation/growth conditions
- MALDI target plate preparation (direct transfer/extended direct transfer/extraction)
- MALDI matrix and solvent composition
- Crystallization conditions (temperature, humidity)

MALDI-TOF MS measurement
- Instrument settings (mass range, detector voltage, and so on)
- Calibration
- Settings for automatic spectra acquisition

Data analysis and library calculation
- Data processing (spectra quality assessment and reference generation)
- Strain selection based on 'MALDI diversity'
- Reliability check

Details on how to proceed in these phases will be described in the following paragraphs.

2.4 Species Identification/Control of Reference Strains to Be Included into a Database

Prior to the database creation, the correct taxonomic identification of all isolates that are intended to be included is mandatory. It is preferable to use a set of methods (morphology, biochemical/physiological tests, gene sequencing, etc.) to identify a strain 'polyphasically' and unambiguously prior to its introduction as a reference entry. Although molecular biology, that is, 16S rRNA gene sequencing, is becoming acknowledged as the gold standard for microorganism identification, there is no "one-fits-all" method which can be applied alone for all species. For many bacteria, 16S rRNA gene sequence analysis is very useful to determine the correct species. But in many cases, closely related species have to be differentiated by at least one further method, preferably the sequences of a protein encoding gene or certain biochemical characteristics or serology.

A database designed for, for example, medical diagnostics should contain, in addition to the spectra of representatives of infection outbreaks, also those of type strains as 'taxonomic marker entries'. As mass spectra contain general species information but may also exhibit strain-specific markers, a species should be represented at best by using strains representing MALDI diversity in the species. Geographical or host origins are usually less important if different geographical variants show a similar MALDI pattern.

As a good starting point for genetic species delineation, the CLSI MM-18a rules [41] can be used to define a certain species based on their 16S rRNA gene sequence. Clear

thresholds on how to interpret 16S rRNA gene sequences and recommendations on which further protein gene sequences can be used – for example, if 16S rRNA gene sequencing does not sufficiently differentiate closely related species – are described in these rules. If after applying these rules, species identity remains uncertain, further tests (morphology, biochemical tests or DNA sequence alignments) have to be deployed to improve reliability.

Although a MALDI-TOF MS reference library may be designed for the same purpose (identification of bacteria) as a 16S rRNA gene library, MALDI-TOF mass spectra are exposed to additional factors of influence compared to 16S rRNA gene sequences. Sequencing of DNA is more or less 'digital' (i.e. it was either successful or unsuccessful for each data point), and the sequencing result is independent of cultivation or other biological influences. MALDI spectra may look 'near-to-perfect', but if calibration, cultivation or sample preparation have been not adequate, the spectra might be not suitable for reference generation.

2.5 Sample Preparation

2.5.1 Microorganism Cultivation

Many microorganisms can be cultivated on quite different cultivation media. In clinical routine laboratories, they are grown commonly on complex media as well as on selective media. Complex media are composed to offer a broad range of nutrients to guarantee effective growth of a broad range of microorganisms. On complex media, microorganisms commonly grow fast and show high reproduction rate with high protein biosynthesis (high concentration of highly abundant ribosomal proteins). Because human specimens are not always derived from originally sterile samples (such as blood or CSF) but from naturally microorganism-populated samples such as stool or skin, complex media are not always applicable. Selective media are commonly designed to prefer the growth of selected species or groups, or the sought-after species or groups show a specific property (e.g. a specific colour). Growth on such media generally is stressful for microorganisms, even for the targeted species. The protein pattern can be slightly changed due to adapted stress physiology, which commonly does not influence MALDI-based identification purposes. For MALDI-TOF MS database reference entry generation, it is recommended to choose cultivation media which are optimal for the growth of selected species (e.g. complex media). Also, the media intended to be used in the later application should be considered. Cultivation generally shows different growth phases such as lag phase, log phase or death phase. The most suitable growth phase for references is the log phase because bacteria show in this phase the highest growth rates. In general, fast-growing species are optimally cultivated overnight, but slow-growing species should be grown until sufficient biological material is visible on agar plates (up to several days or even weeks).

In conclusion, to ensure broad applicability of a MALDI-TOF MS reference database, it is generally recommended to use optimal growth conditions to cultivate reference strains to secure optimal MALDI spectra generation.

2.5.2 MALDI Sample Preparation

The most popular and most widespread sample preparation technique for identification of microorganisms is the direct application of low amounts of biological material to the MALDI target plate. A single bacterial colony is smeared on the MALDI target and either directly overlaid with matrix ('direct transfer') or by 1 µL formic acid which is applied prior to matrix application to improve cell lysis ('extended direct transfer'). After short drying, the samples are ready for MALDI measurement.

In addition to these very fast direct methods, short extraction protocols exist if at all spectra acquisition after such a direct transfer of biological material delivered ambiguous results or failed. This can be observed if strong capsules or cell walls are covering/protecting the organisms (e.g. fungi or mycobacteria) or slimy/mucoid substances interfere with the MALDI process. For such situations, extraction protocols exist where the organisms are subjected to short washing and/or chemical extraction steps. During the extraction, several additional factors contribute to an improvement of MALDI-TOF MS spectra. Components such as salts, carbohydrates or peptides can be effectively removed. In addition, more biological material can be used, and higher acid concentrations can be applied for more effective destruction of cell walls and protein release. Due to careful suspending in extraction fluids, every cell can be reached by the extraction solution, and clumpy pieces can be broken. The interaction time with the solvents can be prolonged and carefully controlled, too. Finally, a homogeneous cell-free extract is generated and can be transferred to the MALDI target. Such a procedure guarantees a higher degree of standardization and therefore a higher reproducibility of mass spectra.

Thus, a short extraction procedure is generally the preferred sample preparation technique for reference entry generation (note: this only applies to the MALDI Biotyper system). Because spectra derived from direct transferred samples are generally very similar or largely contained in extraction spectra, the identification is reliable if DT samples are compared against extracted references.

Although simple sample preparation techniques can be used for most microorganisms, some groups have specific properties hampering successful MALDI measurements. As examples, for mycobacteria as well as filamentous fungi, specific protocols should be applied to improve sample preparation and guarantee reproducible, high-quality biological material for the measurement [3–14,17–19,42]. Filamentous fungi show several growth and germination phases which lead to different MALDI patterns for each phase. Therefore, it is advantageous to use a standardized growth procedure to guarantee similar spectra for strains of a species. Growth in liquid medium has been described as facilitating this, in particular for species which exhibit fast sporulation and when direct mycelium analysis from solid medium fails [13].

Highly pathogenic agents such as *Francisella tularensis* or *Burkholderia mallei* can be measured in the same way as other organisms, but secure inactivation [43–51] has to be shown before any transport out of the high-level security laboratory and introduction into the mass spectrometer.

Several MALDI matrices can be used for microorganism measurement. The most popular matrix is alpha-cyano-hydroxy-cinnamic acid (HCCA). This matrix shows a homogeneous crystallization (important in particular for *fast* MALDI measurement) as well as a good ionization capacity in the commonly used mass range (2,000 to

20,000 m/z). The function of the organic solvent of the matrix solution is to guarantee a good cell lysis and protein release as well as to prove a certain 'crystallization behaviour'. Because an optimal matrix/analyte co-crystallization is mandatory, it is necessary to ensure an adequate duration for this crystallization process. If the matrix crystallizes too fast (e.g. because of too high a temperature), the required co-crystallization does not occur; and if the time for the process is too long (e.g. because of a very low temperature), large matrix crystals can be observed which lead to prolonged measurement time (the laser has to find 'sweet spots' to generate adequate mass spectra). In the typically used mass range of 2,000 to 20,000 m/z, many ribosomal proteins appear. The high concentration of ribosomal proteins and their excellent ionization efficiency lead to the good reproducibility and reliability of MALDI microorganism identification although the described effects can be observed. But for database generation, no negative effects should be tolerated.

2.6 MALDI-TOF MS Measurement

Standard settings of a MALDI-TOF mass spectrometer in the appropriate mass range (e.g. 2,000 to 20,000 m/z) based on the manufacturer's specifications are commonly optimal for identification approaches. No further optimization by users is necessary because IVD-CE-labelled products even inhibit any user interaction.

A 'mass spectrum' covers a certain mass range and contains a certain number of peaks. A 'peak' represents a cell component (i.e. protein, peptide, lipid, …, molecule) displayed by its mass per charge ratio (m/z; its value on the X-axis), its intensity (its value on the Y-axis) and its resolution (i.e. the peak width). The MALDI-TOF MS spectra acquisition depends on several technical instrument settings such as laser power, acceleration voltage, detector voltage, calibration constants or delayed ion extraction.

For measurement of reference entries, optimization and/or control of the MALDI measurement process is strongly recommended. The ability of users to analyze and quality-control the optimal instrument settings should be a precondition for optimal reference generation. An 'optimal mass spectrum' means the presence of many peaks (up to ~150 peaks are possible), high signal-to-noise (S/N) ratios (i.e. low noise and peaks with high intensities), high resolution of peaks over the whole mass range and high mass accuracy. For users who intend to build their own databases, practical training by the respective system manufacturer is advised and cannot be replaced by this article.

The achievable quality of a mass spectrum strongly depends on the taxonomic group of the organisms studied. For example, it is much easier to obtain 'high-quality' mass spectra from fast-growing gram-negative rods than from slow-growing mycobacteria. Nevertheless, it is possible to create mass spectra of good quality from nearly any kind of organism. This has led to the proposal that MALDI-TOF MS might be a "universal identifier" [52]. It is always necessary to achieve the optimum for a particular group of organisms.

To assess the optimal instrument setting, external standards can be used. The Bruker Bacterial Test Standard is such a standard, a bacterial extract produced under stringent quality control and supplemented with a concentration of proteins optimal for checking

the instrument settings. The choice of bacterial extract as the calibrant satisfies the rule that calibration should be performed with a substance which represents the analyte to be measured. The protein concentration of this standard is close to the detection limit and contains substances covering the whole mass range of bacterial MALDI-TOF mass spectra with true bacterial proteins and two additionally spiked proteins. Alternatively, microorganisms with known calibrator masses or non-bacterial proteins can be applied.

2.7 Quality Control during Creation and after Establishment of Reference Libraries

For the establishment or extension of a MALDI-TOF MS fingerprint reference database, it should be clear that the purpose of the acquired mass spectra requires but also enables quality control steps at various times of the process which are different from those applied for routine measurement for microorganism identification. These quality control steps are more extensive and thorough than the automated quality control by the instrument and software during the simple identification measurement. The following table is intended to depict some of the prominent differences between both processes.

The quality of mass spectra is a key point in establishing a database and follows the general principle that if poor data is used, spurious results will be obtained. Therefore, careful standardization and control of the sample preparation, mass spectra acquisition and post-measurement control of acquired spectra is a key factor for the creation of a valuable database.

A checklist of factors (see Table 2.1) which influence the MALDI-TOF measurement and therefore have to be tightly controlled or standardized follows.

2.8 Common Influencing Factors for MALDI-TOF MS

Standardize and control:

Target preparation (quality, thin layer, dried droplet, etc.)
Matrix and solvent
Analyte concentration and analyte:matrix ratio
Crystallization conditions (temperature, humidity)
Instrument parameters (ion source, detector)
Laser energy (see Figure 2.1)
Instrument calibration
Data processing (baseline correction, peak picking, smoothing)

2.8.1 Influencing Factors, Specifically Weighted for MALDI Biotyper

Cultivation conditions (low influence)
Age (low influence)
Vegetative cells/spores
Direct transfer/extended direct transfer/extraction (sample dependent)

A list of recommendations by the authors follows. This is not meant to be mandatory nor complete but might be used as a guideline.

Table 2.1 Different requirements of MALDI-TOF MS measurements and QC for routine identification and database establishment.

Routine ID	Library construction
• Fast measurement with sufficient spectral quality • No (or few) interaction(s) with acquisition software • Fully automated data processing • Mainly automated quality control at data acquisition and interpretation	• High spectral quality • Quality > speed • More interaction with acquisition software possible • Semi-automated data processing • Automated, but also significant manual/visual quality control at data acquisition and post-measurement interpretation

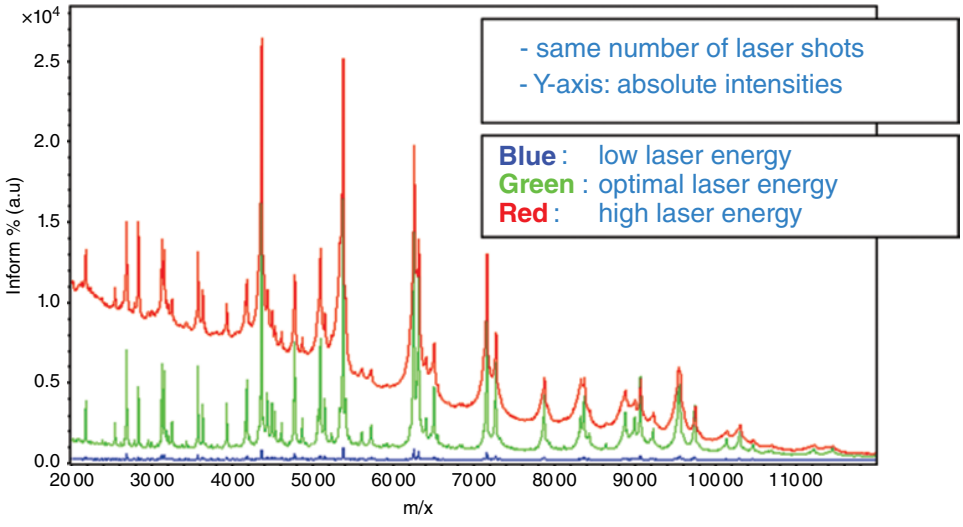


Figure 2.1 Example of spectra acquired with different laser energies. Too low laser energy leads to few peaks with low intensity, and too high energy causes a lift of baseline and broadening of peaks. Broadening of peaks can lead to fusion of closely located signals.

2.8.2 Selection of Strains

- Check: Is identification secure, documented?
- Check: Is influence of growth conditions expected? (Not frequent but, e.g. sporulation)

Generally: let the microorganisms grow in good condition, harvest as fresh, viable cells

2.8.3 Sample Preparation for Measurement

- Use appropriate sample prep protocol (e.g. standard extraction).
- Use a fresh matrix and solvents.
- Always include a Bruker Bacterial Test Standard (BTS) at one position.

- Prepare multiple spots per sample, and if possible multiple measurements per spot too (e.g. 24 spectra per sample; at least 20 have to pass QC later) have to be performed.
- Control and document spotting.

2.8.4 Mass Spectrometry Measurement

- Use a *well-'tuned' and calibrated* instrument.
- Take care that your acquisition/processing parameters are the same as for subsequent identification procedures.

Finally, the already acquired mass spectra as well as the library that has been calculated based on those should be controlled again. A proven control procedure for the mass spectra in a team dedicated to reference library construction is spectra control by two different specialists; that is, the person responsible for the measurement performs a first check, and the second analysis (as a final quality gate) is done by a different specialist. Standards and requirements for the mass spectra may differ depending on the intended library.

2.8.5 Spectra Analysis/Quality Control

- For MALDI Biotyper: Check BTS; Identified as *E. coli* with high log (score) (e.g. >2.2)? Spectral quality?
- Investigate and compare mass spectra using adequate mass spectrometry software, for example, FlexAnalysis for the MALDI Biotyper (intensity, resolution, number of peaks).

On the basis of the acquired mass spectra, the reference entries/library have to be calculated depending on and in accordance with the MALDI-TOF system used. For the MALDI Biotyper, this means calculation of the so-called main spectra (MSP). In many cases, the standard setting given by the manufacturer should be applied, but this depends on the intended purpose of the database. Finally, the established library should be controlled for agreement with and discrepancies from the contained references. This can be done in different, complementary ways, also depending on the mass spectrometry system which is used. One kind of control which is specific for the MALDI Biotyper is calculation of 'local MSP dendrograms'. For this purpose, new reference entries are included in a similarity dendrogram calculation which contains established and confirmed database entries, some of them more closely related, others more distinct. Figure 2.2 shows an example where a wrong pre-identification of a strain could be detected with this simple control step. As such a procedure is a push-button process in the MALDI Biotyper software, it should be mandatory in quality control for each database creation and extension. Discrepancies detected in this control step have to be further investigated and resolved by other methods, subsequently.

A second very valuable control step is the comparison of newly established database entries with as many as possible existing references, again to check for congruence with the existing data or discrepancies. For the MALDI Biotyper, the new MSPs can simply be loaded in the software and matched against each available database entry. This can be facilitated using the 'Explorer' module for so-called 'offline analysis'; details can be found in the manual supplied by the manufacturer. Using the standard database that

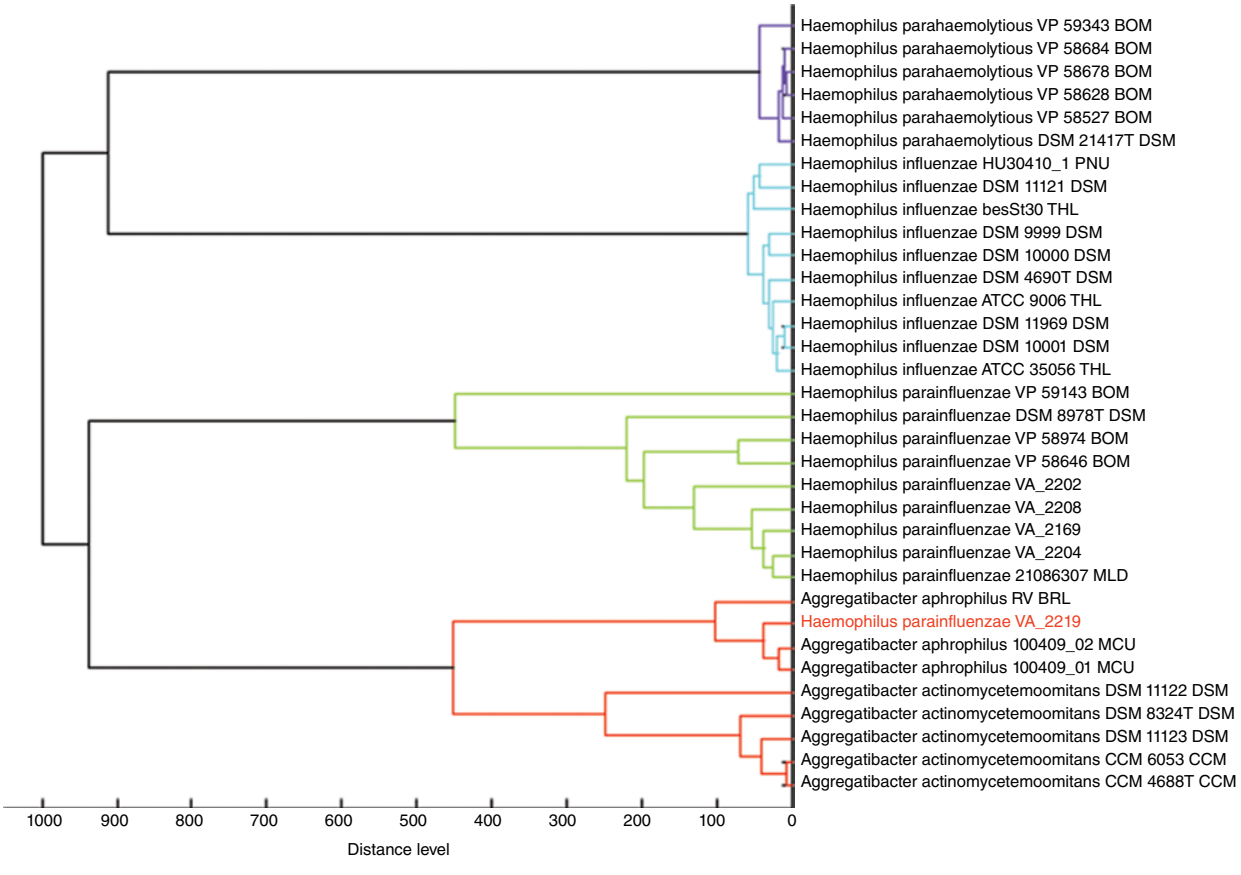


Figure 2.2 Dendrogram of a novel MSP (strain VA2219, depicted in red) with old database references; wrong pre-identification has been detected by cluster analysis. The strain could be confirmed as *Aggregatibacter aphrophilus*, subsequently.

accompanies the system, currently this enables comparison with about 6000 database entries representing more than 2300 species. As an example, a new database entry which is created from a species not known to be related to any of the species covered by the existing database should not result in a close match with the existing MSPs. Further, this step as well as the dendrogram calculation may give a hint regarding a lack of differentiation of closely related species. If so, this has to be further investigated by more extensive tests (e.g. validation with a larger set of well-characterized strains) and resolved or – if resolution of the problem is not possible – it may point to the incapability of the method to distinguish these species.

A third step to control a newly established database, or a database extension can be the control for spectra acquired from strains not contained in the database. For repeated database extensions, it makes sense to use an established set of such spectra, for example, from a large validation study, as a standard. For the MALDI Biotyper, again this can be accomplished using the Explorer module of the program. Spectra from new species in a database can be controlled further by measurement and matching of new, database-independent strains (if available).

As a summary, quality control is possible and should be performed for MALDI-TOF MS profile libraries at various stages in the process of database entry generation. Research databases may be controlled with less effort, and the more a database will be used in a routine setting, the more control is recommended. A proven principle for increased quality assurance is the 'principle of four eyes': two people perform an independent control for each step.

2.8.6 MSP Creation and Analysis/Quality Control

- Create MSP using the appropriate parameters (e.g. Bruker standard).
- Create dendrograms – check if reasonable (e.g. grouping of closely related/unrelated organisms).
- Match against the Bruker database and/or your own library – check if reasonable (no match with unrelated organism, match with closely related organisms).
- Match with quality control spectra set which was not part of the library creation.

2.9 User-Created and Shared Databases: Examples and Benefits

It is not the intention of this chapter to promote the distribution of databases which have been created by users of a system. One should be aware that the person or institution creating such a MALDI-TOF MS fingerprint database is fully responsible for its quality and applicability. As described above, careful quality control during the establishment of a library is necessary, but this is an elaborate process. Databases intended and validated for diagnostic purposes are 'closed' by manufacturers. They cannot be extended or modified by design. On the other hand, in particular for research purposes but also for routine applications which are not in the clinical diagnostic field, it may make sense for a user to creates his or her own dedicated database or to extend a database delivered by the manufacturer to close gaps in species coverage. As an example, strain collections can create references for stored strains which in a later quality control

process can help to secure strain identity. For example, this has been the practice at the German strain collection in Braunschweig (DSMZ) and the strain collection of the Institute Pasteur in Paris (CIP) for years (personal communication Dr. Schumann, DSMZ, and Dr. Bizet, CIP, respectively). Further, a consortium working on the exploration of bacteria from the genus *Vibrio* has developed a library which complements the MALDI Biotyper database with many strains from their research work [53]. This library is also offered to other users for their research purposes. There are more examples, and most are not published; therefore, it is not intended to reference a broad library collection here. But it should be mentioned also that the manufacturers can learn from the work which is done in this field, and obtain examples of possible database improvement. Manufacturers should observe such research and carefully investigate if there are suggestions for improvement even for their own diagnostic databases. Although MALDI-TOF MS fingerprinting already can be considered the laboratory standard for microorganism identification in clinical microbiology laboratories, there still are opportunities for improvement which should be exploited to benefit users and patients.

References

1 Welker, M. and Moore, E. R. B. (2011). Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst. Appl. Microbiol.*, **34** (1), 2–11.

2 Rychert, J., Burnham, C.-A. D., Bythrow, M. *et al.* (2013). Multicenter evaluation of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system for identification of gram-positive aerobic bacteria. *J. Clin. Microbiol.*, **51** (7), 2225–2231.

3 Becker, P. T., de Bel, A., Martiny, D. *et al.* (2014). Identification of filamentous fungi isolates by MALDI-TOF mass spectrometry: Clinical evaluation of an extended reference spectra library. *Med. Mycol.*, **52** (8), 826–834.

4 Becker, P. T., Stubbe, D., Claessens, J. *et al.* (2015) Quality control in culture collections: Confirming identity of filamentous fungi by MALDI-TOF MS. *Mycoscience*, **56** (3), 273–279.

5 Cassagne, C., Ranque, S., Normand, A.-C. *et al.* (2011). Mould routine identification in the clinical laboratory by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *PLoS ONE*, **6** (12), e28425.

6 Chalupová, J., Raus, M., Sedlárová, M. and Sebela, M. (2013). Identification of fungal microorganisms by MALDI-TOF mass spectrometry. *Biotechnol. Adv.*, **51** (3), 828–834.

7 Gautier, M., Ranque, S., Normand, A.-C. *et al.* (2014). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: Revolutionizing clinical laboratory diagnosis of mould infections. *Clin. Microbiol. Infect.*, **20** (12), 1366–1371.

8 Lau, A. F., Drake, S. K., Calhoun, L. B. *et al.* (2012). Development of a clinically comprehensive database and simple procedure for the identification of molds from solid media by matrix-assisted laser desorption/ionization time of flight mass spectrometry. *J. Clin. Microbiol.*, **51** (3), 828–834.

9 Normand, A.-C., Cassagne, C., Ranque, S. *et al.* (2013). Assessment of various parameters to improve MALDI-TOF MS reference spectra libraries constructed for the routine identification of filamentous fungi. *BMC Microbiol.*, **13**, 76.

10 Panda, A., Ghosh, A. K., Mirdha, B. R. *et al.* (2015). MALDI-TOF mass spectrometry for rapid identification of clinical fungal isolates based on ribosomal protein biomarkers. *J. Microbiol. Methods*, **109**, 93–105.
11 Posteraro, B., De Carolis, E., Vella, A. and Sanguinetti, M. (2013). MALDI-TOF mass spectrometry in the clinical mycology laboratory: identification of fungi and beyond. *Expert Rev. Proteomics*, **10** (2), 151–164.
12 Ranque, S., Normand, A.-C., Cassagne, C. *et al.* (2014). MALDI-TOF mass spectrometry identification of filamentous fungi in the clinical laboratory. *Mycoses*, **57** (3), 135–140.
13 Schrödl, W., Heydel, T., Schwartze, V. U. *et al.* (2012). Direct analysis and identification of pathogenic *Lichtheimia* species by matrix-assisted laser desorption ionization–time of flight analyzer-mediated mass spectrometry. *J. Clin. Microbiol.*, **50** (2), 419–427.
14 Buckwalter, S. P., Olson, S. L., Connelly, B. J. *et al.* (2015). Evaluation of MALDI-TOF mass spectrometry for the identification of *Mycobacterium* species, *Nocardia* species and other Aerobic Actinomycetes. *J. Clin. Microbiol.*, JCM.02128–15.
15 El Khéchine, A., Couderc, C., Flaudrops, C. *et al.* (2011). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of Mycobacteria in routine clinical practice. *PLoS ONE*, **6**, e24720.
16 Lotz, A., Ferroni, A., Beretti, J.-L. *et al.* (2010). Rapid identification of Mycobacterial whole cells in solid and liquid culture media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.*, **48** (12), 4481–4486.
17 Mather, C. A., Rivera, S. F. and Butler-Wu, S. M. (2013). Comparison of the Bruker Biotyper and VITEK MS matrix assisted laser desorption ionization time-of-flight mass spectrometry systems for the identification of Mycobacteria using simplified protein extraction protocols. *J. Clin. Microbiol.*, **52**(1), 130–138.
18. Saleeb, P. G., Drake, S. K., Murray, P. R. and Zelazny, A. M. (2011). Identification of mycobacteria in solid-culture media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.*, **49** (5), 1790–1794.
19 Seng, P., Abat, C., Rolain, J. M. *et al.* (2013). Identification of rare pathogenic bacteria in a clinical microbiology laboratory: Impact of MALDI-TOF mass spectrometry. *J. Clin. Microbiol.*, **51** (7), 2182–2194
20 Wybo, I., Soetens, O., De Bel, A. *et al.* (2012). Species identification of clinical *Prevotella* isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.*, **50** (4), 1415–1418.
21. Barreau, M., Pagnier, I. and La Scola, B. (2013). Improving the identification of anaerobes in the clinical microbiology laboratory through MALDI-TOF mass spectrometry. *Anaerobe*, **22**, 123–125.
22 Fedorko, D., Drake, S., Stock, F. and Murray, P. (2012). Identification of clinical isolates of anaerobic bacteria using matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Eur. J. Clin. Microbiol. Infect. Dis.*, **31** (9), 2257–2262.
23 Coltella, L., Mancinelli, L., Onori, M. *et al.* (2013). Advancement in the routine identification of anaerobic bacteria by MALDI-TOF mass spectrometry. *Eur. J. Clin. Microbiol. Infect. Dis.*, **32** (9), 1183–1192.
24 Murugaiyan, J., Ahrholdt, J., Kowbel, V. and Roesler, U. (2011). Establishment of a matrix-assisted laser desorption ionization time-of-flight mass spectrometry database for rapid identification of infectious achlorophyllous green micro-algae of the genus

Prototheca. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.*, **18** (5), 461–467.

25 Karger, A., Kampen, H., Bettin, B. *et al.* (2012). Species determination and characterization of developmental stages of ticks by whole-animal matrix-assisted laser desorption/ionization mass spectrometry. *Ticks Tick-Borne Dis.*, **3** (2), 78–89.

26 Kaufmann, C., Schaffner, F., Ziegler, D. *et al.* (2011). Identification of field-caught Culicoides biting midges using matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Parasitology*, **139** (02), 248–258.

27 Yssouf, A., Flaudrops, C., Drali, R. *et al.* (2013). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of tick vectors. *J. Clin. Microbiol.*, **51** (2), 522–528.

28 Yssouf, A., Socolovschi, C., Flaudrops, C. *et al.* (2013). Matrix-assisted laser desorption ionization – time of flight mass spectrometry: An emerging tool for the rapid identification of mosquito vectors. *PLoS ONE*, **8** (8), e72380.

29 Yssouf, A., Socolovschi, C., Leulmi, H. *et al.* (2014). Identification of flea species using MALDI-TOF/MS. *Comp. Immunol. Microbiol. Infect. Dis.*, **37** (3), 153–157.

30 Mathis, A., Depaquit, J., Dvořák, V. *et al.* (2015). Identification of phlebotomine sand flies using one MALDI-TOF MS reference database and two mass spectrometer systems. *Parasit. Vectors*, **8** (1), 266–276.

31 Schaffner, F., Kaufmann, C., Pflüger, V. and Mathis, A. (2014). Rapid protein profiling facilitates surveillance of invasive mosquito species. *Parasit. Vectors*, **7** (1), 142.

32 Laakmann, S., Gerdts, G., Erler, R. *et al.* (2013). Comparison of molecular species identification for North Sea calanoid copepods (Crustacea) using proteome fingerprints and DNA sequences. *Mol. Ecol. Resour.*, **13** (5), 862–876.

33 Boehme, K., Fernández-No, I. C., Pazos, M. *et al.* (2013) Identification and classification of seafood-borne pathogenic and spoilage bacteria: 16S rRNA sequencing vs. MALDI-TOF MS fingerprinting. *Electrophoresis*, **34**, 877–887.

34 Kern, C. C., Vogel, R. F. and Behr, J. (2014). Differentiation of Lactobacillus brevis strains using matrix-assisted-laser-desorption-ionization-time-of-flight mass spectrometry with respect to their beer spoilage potential. *Food Microbiol.*, **40**, 18–24.

35 Nicolaou, N., Xu, Y. and Goodacre, R. (2012). Detection and quantification of bacterial spoilage in milk and pork meat using MALDI-TOF-MS and multivariate analysis. *Anal. Chem.*, **84** (14), 5951–5958.

36 Schurr, B.C., Behr, J. and Vogel, R. F. (2015). Detection of acid and hop shock induced responses in beer spoiling Lactobacillus brevis by MALDI-TOF MS. *Food Microbiol.*, **46**, 501–506.

37 Šedo, O., Márová, I. and Zdráhal, Z. (2012). Beer fingerprinting by matrix-assisted laser desorption-ionisation-time of flight mass spectrometry. *Food Chem.*, **135** (2), 473–478.

38 Turvey, M. E., Weiland, F., Meneses, J. *et al.* (2016). Identification of beer spoilage microorganisms using the MALDI Biotyper platform. *Appl. Microbiol. Biotechnol.*, **100** (6), 2761–2773.

39 Usbeck, J. C., Kern, C. C., Vogel, R. F. and Behr, J. (2013). Optimization of experimental and modelling parameters for the differentiation of beverage spoiling yeasts by matrix-assisted-laser-desorption/ionization–time-of-flight mass spectrometry (MALDI–TOF MS) in response to varying growth conditions. *Food Microbiol.*, **36** (2), 379–387.

40 Wieme, A. D., Spitaels, F., Aerts, M. *et al.* (2014). Identification of beer-spoilage bacteria using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Int. J. Food Microbiol.*, **185**, 41–50.

41 Wilson, J. A. and Clinical and Laboratory Standards Institute (2008). *Verification and Validation of Multiplex Nucleic Acid Assays: Approved Guideline*, Clinical and Laboratory Standards Institute, Wayne, PA.

42 Buchan, B. W., Riebe, K. M., Timke, M. *et al.* (2014). Comparison of MALDI-TOF MS with HPLC and nucleic acid sequencing for the identification of Mycobacterium species in cultures using solid medium and broth. *Am. J. Clin. Pathol.*, **141** (1), 25–34.

43 Cunningham, S. A. and Patel, R. (2015). Standard MALDI-TOF MS reagents inactivate potentially hazardous bacteria. *J. Clin. Microbiol.*, JCM.00957–15.

44 Drevinek, M., Dresler, J., Klimentova, J. *et al.* (2012) Evaluation of sample preparation methods for MALDI-TOF MS identification of highly dangerous bacteria. *Lett. Appl. Microbiol.*, **55** (1), 40–46.

45 Karger, A., Stock, R., Ziller, M. *et al.* (2012). Rapid identification of *Burkholderia mallei* and *Burkholderia pseudomallei* by intact cell matrix-assisted laser desorption/ionisation mass spectrometric typing. *BMC Microbiol.*, **12**, 229.

46 Lasch, P., Drevinek, M., Nattermann, H. *et al.* (2010). Characterization of Yersinia using MALDI-TOF mass spectrometry and chemometrics. *Anal. Chem.*, **82**, 8464–8475.

47 Lasch, P., Nattermann, H., Erhard, M. *et al.* (2008). MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. *Anal. Chem.*, **80** (6), 2026–2034.

48 Lasch, P., Wahab, T., Weil, S. *et al.* (2015). Identification of highly pathogenic microorganisms using MALDI-TOF mass spectrometry – results of an inter-laboratory ring trial. *J. Clin. Microbiol.*, JCM.00813–15.

49 Tracz, D. M., Antonation, K. and Corbett, C. R. (2015). Verification of a MALDI-TOF mass spectrometry method for diagnostic identification of high-consequence bacterial pathogens. *J. Clin. Microbiol.*, JCM.02709–15.

50 Tracz, D. M., McCorrister, S. J., Westmacott, G. R. and Corbett, C. R. (2012). Effect of gamma radiation on the identification of bacterial pathogens by MALDI-TOF MS. *J. Microbiol. Methods*, **92** (2), 132–134.

51 Weller, S. A., Stokes, M. G. and Lukaszewski, R. A. (2015). Observations on the inactivation efficacy of a MALDI-TOF MS chemical extraction method on *Bacillus anthracis* vegetative cells and spores. *PLoS ONE*, **10** (12), e0143870.

52 Dridi, B., Raoult, D. and Drancourt, M. (2012). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of Archaea: Towards the universal identification of living organisms. *APMIS Acta Pathol. Microbiol. Immunol. Scand.*, **120** (2), 85–91.

53 Erler, R., Wichels, A., Heinemeyer, E.-A. *et al.* (2014). VibrioBase: A MALDI-TOF MS database for fast identification of Vibrio spp. that are potentially pathogenic in humans. *Syst. Appl. Microbiol.* **38**, 16–25.

3

Applications of MALDI-TOF Mass Spectrometry in Clinical Diagnostic Microbiology

Onya Opota, Guy Prod'hom and Gilbert Greub

Institute of Microbiology, University Hospital Center, and University of Lausanne, Lausanne, Switzerland

3.1 Introduction

Laboratory results account for a large part in physicians' management of patients. Although the majority of the laboratory results are available the same day of hospital admission, microbiology analysis may be associated with longer time-to-results. A suspicion of infection is generally based on clinical symptoms and signs, some of them highly unspecific [1,2]. Indeed, a similar clinical presentation can be caused by different etiologies and different microorganisms. As a consequence, empirical treatments comprising broad-range antibiotics are often started [3,4]. In bacteriology and mycology, a large part of the diagnostic is culture-based, a sensitive approach (when the sampling is achieved before the introduction of antibiotics) that is semiquantitative or quantitative and provides a pure isolate for species identification and antibiotic susceptibility testing (AST), which will allow adapting the antibiotic treatment if needed. The main disadvantage of culture-based approaches is the time to positivity, which varies according to the microorganism growth. When a culture is positive, the identification of the incriminated microorganism may rely on (1) colony morphology and microscopic appearance determined by Gram staining or other specific staining, and (2) biochemical identification; some rapid enzymatic and antigenic detection assays allow a rapid identification; however, these methods are restricted to a limited number of microorganisms. High-throughput automated systems for subculture-based microorganism identification, which represents a large majority of biochemical identifications, are still time consuming because of the need for a subculture.

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) represents one of the most accurate, reliable, and fast methods for the identification of bacterial strains from positive cultures, and therefore it has largely replaced all other previously used approaches for microbial identification. It has proved to be much cheaper than nucleic-acid-based methods such as polymerase-chain reaction plus sequencing, fluorescence in situ hybridization (FISH), or microarrays.

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

The principle of MALDI-TOF MS is to generate the mass spectrum profile of a sample that consists of a mix of proteins with different masses. Matrix-assisted laser desorption ionization (MALDI) is a method in which proteins are embedded in a specific matrix that will facilitate the ionization achieved by a laser. The ionized proteins are subsequently separated according to their time of flight (TOF), which is a function of their mass and charge. A detector will generate a mass spectrum according to the calculated TOF. MALDI-TOF MS was initially applied for the identification of microorganisms directly from whole cells, which made it easy to implement in the routine workflow of clinical microbiology laboratory from colonies obtained on agar plates. The success of MALDI-TOF MS is largely due to easy-to-use instruments with friendly software accessible to nonchemists. MALDI-TOF MS was further applied for the identification of microorganisms from other samples, including positive blood cultures and for various applications such as typing and antibiotic resistance determination. In this chapter, we intend to summarize the main applications of MALDI-TOF MS in clinical microbiology.

3.2 Principle of Microorganisms Identification using MALDI-TOF MS

3.2.1 Soft Ionization and MS Applied to Microorganisms Identification

MALDI-TOF MS identification includes three major steps:

1) The sample (whole cells or protein extract) is deposited on a metal plate and embedded in a matrix that crystallizes the analytes. The sample is then bombarded by brief laser pulses that achieve the ionization by proton transfer from the matrix, which results in positively charged analytes (MALDI).
2) The ions formed by this process are accelerated in an electrostatic field and directed in the flight tube in which they are separated according to their TOF, a function of their mass-to-charge ratio (m/z), which is proportional to the square of the drift time $m/z = 2t^2K/L^2$ (m = mass, z = number of charges on ion, t = drift time, L = drift length, K = kinetic energy of the ion) (Figure 3.1). To avoid ion collisions in the flight tube, a high vacuum is generated by a pump, before laser pulses, which this takes 1–2 min.
3) A mass spectrum is eventually generated by protein detection at the exit of the flight tube. The mass spectrum is composed of peaks of specific mass-to-charge ratios with different intensities, which correspond to a reproducible fingerprint of a defined microorganism [5]. Some characteristic mass peaks are shared by phylogenetically related bacteria and assist identification at the genus of species level [6,7]. The identification is achieved by a comparison of the mass spectrum with a database of reference mass spectra generated with well-classified bacterial strains.

3.2.2 Biomarker Proteins

MS is a technology initially developed for the analysis of small molecules in chemistry. The use of MS for bacterial identification was first proposed in 1975, based on the fact that different bacterial extracts were associated with unique protein spectra [8]. Rapid identification of microorganisms from intact colonies using MS relies on the ability to

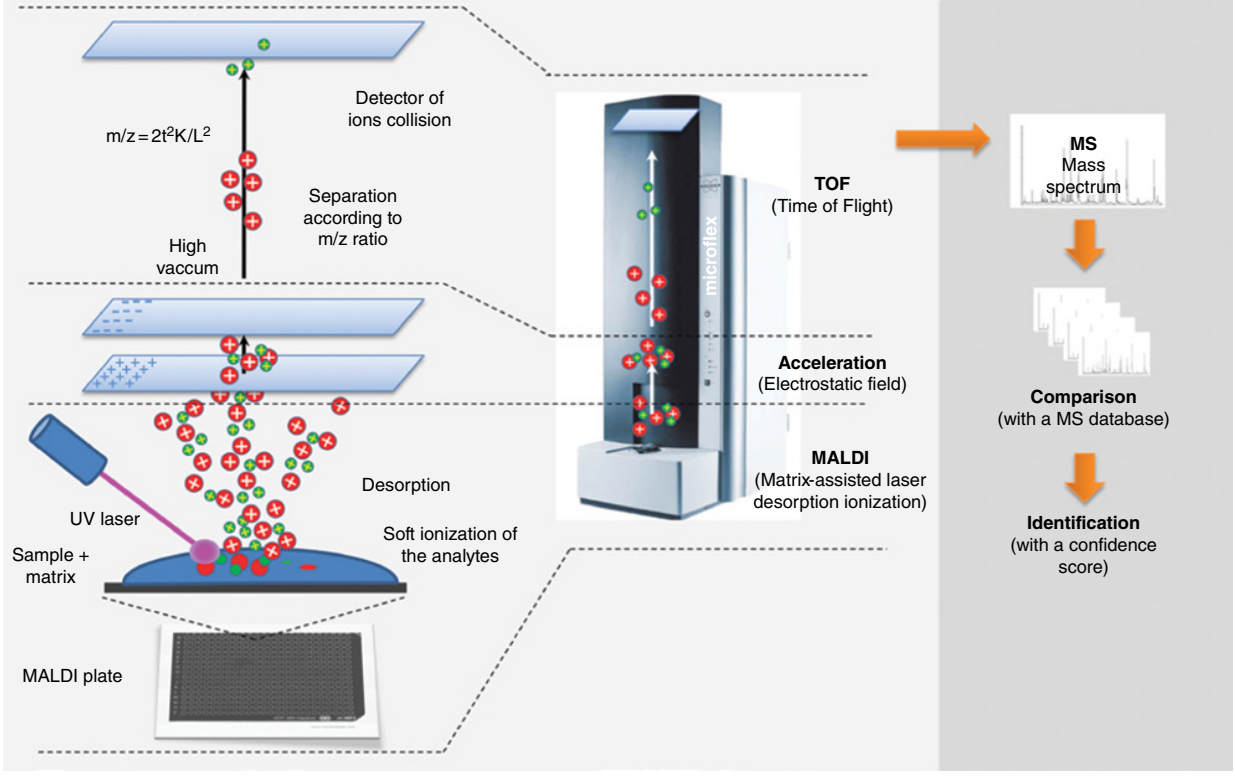


Figure 3.1 **Principle of MALDI-TOF MS identification of microorganisms.** The sample is first deposited on a metal plate and embedded in the matrix that crystallizes the analytes; it is then bombarded by brief laser pulses that achieve the ionization (MALDI) by proton transfer from the matrix, which results in positively charged analytes. The desorbed ions are then accelerated by an electrostatic field and directed in the flight tube, in which they are separated according to their time of flight (TOF) in the flight tube, in which a high vacuum is generated by a pump. Ions are detected at the exit of the flight tube, and a software generates a mass spectrum. The identification is achieved by comparison of the mass spectrum with a database of reference mass spectra (adapted from Croxatto, A. *et al.* (2012). *FEMS Microbiology Reviews* 36: 380–407).

generate ions from various proteins including ribosomal proteins. MALDI, a soft ionization technique that uses short UV laser pulses (generally around 337 nm), does not degrade extracted proteins. Holland and colleagues reported for the first time that MALDI TOF MS could be applied directly to whole cells to achieve identification by comparison with the MS obtained with either archived reference spectra or with spectra co-generated with cultures of known bacteria [7]. Batches of blinded bacterial colonies from agar plate were mixed with the matrix and air-dried, before MALDI-TOF MS. The obtained spectra were compared to the blinded spectra. A number of high-mass ions corresponding to bacterial proteins could be used for species–species matches (7). This method was confirmed by several other studies and used for rapid bacterial identification from whole bacterial cells [6,9,10].

Biomarker proteins that constitute the mass spectrum have a molecular mass lower than 15 KDa. A large majority (~50%) of the proteins used for identification are ribosomal proteins owing to their abundance and because the organic solvents and the acidic conditions that are used for cell lysis promote their extraction [11,12]. The results proposed by the identification software correspond to the more closely related species in the database and are associated with a score that integrates the number of concordant peaks as well as the quality of the spectrum. This score enables acceptance of the identification at the genus or at the species level or of its rejection according to the threshold established by the manufacturer.

3.2.3 Current Commercial MALDI-TOF MS Instruments

The success of MS in clinical microbiology relied on the commercialization of instruments requiring limited hands-on-time for the sample preparation and overall requiring limited skills for data analysis as they integrated identification software and databases. This allowed the use of MS by clinical microbiologists and not just by chemists or biochemists.

The Biotyper mass spectrometer systems (Bruker Daltonics) are the most widely used. They allow identification from bacteria or fungal cells and make it possible to implement the database with spectra generated with the users' own isolates. The identification is associated with a log scale of 0 to 3. A score above 2.3 corresponds to a high level of confidence of the identification at the species level. A score between 2.0 and 2.3 allows identification at the species level with a good probability. A score between 1.7 and 2.0 suggests a correct identification at the genus level. Scores above 1.7 have limited accuracy.

The VITEK MS (bioMérieux, France) is the former Axima Assurance system (Shimadzu Corporation, Kyoto, Japan). The identification are rated as "ready to report," "requiring further review," or "no identification made" [13,14].

The Autoflex II mass spectrometer (Bruker Daltonics) and the Axima Assurance system (Shimadzu), which are the most widespread, display similar performance for routine identification of microorganisms both in term of percentage of valid identifications and percentage of correct identifications. A comparative study between the two systems performed on 720 consecutive bacterial colonies obtained in routine clinical laboratory conditions and using the sequencing of the 16S rRNA gene as gold standard reported 99.1% (674/680) of correct identifications for the Bruker MS system (Bruker) and 99.4% (635/639) for the Shimadzu MS system [15]. In a similar study, Martiny *et al.*

reported 92.7% and 93.2% of correct identifications for the Biotyper and the VITEK MS systems, respectively ($n = 986$) [16]. Both systems displayed equivalent performance for identification of yeasts: 93.0% (175/188) for the VITEK MS and 92.6% (174/188) for the Bruker Biotyper [17]. Other studies comparing the VITEK MS and the Bruker systems confirmed the equivalent performances of the two instruments [18–21]. The Andromas system (Andromas SAS, Paris, France) uses a software that determine the percentage of similarity of the obtained spectrum with reference strains present in the database which provide a "good identification," an "identification to be confirmed," or an "absence of identification" [22–24]. To date only a few studies have been published.

3.2.4 Automated Colony Picking

Manual colony picking is a major limitation of the throughput of MS identification and causes major errors due to inversions, which are estimated to occur at the rate of 0.25% [25]. The throughput as well as the occurrence of major errors should be improved by the emergence of new automated systems for clinical microbiology laboratories, such as the Total Laboratory Automation (TLA) system from BD Kiestra (the Netherlands) and the WASPLab from Copan Diagnostics (Italy), including automated colony picking systems [26]. Because MALDI-TOF MS is currently used to identify more than 95% of all isolated strains in laboratories, should they be implemented, such systems will undoubtedly have a major positive impact on the laboratory technicians' overall workload.

3.3 Factors Impacting the Accuracy of MALDI-TOF MS Identifications

3.3.1 The Importance of the Database

MALDI-TOF MS accuracy for a microorganism's identification largely relies on the representativeness and the quality of the reference spectra database [27,28]. Commercial databases are made of reference spectra that are expected to cover the maximum number of clinically relevant isolates. These spectra are generally made of an accumulation of several spectra obtained for a single isolate or for different isolates. Reference spectra can also be made with isolates from different origins as regards to potential geographic diversity. However, a comprehensive database does not exist yet.

Some misidentifications can be due to an insufficient number of reference spectra for a specific species or to the absence of reference spectra. The databases need to be frequently updated with new spectra. Most of the systems allow users to expand the commercial database with spectra generated with "local" isolates. For instance, one could implement the database with the spectrum of an isolate unsuccessfully identified using MALDI-TOF MS but that has been unambiguously identified using other approaches such as 16S rDNA sequencing or other molecular approaches; this has been shown to significantly increase the identification rate [29,30]. To reduce inter-analysis variability, new references spectra should be generated by the addition of several spectra obtained with a defined isolate. Errors in the database due to a wrong identification of the strain that will serve as reference or taxonomic errors constitute a

major source of misidentifications. Thus, implementing one's own spectra with local isolates requires specific care on the part of the user, especially regarding the methods used for the identification.

Some BSL3 agents such as *Burkholderia pseudomallei, Brucella suis, Brucella melitensis, Francisella tularensis*, or *Bacillus anthracis* are not present in all commercial databases and are included in dedicated databases [31].

3.3.2 Quality of the Spectrum and Standardization of the Pre-analytic

Both for bacteria and fungi, the mass spectrum is dependent on the age of the colony and the growth medium. Therefore, it is recommended to perform MS identification from fresh colonies. In addition, when implementing the database, it is important that the new reference spectra be made in similar conditions to those used for routine identifications. The quality of the spectrum may also vary according to the investigated species. For instance, difficult-to-lyse bacteria such as *Klebsiella pneumoniae*, which is encapsulated, can lead to poor identification scores. Other organisms such as *Mycobacteria* spp., *Nocardia* spp., Actinomycetes, yeasts, and fungi need specific extraction protocols [32,33].

3.3.3 Limit of Detection

A sufficient biomass is required to generate a spectrum whose quality is sufficient to provide identification. MALDI-TOF MS was initially restricted to colonies obtained from culture on solid or liquid media. A major improvement in the diagnosis of bloodstream infections and patient's management is the identification of bacteria and fungi directly from positive blood cultures allowed by the validation of specific protocols aimed at removing erythrocytes and nonmicrobial cells elements and to concentrate bacterial or fungal cells [34]. Recent studies also support the potential use of MALDI-TOF MS for a microorganism's identification directly from clinical samples such as urine [27].

3.3.4 Errors and Misidentifications

Absence of identification can be due to incomplete databases; however, a large number of identification failures result from the poor quality of the mass spectrum obtained for the microorganism of interest, which can impair the identification of a concordant spectrum in the database. This can be due to a poor biomass or to inefficient cell lysis or inefficient protein extraction. Erroneous identifications of the reference spectra in the database can be a cause of major errors [25,28]. Misidentifications often stem from similar analytes displaying different TOFs as a result, for instance, of an incomplete ionization. This can be avoided by using a fresh matrix. Finally, the routine identification procedure has to be similar to the procedure used to obtain the mass spectrum for the database as these could impact the quality and quantity of the peaks and impair identification.

3.3.5 Mixed Bacterial Populations

MALDI-TOF MS has been developed for a microorganism's identification from pure culture; identification of a mixed bacterial population can lead to unexpected results.

The Gram staining that remains mandatory could disclose a mixed infection. Indeed, identification from mixed populations can generate a nonexistent mass spectrum resulting from the sum of two or more spectra from phylogenetically distinct bacteria. Alternatively, the identification software could propose two distinct strains that are not genetically related. In general, the identification scores hardly reach the threshold for high confidence identification at the species level. It is essential that the MALDI-TOF MS identification match the microorganism's morphotype revealed by the Gram staining and/or suggested by phenotypic characteristics. Thus, identifications obtained from mixed populations should be considered as not definitive; a subculture is recommended for the isolation of the microorganisms and further identification on monobacterial colonies.

3.3.6 Closely Related Species

Some closely related species cannot be differentiated using MALDI-TOF MS even with high-quality spectrum using routine procedures (Table 3.1). This is the case of *Streptococcus* spp. of the "viridians" group. For *Streptococcus pneumoniae* identification, additional phenotypic tests such as the Optochin susceptibility test or the bile solubility test are required for definitive identification [35–40]. Some strategies, based on the analysis of specific peaks (6949, 9876, and 9975 m/z), allow discrimination between *S. pneumoniae* and group mitis *Streptococcus* species; however, they are hardly suitable for routine application [39]. The same issue is true for the distinction between *Escherichia coli* and *Shigella* spp., for which even when specific peaks are used, misidentifications still occur [16; 41–43]. In the absence of phenotype confirmation such as lactose fermentation or lysine decarboxylase activity, the identification of *E. coli* should be presumptive. In practice, one may report the identification of some bacteria as a "complex" or "group" of bacteria when the clinical relevance of differentiating some species is questionable and more than one species of these complex are generally obtained with an identification score greater than 2 (Table 3.1) [44]. Nevertheless, such strategies should be updated according to improvements in database availability, new bioinformatic algorithms, and taxonomy evolutions. On the basis of empirical observations, some authors have proposed an alternative acceptance criteria for gram-negative bacteria, which consist in the acceptance of identification at the species level when at least a 0.200 log difference is observed between multiple species present besides a log score ≥2.0 [45].

3.4 Identification of Microorganisms from Positive Cultures

3.4.1 Identification from Positive Cultures on Solid Media

The main application of MALDI-TOF MS in clinical microbiology laboratories is the identification of bacteria from colonies recovered from solid culture media (Figure 3.2). A standard procedure is suitable for a large majority of the bacteria, which is convenient for routine identification. Specific identification procedures, which will be discussed later in this chapter and are presented in detail in chapters of this book, are needed for some bacteria, such as *Actinomycetes* and *Mycobacteria*.

Table 3.1 Example of interpreted results as currently reported in our laboratory.

Complex of bacteria	Species
Acinetobacter baumanii complex	*A. baumanii* *A. calcoaceticus* *A. nosocomialis* *A. pittii*
Bacteroides fragilis complex	*B. caccae* *B. eggerthii* *B. fragilis* *B. ovatus* *B. stercoris* *B. thetaiotaomicron* *B. uniformis* *B. vulgatus*
Burkholderia cepacia complex (depending on the clinical situation, identification to species level may be clinically relevant)	*B. ambifaria* *B. anthina* *B. arboris* *B. contaminans* *B. cenocepacia* *B. diffusa* *B. dolosa* *B. lata* *B. latens* *B. metallica* *B. multivorans* *B. pyrrocinia* *B. seminalis* *B. stabilis* *B. vietnamiensis* *B. ubonensis*
Citrobacter freundii complex	*C.brakii* *C. freundii* *C. gillenii* *C. murliniae* *C. rodentium* *C. sedlakii* *C. serkmanii* *C. youngae*
Enterobacter cloacae complex	*E. asburiae* *E. cancerogenus* *E. cloacae* *E. dissolvens* *E. hormaechei* *E. kobei* *E. ludwigii* *E. nimipressuralis*

Table 3.1 (Continued)

Complex of bacteria	Species
Haemophilus influenzae complex	*H. haemolyticus* *H. influenzae*
Klebsiella oxytoca complex	*K. oxytoca* *Raoultella ornithinolytica* *Raoultella planticola* *Raoultella terrigena*
Proteus vulgaris complex	*P. hauseri* *P. penneri* *P. vulgaris*
Streptococcus bovis complex	*S. equinus* *S. gallolyticus* *S. infantarius* *S. lutetiensis* *S. pasteurianus*
Streptoccus mitis complex	*S. gordonii* *S. infantis* *S. massiliensis* *S. mitis* *S. oralis* *S. peroris* *S. parasanguinis* *S. sanguinis* *S. tigurinus*
Streptococcus salivarius complex	*S. salivarius* *S. vestibularis* *S. thermophilus*

For routine identification of bacteria, colonies are picked from agar plates using a sterile loop, deposited on the metal MALDI plate, and overlaid with a matrix [27,46,47]. The mixture is air-dried (a step that can be accelerated using a warming plate) and then introduced in the device and subjected to laser pulses. Matrices with different properties can be used: 2,5-dihydroxybenzoic acid (DHB), a-cyano-4-hydroxycinnamic acid (CHCA), sinapinic acid (SA), ferulic acid (FA), and 2,4-hydroxy-phenyl benzoic acid [27,46]. Each device is generally associated with a specific matrix recommended by the manufacturers. This procedure is straightforward, simple, and fast: less than 5 min from the colony picking to the final identification. To improve the quality of the spectrum and to increase the accuracy (confidence score) of the identification, the analysis can be achieved following protein extraction using an acetonitrile/formic acid procedure that promotes the generation of positive ions and increases the number of peaks available for the analysis [25,48]. The formic acid can be added directly on the colonies on the MALDI plate and air-dried before the embedding into the matrix [27]. If no identification is obtained, the third option is to perform a full protein extraction using ethanol and formic acid.

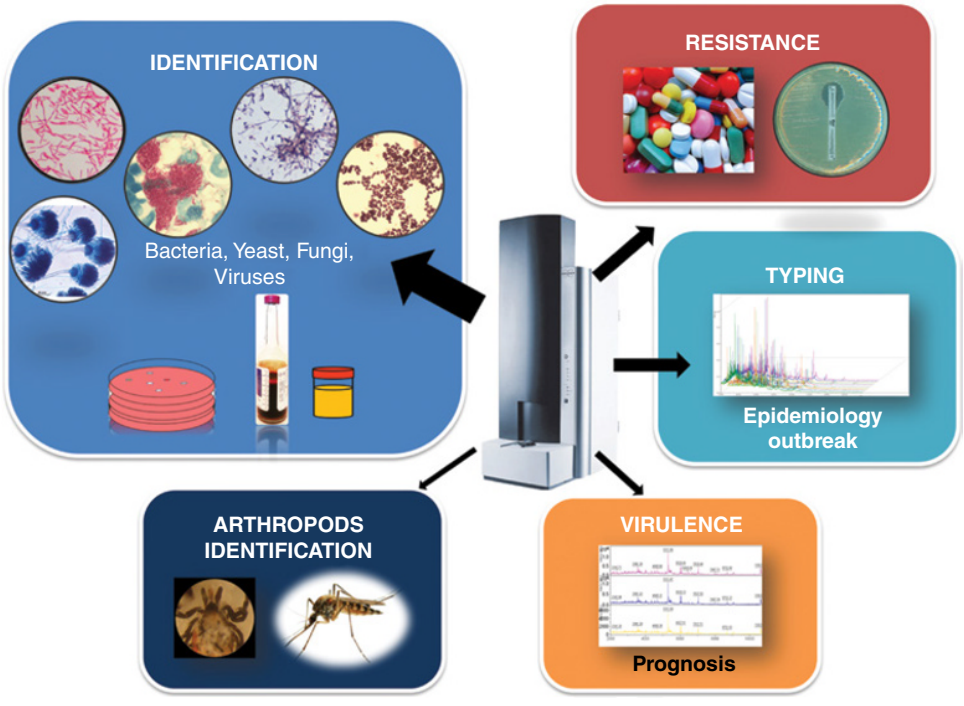


Figure 3.2 Application of MALDI-TOF MS in clinical microbiology.

The performance of MALDI-TOF MS identification relies on (1) the number of mass spectra that reach the quality allowing identification and (2) the number of correct identifications. For these two parameters, all the commercial devices available display similar performances. The percentage of correct identifications ranges from 95% to 99% with the Bruker system [15,25,28], whereas the VITEK MS system provided 96.2% of correct identifications at the species level [49]. It is recommended to avoid working with old colonies and to minimize the number of passages.

3.4.2 Identification from Positive Blood Cultures

Bloodstream infection (BSI) is suspected by physicians on the basis of nonspecific criteria, which leads to the introduction of empiric antimicrobial treatments. The time to identification of the etiologic agent of a BSI is crucial as the mortality and morbidity of patients are directly dependent on the introduction of the first efficient anti-infectious treatment. Empiric treatments, introduced prior to the identification of the incriminated microorganism and its susceptibility profile, are generally based on broad-spectrum antibiotics with a deleterious impact on the profitable microbiota.

3.4.2.1 Identification from Positive Blood Cultures Via a (Short) Subculture

Positive blood cultures are not suitable for direct identification using MALDI-TOF MS as the bacterial concentration is generally too low (10^6 to 10^9 cells/ml) and because of the presence of an excess of nonbacterial proteins that would impair MS identification [50]. A first strategy is to perform a subculture of the positive blood culture (BC) in order to

obtain a pure culture on agar-based medium, on which the standard MALDI-TOF MS identification methods described above can be applied. However, this is a time-consuming method. Indeed, considering that MALDI-TOF MS identification turnaround time (TAT) is lower than 5 min, it is regrettable that the time to result because the positive BC is detected is delayed by several hours because of a long subculture. To accelerate the time to result from the detection of the positive BC, an alternative strategy is to achieve a short subculture on an agar plate. The efficiency of this procedure is, however, dependent on the bacterial strain growth rate [27]. This approach is particularly suitable for gram-negative bacteria for which subcultures of ≤4, ≤6, ≤8, and ≤12 h yielded 95.2%, 97.6%, 97.6%, and 97.6% of correct identifications; in contrast, the same incubation time yielded only 18.6%, 64.0%, 96.5%, and 98.8% of correct identifications for gram-positive bacteria [51]. With an incubation period of 5 h, Verroken *et al.* reported 81.1% (727/896) for monomicrobial cultures with misidentifications mainly occurring for yeasts and anaerobes [52]. A subculture is especially suitable for laboratories with a large volume of the positive BC and with limited technicians, as up to now subculture-independent methods that will be presented below require a lot of hands-on time (Table 3.2) [34].

3.4.2.2 Directly from the Positive Blood Culture Vial

In order to reduce the TAT to final identification from positive BCs, several subculture-independent strategies, whose principle is to remove nonmicrobial material (erythrocytes, cells debris) and to concentrate the microorganisms, have been developed and reported in numerous publications. The first method consists of bacterial enrichment from the positive BC by centrifugation and erythrocyte lysis with ammonium chloride solution. This procedure gives a successful identification for 78.7% of the bacterial pellets tested with 99% of correct identifications [53,54]. A method using collection tubes with separator gels in which 1.5 ml from the positive BC is injected before bacterial protein extraction and MALDI-TOF MS analysis gives more than 90% of correct identifications [35]. A similar method using a tube containing a separator gel and a clot activator leads to 95.3% of correct identifications [36]. Equivalent results can be obtained with commercial pellet preparation kits [37,55–57].

The pellet enriched in bacterial cells generated from a positive BC for MALDI-TOF MS identification can also be used for ASTs using disk diffusion methods, E-test (bioMérieux, France), or automated systems. Prod'hom *et al.* reported 99% and 74% of correct identifications for Enterobacteriaceae and for staphylococci, respectively, and less than 1% of errors for AST using the VITEK 2 automated system [58]. More recently, another group obtained similar results using saponin and chloride ammonium. This strategy has been transposed to the BD Phoenix automated system with 97.9% of correct AST [59]. Using a commercial lysis-filtration method of bacterial enrichment from a positive BC associated with the VITEK MS system, Machen *et al.* reported a TAT of 11.4 h until the AST result, when the TAT for conventional methods was 56.3 h [60].

3.5 Identification of Microorganisms Directly from Samples

3.5.1 Urine

Ferreira *et al.* first reported the possibility of diagnosing urinary tract infections by direct detection of bacteria in urinary samples [61]. The procedure consisted of an initial

Table 3.2 MALDI-TOF MS applications in clinical microbiology and references.

Application	Comments	Selected references
Standard identification from colonies	Fast, accurate; misidentification associated with absence or incorrect identification of the reference spectra in the database.	[25,27,28,49]
Identification from positive blood culture via a short subculture	Suitable for gram-negative bacteria; not suitable for gram-positive bacteria, anaerobes, and yeast.	[51,52]
Identification from positive blood culture via bacterial pellet	Positive impact on patient management; misidentification or absence of identification is generally associated with insufficient biomass or mixed bacterial population. High hands-on time.	[34–37,144–146]
Identification from urine	For culture positive with more than 10^3–10^5 CFU/ml depending on the study; high workload, unknown impact on patient management.	[61–63,147]
Identification from cerebrospinal fluid	Proof of concept provided for a meningitidis case.	[65]
Alpha-hemolytic streptococci	Misidentification of group mitis streptococci and *S. pneumonia*.	[39,148]
Beta-hemolytic streptococci	Fastest method; more accurate than conventional methods. Errors can occur with rare species.	[133]
Aerobic gram-positive bacilli	Accuracy depends on the database.	[23,90]
Shigella* spp./*E. coli	Routine distinction methods with 100% sensitivity and specificity are not available.	[16,41–43]
Anaerobes	Important added value due to the long incubation time required for these microorganisms.	[127–132]
Mycobacteria	Rapid for the identification of NTM. No added value for MTBC complex identification. Needs specific inactivation and cell lysis procedure, no added value for laboratory with a molecular diagnosis platform.	[69–71,149]
Yeast	Accurate; same procedure as for bacteria.	[17,24,150–154]
Fungi	Requires specific extraction procedure to achieve cell lysis. Impacted by the age of the colony and the growth stage.	[22,75]
Dermatophytes	Variable sensitivity according to the species, limited clinical impact of identification at the species level.	[76]
Protozoan parasites	Amoeba. *Leishmania* species. *Giardia* species.	[139] [137] [136]

Table 3.2 (Continued)

Application	Comments	Selected references
Ticks and fleas	Proof of concept from hemolymph (ticks) and various body parts (fleas).	[140,143,155,156]
Typing	Can be accurate but is still time consuming.	[81,82,91–93]
Resistance	Essentially for carbapenemase-producing strain identification.	[78–80]
Toxin	Requires identification of exact virulence factor-associated peak. Misidentification can be due to type-associated peaks rather than toxin-associated peaks.	[85,88,89]
Viruses	Detection, identification (including simultaneous detection), typing, nucleic-acid methylation.	[96,97,106–108,157]
PCR-ESI MS	Universal (>800 pathogens), detection of mixed bacterial populations, semiquantitative. Low throughput (8 samples/8 h), expensive, no interventional studies.	[50,111,118,158]

centrifugation of 4 ml of urine at low-speed (2000 g) to remove leukocytes, followed by a high-speed centrifugation (15,500 g) to obtain a bacterial pellet suitable for MALDI-TOF MS analysis. The performance of MALDI-TOF MS was compared to the performance of the culture for the detection and identification of 260 positive urine samples using the automated screening device UF-1000i (bioMérieux, France) based on flow cytometry as the reference. Among the 260 samples, 235 corresponded to monobacterial cultures with $> 10^5$ CFU/ml. MALDI-TOF MS correctly identified 91.8% (202/235) microorganisms at the species level and 92.7% (204/235) at the genus level [61]. In another study, March *et al.* also reported a threshold of 10^5 CFU/ml for successful identification directly from urine [62]. In contrast, Kohling *et al.* reported that the identification of bacteria directly from urine samples is possible for concentrations as low as 10^3 CFU/ml. The authors also noticed that the presence of human defensin that gets inserted into the membrane of bacteria and cannot be removed by washing had a negative impact on MALDI-TOF MS identification [63]. A screening of positive urine using the automated system UF-1000i followed by direct analysis using MALDI-TOF MS has been reported to give 94.8% (1381/1456) of correct identifications [64]. Altogether, these studies demonstrate the potential of MALDI-TOF MS to detect pathogens directly from urine. However, given the workload associated with such procedures and the number of urine samples to investigate, this approach is not widely used. Moreover, the impact of such an approach on patient management still needs to be investigated and might be limited if MALDI-TOF MS rapid identification is not associated with rapid AST.

3.5.2 Cerebrospinal Fluid

To date, no study has addressed the reliability and usefulness of MALDI-OF MS for the detection of pathogens directly from cerebrospinal fluid (CSF). Nevertheless, a case

report discussed the feasibility of such a strategy. In this report, CSF from an unconscious febrile patient who presented with a nuchal rigidity, was sent to the clinical microbiology laboratory and analyzed by MALDI-TOF MS using the same protocol that was described above for urine: low-speed centrifugation (2000 rpm for 30 s) to remove leucocytes and then high-speed centrifugation (13,000 rpm for 5 min) to obtained a bacterial pellet deposited onto the MALDI plate for analysis. This approach proposed *Streptococcus pneumoniae* as the etiological agent [65]. There is a need for additional studies to determine the reliability and the added value of such approaches.

3.6 Microorganisms Requiring a Specific Processing for MALDI-TOF MS Identification

Although most of the bacteria can be identified from positive cultures using MALDI-TOF MS, some of them would require specific processing mainly aimed at promoting cell lyses. In addition, some microorganisms require an inactivation step for biosafety reasons. Several chapters of this book are dedicated to the identification of specific bacteria using MALDI-TOF MS; nevertheless, in this section, we have selected and emphasized some of these aspects.

3.6.1 *Nocardia* and Actinomycetes

Nocardia spp. and Actinomycetes have a difficult-to-lyse cell wall requiring specific processing prior to MALDI-TOF MS identification. Verroken *et al.* developed a method for identification of *Nocardia* spp. that consists of 30 min of boiling, followed by a full protein extraction using ethanol and formic acid [66]. Using this procedure together with an extended homemade database made of 110 isolates, the authors could accurately identify 88% (38/43) of the *Nocardia* strains including 34/43 correct identifications at the species level [66]. In contrast, only 44% (19/41) of correct identifications, including 10 correct identifications at the species level, were obtained using the MALDI Biotyper commercial database [66].

3.6.2 Mycobacteria

The identification of *Mycobacterium tuberculosis* from a positive culture is generally achieved by antigen detection and by PCR. Similarly, the identification of non-tuberculosis mycobacteria (NTM) can also be achieved by PCR or other nucleic-acid-based methods such as DNA–DNA hybridization. Mycobacteria identification using MALDI-TOF MS has been proposed to decrease the TAT, labor, and cost compared to nucleic-acid-based methods. Several procedures including bacteria inactivation and cell disruption have been developed for the different MALDI-TOF MS systems commercially available with similar performances [67; see Chapter 4]. However, the concordance of *Mycobacterium* species identification using MALDI-TOF MS with molecular-based methods varies according to the studies and the protocols. This is because some closely related strains can hardly be distinguished using this method [68]. However, MALDI-TOF MS may interest laboratories lacking a molecular diagnosis platform for nucleic acid extraction, amplification, and sequencing or for nucleic acid hybridization.

Prior to MS identification, *Mycobacterium* species need specific sample processing to inactivate the microbes and to achieve cell disruption. Inactivation can be achieved by incubating the bacteria at 95 °C for 30 min in 70% ethanol. A sonication step is necessary to optimize cell disruption before protein extraction. Using this procedure on cells collected from 7H11 agar medium, Machen *et al.* reported 82.2% of correct identifications at the species level and 88.8% at the genus level [69]. Kodana *et al.*, proposed a slightly different procedure in which the heat inactivation is achieved in water with a subsequent ethanol treatment. Alternatively, El Khechine *et al.* proposed a heat inactivation performed in the presence of Tween-20 followed by a mechanic cell lysis and a full protein extraction; using this method, the authors reported 100% of agreement with conventional molecular methods for 124 clinical isolates, including 87 *M. tuberculosis* strains and 37 NTM strains [70]. Cell inactivation and disruption can also be achieved simultaneously by vigorous vortex of the cells during 15 min in 70% ethanol in the presence of glass bead, which gave 88.8% of correct identifications [69]. A heat-independent inactivation/extraction method has been proposed by Lotz *et al.* in which a cell pellet obtained from positive MGIT or Lowenstein–Jensen culture was inactivated in 70% ethanol, which was sufficient for work outside Biosafety Level 3 (BSL3) laboratories [71]. The cell suspension was deposited on the MS plate and air-dried before the addition of the matrix (SA, 20 mg/ml; acetonitrile, 30%; trifluoroacetic acid, 10%). After the addition of 10 mM ammonium phosphate on the crystallized mixture, the sample was analyzed by MS, giving 97% of correct identifications from Lowenstein–Jensen and 77% of correct identifications from MGIT [71]. Very recently, another group demonstrated that a 5 min 70% ethanol inactivation may be sufficient for mycobacterial identification [33].

In conclusion, MALDI-TOF MS can provide a same day result from positive mycobacteria cultures. Inactivation of the cells should be ensured in order to perform the analysis outside BSL3 laboratories. The inactivation procedure is generally not sufficient for cell disruption, which requires a specific step. However, misidentifications can occur for closely related strains or for rare species for which there are not enough reference mass spectra in the database [68,71]. This is the case for *M. abscessus*, *M. massiliense*, and *M. bolletii* and for *M. tuberculosis*, *M. bovis*, *M. bovis BCG*, *M. microti*, and *M. africanum*. In conclusion, *Mycobacteria* spp. can accurately be identified using MALDI-TOF MS.

3.6.3 Yeast and Fungi

Yeast identification using MALDI-TOF MS can be achieved directly from whole cells using a procedure similar to the standard bacterial identification procedure from agar plates (see Chapter 9). The two main commercial MALDI-TOF MS systems display similar performances for the identification of yeast at the species level with up to 89.8% of correct identifications for the Bruker Biotyper MS system and 84.3% for the VITEK MS system [17,72]. These performances are comparable to the VITEK 2 system [17]. A full extraction method that consists of a pretreatment in 70% ethanol and formic acid/acetonitrile protein extraction gave 99% of correct identifications using the Bruker Daltonics MALDI Biotyper software [73]. The age of the colony is expected to impact the accuracy of yeast identification; however, Goyer *et al.* reported similar results for colonies grown in 48 h or 72 h, with respectively 95.1% and 96.6% of correct identifications

from chromogenic agar media [74]. Nevertheless, caution is warranted for much older colonies (≥5 days).

The identification of invasive filamentous fungi using protein MALDI-TOF MS is more complex than for bacteria and yeast, because of their complex life cycle (see Chapter 9). Indeed, different life stages, spores, conidiophores, and mycelium can be present on the same colony. A first strategy is to mix the different growth stages present in a single colony, and then to perform protein extraction for MALDI-TOF MS analysis [22]. Specific sample preparation methods are also required to achieve cell lysis and to improve protein extraction. In contrast to yeast identification, the accuracy of MALDI-TOF MS for the identification of filamentous fungi is largely dependent on the growth medium, the age of the colony, and the growth stage. Repeated identification can be necessary. Using a simplified extraction method and an extended database generated with 55 species of *Aspergillus* spp., *Fusarium* spp., and *Mucorales* spp., De Carolis *et al.* reported 96.8% (91/94) of correct identifications [75]. Samples were prepared by placing 1 µl of a suspension of mycelium and conidia onto the MALDI plate, from which direct protein extraction was achieved using ethanol/alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile/2.5% trifluoroacetic acid [75].

Dermatophyte identification at the species level is generally not necessary because the antibiotic susceptibility profile is generally shared by all the members of a defined group. If identification at the species level is required, an in situ identification by microscopy and PCR is possible. The identification of dermatophytes using MALDI-TOF MS is complicated by the fact that microorganisms are difficult to recover after culture. Dermatophytes require a long (4 to 5 d) culture, with a significant impact on protein expression, which can impact the MALDI-TOF MS analysis. Nevertheless, some protocols have been proposed. As for other filamentous fungi, it is recommended to pool the different growth stages present in the same colony or to multiply the identifications within the same colony. Using an extended database, De Respinis *et al.* reported 60% to 100% of correct identifications of dermatophytes depending on the species [76]. More recently, Wang *et al.* also successfully used MALDI-TOF MS, not only to identify yeasts at the species level, but also dermatophytes [77].

3.7 Detection of Antimicrobial Resistance

3.7.1 Carbapenemase Detection

Carbapenemases are enzymes produced by some *Enterobacteriaceae* and some nonfermenting gram-negative bacteria that can hydrolyze carbapenems, a class of broad-spectrum antibiotics. The first approach to the identification of carbapenemases-producing strains using MALDI-TOF MS is based on the detection of ertapenem hydrolysis corresponding to a shift or a disappearance of the antibiotic-specific peak on the mass spectrum. Kempf *et al.* focused on *Acinetobacter baumannii*, a bacterium that mainly produces three families of carbapenemases: OXA-23-like, OXA-24-like, and OXA-58-like. On the mass spectrum obtained by MALDI-TOF MS, carbapenemase production resulted in the reduction of a 300.0 m/z peak corresponding to native imipenem and the increase of a 254.0 m/z peak corresponding to the natural metabolite of imipenem in the supernatant of bacteria incubated with

imipenem during 4 h [78] (see Chapter 15). This method displayed 100% sensitivity and 100% specificity for the detection of carbapenemase-producing strains. Carvalhaes *et al.* monitored ertapenem degradation through the disappearance of the 475 and 497 m/z peaks corresponding to the native form and to the monosodium salt form of the molecule. This strategy identified 72.4% (21/29) of the carbapenemase-producing strains including 100% of the KPC-2 and 100% of the SPM-1 (100%) but hardly any OXA-23-producing *Acinetobacter baumannii* strains [79]. A similar method described by Vogne *et al.*, monitoring ertapenem degradation, showed 100% sensitivity and 100% specificity for the detection of carbapenemase-producing strains among *Klebsiella pneumoniae*, *Escherichia coli*, *Enterobacter cloacae*, *Providencia stuartii*, *Serratia marcescens*, *Enterobacter aerogenes*, *Hafnia alvei*, *Klebsiella oxytoca*, *Proteus vulgaris*, *Morganella morganii*, *Pseudomonas aeruginosa*, *Acinetobacter baumanii*, and *Aeromonas* species [80]. The simple approach proposed by Vogne *et al.* exhibited a short TAT (~60 min) as it used a commercially available antibiotic disk, which reduced the hands-on-time. In addition, this approach could benefit from the high stability of the antibiotic molecule in such commercial formulations.

3.7.2 Methicillin-Resistant *S. aureus*

Approaches, more related to typing, aim to identify genotype associated with specific resistance profiles using MALDI-TOF MS. They have been mainly used to distinguish between methicillin-susceptible *S. aureus* (MSSA) and methicillin-resistant *S. aureus* (MRSA) [81,82]. These approaches correspond to the detection of biomarker proteins associated with resistance strains. However, Edwards-Jones *et al.* suggested that the identification of some MALDI-TOF MS characteristic peaks could be used for the distinction between methicillin susceptibility and methicillin-resistant *S. aureus* [83].

3.7.3 Vancomycin-Resistant Enterococci

By comparing spectra obtained from vancomycin-sensitive *Enterococcus faecium* and vancomycin-resistant isolates (VRE), Griffin *et al.* demonstrated that a 5945 Da peak could distinguish sensitive and resistant strains. Another peak at 6603 Da could distinguish vanA- and vanB-positive isolates [84]. This likely represents strain-type identification rather than antibiotic susceptibility markers and should be treated with caution.

Other aspects of the antibiotic resistance detection based on strain typing will be addressed in chapters 10 and 15. Altogether these studies suggest that MALDI-TOF MS is suitable for antibiotic resistance determination. However, these studies need to be extended to more bacterial isolates and need to be reproduced in other locations to determine if the accuracy of these approaches is impacted by the local epidemiology of resistance isolates.

3.8 Detection of Bacterial Virulence Factors

Because of its accuracy, MALDI-TOF MS has the potential to detect specific peaks corresponding to virulence factors such as toxins. Gagnaire *et al.* demonstrated that MALDI-TOF MS applied to *S. aureus* whole cells could detect two peaks at 3005 Da and

3035 Da corresponding to the *S. aureus* delta toxin, a biomarker of the activity of the *agr* (accessory gene regulator) system [85]. In *S. aureus*, the *agr* system regulates the expression of numerous virulence factors and pathogenesis-associated determinants, which represents a therapeutic target [86,87].

MALDI-TOF MS has also been proposed for *S. aureus* strain typing or for the detection of biomarkers of the most virulent toxigenic isolates. The Panton–Valentine leukocidin (PVL), a toxin produced by some strains of MSSA and MRSA, is associated with recurrent skin and soft tissue infections as well as necrotizing pneumonia. The gene encoding the PVL brought by phages has a high potential to spread among isolates. Bittar *et al.* initially reported the identification of two peaks of 4448 and 5302 Da associated with PVL-positive isolates [88]. However, a second study revealed that there were no reliable associations between the 4448 and 5302 Da peaks and the presence of PVL; PVL-positive strains were identified by chance because of similar mass profiles due to their high relatedness [89].

Konrad *et al.* reported the correct identification of 99.1% (116/117) of toxigenic *Corynebacterium* sp. (*C. diphtheriae*, *C. ulcerans*, and *C. pseudotuberculosis*) among non-toxigenic species using the MALDI-TOF MS system Microflex LT mass spectrometer (Bruker Daltonics) and the Biotyper 2.0 identification software. This result was not obtained directly from colonies but after rapid protein extraction using ethanol and formic acid. The discordant result was a *C. tuberculostearicum* isolate that could neither be identified at the species level by using the MALDI-TOF MS (log (score) of 1.8) nor by sequencing the *rpoB* gene or by using the API Coryne gallery [90]. When the score was above 2.0, the negative predictive value and the positive predictive value were both 100%, attesting to the high reliability of the detection of potential toxigenic *Corynebacterium* species.

3.9 Typing and Clustering

At present, the reference methods for isolate typing are nucleic-acid-based methods such as multi-locus sequence typing (MLST) and pulsed-field gel electrophoresis (PFGE) because of their excellent discriminatory power that only whole genome sequencing can surpass. However, these methods can be time consuming and expensive, and so more rapid and cost-effective methods are needed.

3.9.1 MRSA Typing

S. aureus typing by MLST or PFGE generally relies on the *spa* gene (*S. aureus* protein A gene). Wolters *et al.* reported that the analysis of 13 characteristic peaks allows discrimination of the major *S. aureus* MRSA lineages by MALDI-TOF MS with a discriminatory index of 0.770 (95% CI 0.671–0.869), which is comparable to the discriminatory indices of genotyping using the *spa* gene 0.818 (95% CI 0.721–0.915) [82]. Lu *et al.* proposed a peak-based approach allowing discrimination between *S. aureus* SCC*mec* types IV and V isolates that are mainly community-associated (CA) MRSA and SCC*mec* types I–III isolates that are hospital-associated (HA) MRSA [81]. Josten *et al.* analyzed the mass spectrum profile of 401 MRSA and MSSA strains and identified peaks that differentiate between the main *S. aureus* MRSA and MSSA clonal

complexes and could be used to distinguish between sensitive and resistant strains [91]. To conclude, these studies suggest that MALDI-TOF MS is a reliable method for the typing

3.9.2 Enterobacteriaceae Typing

Noteworthy typing using MALDI-TOF MS requires a precise calibration of the instrument using highly conserved peaks and generally requires a database of references spectra representing the major lineages of the species of interest. However, Christner *et al.* developed an approach that includes the identification of biomarker peaks for MALDI-TOF MS typing during an outbreak. This approach identified two peaks, one at m/z 6711 and the other at m/z 10883, that correctly classified 292/293 isolates of Shiga-toxin producing *E. coli* during the 2011 outbreak that occurred in northern Germany [92]. The reliability of MALDI-TOF MS for typing isolates directly from whole cells has also been demonstrated for *Yersinia enterocolitica* [93] and *Salmonella enterica* [94]. Interestingly, strain typing using MALFI-TOF MS generally relies on the analysis of selected specific biomarkers only.

3.9.3 Typing *Mycobacterium* spp.

By comparing the mass spectra of *M. abscessus*, *M. massiliense*, and *M. bolletii* isolates, Suzuki *et al.* identified a cluster of *M. massiliense* strains closely related to *M. abscessus* due to the similarity of their mass spectra, indicating that MALDI-TOF MS could also be used for *Mycobacterium* strain typing and clustering [95].

3.10 Application of MALDI-TOF MS in Clinical Virology

Clinical virology aim at (1) virus detection, (2) identification (including typing and epidemiology), and (3) resistance detection. For these purposes, the clinical virology laboratories rely on cell culture, electron microscopy, serology, and PCR with satisfactory results. Nevertheless, several groups have addressed the possibility of using MALDI-TOF MS in these different aspects of clinical virology, and when possible, they address the added value [96] (see Chapter 8).

The first approach is the detection of biomarker viral proteins from infected cell cultures. Using this strategy, Calderaro *et al.* reported the detection of *Picornaviridae* from infected cell culture using the viral protein VP4 as a biomarker [97].

La Scola *et al.* demonstrated that giant viruses (capsid sizes of 150 to 600 nm) could be characterized using MALDI-TOF MS analysis of viral particles. The results obtained correlated with results provided by the sequencing of the *polB* gene [98]. However, this method requires specific equipment and is time consuming (cell lysis, filtration, concentration by centrifugation, rinsing, ultracentrifugation), which makes routine use difficult.

Another approach applied for influenza viruses detection consisted in the capture of viral particles using magnetic nanoparticles functionalized with H5N2 specific antibodies with subsequent MALDI-TOF MS analysis of the complexes nanoparticles/viruses to detect the hemmaglutannin protein HA [99]. This allowed H5N2 detection with no cross reactivity with H5N1 viruses.

3.11 PCR-Mass Assay

MS technologies have been applied to the analysis of nucleic acid post amplification. This approach that combines PCR and MALDI-TOF MS, also known as PCR-Mass assay, allows the simultaneous detection of several amplicons at the same. In addition, this technology is sensitive enough to detect a single-nucleotide polymorphism or a nucleic acid modification such as methylation.

3.11.1 Application of PCR-Mass Assay in Clinical Bacteriology

The Sequenom MassARRAY platform (Sequenom) is an instrument dedicated to nucleic acid analysis using the PCR-Mass assay approach [100]. The amplification is achieved in the presence of deoxyuridine triphosphate (dUTP) instead of deoxythymidine triphosphate (dTTP). The amplicons are then fragmented with uracil-DNA-glycosylase, and the obtained fragments are analyzed by MALDI-TOF MS [101]. The mass pattern is then compared with a database or with a reference spectrum. Bacterial identification can be achieved by using PCR targeting the 16S rDNA [101]. This technique is also reliable for microbe typing or to detect mutations at the level of a single nucleotide that are associated with antibiotic resistance [102]. The sensitivity also allows detection of methylation variations [103].

This technology can be applied to the detection of uncultivable microorganisms directly from clinical samples as demonstrated for *Bordetella* species detection by targeting the 16S rRNA gene [101].

3.11.2 Application of PCR-Mass Assay in Clinical Virology

The PCR-Mass assay is an extremely sensitive and specific approach that has been applied to clinical virology for the detection, identification, and quantification of viruses as well as for the genotyping and the detection of drug resistance [96]. It is noteworthy that this technology allows simultaneous analysis of several amplicons. Using two multiplex PCRs followed by MALDI-TOF MS analysis of the amplicon, Sjoholm *et al.* could detect and identify any herpesviruses directly from various body fluid including bronchoalveolar lavage, conjunctival fluid, sore secretion, blister material, plasma, serum, and urine with an average concordance of 95.6% (86.4%–97.2%) with PCR results [104]. Piao *et al.* reported the simultaneous detection and identification of eight enteric viruses (enterovirus 71, coxsackievirus A16, reovirus, poliovirus, hepatitis E virus, norovirus, astrovirus, and hepatitis E virus) with a detection limit of 100–1000 copies per reaction [105]. Discrepant results during the validation of these methods have been associated with false negative results due to RNA degradation [105].

Mass assay technology was also demonstrated to be useful for the high-throughput diagnosis of HPV infections with simultaneous genotyping to differentiate between the high-risk and low-risk types [106,107]. The mass assay has been successfully used for the detection of Cytomegalovirus (CMV) resistance to ganciclovir in patients treated with this drug by detecting a single point mutation of the viral phosphotransferase (UL97) or the viral polymerase UL54 associated with ganciclovir resistance [108].

3.12 PCR-ESI MS

PCR electrospray ionization mass spectrometry (PCR-ESI MS) is a technology that associates DNA amplification by PCR and performs subsequent analysis of the obtained amplicon by MS [109,110]. Basically, nucleic acids extracted from clinical samples are amplified using multiple broad-range PCRs. Then, the precise molecular mass of the amplicon(s) is(are) determined by ESI/MS. This mass integrated to the amplicon length and to the DNA complementary rule provides the exact base composition (%A, %C, %G, and %T) of the amplicons. Pathogen identification is then achieved by comparison with a database [110,111]. This technology allows the detection of a broad range and a high number of pathogens: more than 750 bacteria and fungi, more than 200 fungi, and more than 130 viruses. The bacterial panel includes primers for the detection of the resistance genes *mecA*, bla_{KPC}, *vanA*, and *vanB*. Compared to current methods of clinical microbiology, PCR-ESI MS has a short TAT (~8 h) because (1) it can be applied directly on clinical samples without culture due to its high sensitivity and (2) the analysis of the amplicon is achieved by ESI-MS rather than by Sanger sequencing. PCR-ESI MS has been tested on various clinical samples such as CSF [112–114] and respiratory tract samples [115–117] with an increased sensitivity when compared to culture. The last platform developed allows diagnosis of bloodstream infections as it can accept up to 5 ml of sample and because of the optimization of the nucleic acid extraction and the amplification step [118]. This device, known as the PLEX-ID (Abbott Molecular, Des Plaines, IL), has recently been commercialized under the name Iridica. The Iridica system can detect more than 800 hundred pathogens directly from blood with a sensitivity of 50%–91% and a specificity of 98%–99% [50]. PCR-ESI/MS can detect polymicrobial infections and provide information on the abundance of each microorganism based on a semiquantitative analysis. The broad spectrum of this technology together with its extremely low detection limit makes it highly sensitive to contamination. Interventional studies will determine the exact clinical impact of this new technology, especially the impact on antibiotic stewardship.

3.13 Impact of MALDI-TOF MS in Clinical Microbiology and Infectious Disease

3.13.1 Time to Result

A challenge for clinical microbiology is to accelerate diagnosis when an infection is suspected, especially when most of the other laboratory results are obtained within the first hours of hospitalization [119]. Despite recent developments, especially in the domain of molecular diagnostic, culture-based methods remain predominant. Thus, MALDI-TOF MS dramatically reduces the time interval between the positive culture and the final identification when compared to subculture-based phenotypic identifications. In a large prospective study, Seng *et al.* estimated that MALDI-TOF MS identification took approximately 6 to 8.5 min when the time to result was 5 to 8 h for the VITEK system (BioMérieux, France) and 5 to 20 h for the Phoenix system (BD Diagnostic), which rely on phenotypic methods [28] (see Table 3.3). Tan *et al.* also reported a reduction of the time to result for a set of 20 bacteria, with a mean time gain

Table 3.3 Added value of MALDI-TOF MS in clinical microbiology.

	Comment	Selected references
Reduced time-to-results	Gain of 24 h and more over conventional methods.	[15,28]
Reduced cost	Cost-effective due to reduced cost of identification of a given isolate.	[15,120,134]
Improved therapeutic management of sepsis	Adaptation of the antibiotic treatment in ~35% of bacteremia involving gram-negative bacteria.	[34,121]
Increasing accuracy	Less than 2% of errors at species level.	[25,27,28,49]
Improved identification of specific pathogenic agents	Reduced need for 16S rRNA gene sequencing. Discovery of new pathogens.	[20,123,124,126]

of 1.45 d [120]. In particular, the reduction monitored by the authors was ~1.35 d for *S. aureus* and *Enterobacteriaceae* identification. The maximum gain was observed for gram-positive rods, with a reduction of 4.13 d (Table 3.3).

3.13.2 Impact on Patient Management

With a bloodstream infection, delaying the introduction of an efficient anti-infectious treatment directly impacts patient survival. Despite the emergence of culture-independent methods, blood cultures remain the gold standard to determine the etiologic agent of a bloodstream infection [34,50]. A prospective observational study performed in a tertiary hospital that evaluated the impact of MALDI-TOF MS on BC positive with gram-negative bacteria using a rapid bacterial pellet method reported that the empirical therapy was often inappropriate or too broad. In 35.1% of cases, the rapid identification led to a modification of the empirical therapy [121]. In this study, a correct identification at the genus level was obtained in 86.7% (143/165) of monomicrobial infections [121]. In their study, Martiny *et al.* demonstrated that MALDI-TOF MS was helpful in confirming suspected contamination in 37.50% of cases for pediatric patients [122].

3.13.3 Impact on Rare Pathogenic Bacteria and Difficult-to-Identify Organisms

Rare microorganisms can be defined as microorganisms with less than 10 reports designating them as human pathogens on the Pubmed database [123]. Alternatively, some organisms were difficult to identify accurately because of the phenotypic proximity with other organisms [124]. The identification of these pathogens is generally time and labor consuming as it often requires gene sequencing. MALDI-TOF MS allows us to report the first case of bacteremia caused by *Comamonas kerstersii*, a nonfermenting microorganism that was difficult to distinguish from other *Comamonas* species or closely related *Delftia* spp. in the pre-MALDI era [125]. For both rare microorganisms and difficult-to-identify microorganisms, MALDI-TOF MS significantly reduces the use of DNA sequencing as we reported in a study where we obtained nearly 50% of accurate identifications for this type of organism [126].

The accuracy of MALDI-TOF MS identification might reveal the real incidence and the pathogenic role of some rare/difficult-to-identify microorganisms [20,123,124].

3.13.4 Anaerobes

The diagnosis of anaerobic bacteria is generally a long process because of their slow growth, and its accuracy is often limited owing to their limited biochemical activity. In addition, anaerobes often occur in polymicrobial infections, resulting in multiple different species in the same sample. In a study conducted between 2010 and 2011 on 283 anaerobic bacteria, Nagy *et al.* reported 77% of correct identifications at the species level and 10.95% of correct identifications at the genus level using the MALDI-TOF MS [127] (see Chapter 5). Moreover, among 544 anaerobic bacteria isolated from clinical samples, La Scola *et al.* reported 61% of identifications at the species level using MALDI-TOF MS compared to 39% of successful identifications by 16S rRNA gene sequencing [128]. This demonstrated the potential of MALDI-TOF to reduce the number of time-consuming and expensive 16S rRNA gene sequencing for anaerobe identification as reported for rare microorganisms' identification (Table 3.3). Significantly, the rate of successful identification of anaerobes using MALDI-TOF MS, meaning the percentage of identifications that obtain a sufficient confidence score, is dependent on the representativeness of the database [129–132].

3.14 Identification of Protozoan Parasites

While waiting for confirmatory studies, several publications suggest that MALDI-TOF MS may be used for the identification of protozoa. The analysis of mass spectra obtained with intact *Giardia lamblia* and *Giardia muris* cysts identified common peaks as well as species-specific peaks that could be used for MALDI-TOF MS identification of these two microorganisms [136]. Cassagne *et al.* demonstrated that *Leishmania* promastigotes could successfully be identified at the species level from in vitro culture using a homemade reference mass spectra database comprising the main *Leishmania* species known to cause infection in humans. This strategy correctly identified correctly identified 66 of the 69 *Leishmania* promastigote isolates tested with log score values greater than 2. Two isolates failed to generate interpretable mass spectra, whereas one isolate identified a *Leishmania braziliensis* isolate as the closely related *Leishmania peruviana* isolate [137]. To evaluate the ability of MALDI-TOF MS to determine the subtype of enteric *Blastocystis*, Martiny *et al.* constructed a database of specific protein signatures of five *Blastocystis* subtypes. This approach gave correct subtype determination for 19 axenic cultures of various *Blastocystis* subtypes used to challenge the database [138]. By analyzing the MALDI-TOF MS spectra of *Entamoeba histolytica* and *Entamoeba dispar*, Calderaro *et al.* identified five peaks that could be used to discriminate the two species [139].

3.15 Identification of Ticks and Fleas

Yssouf *et al.* demonstrated that mass spectra obtain from leg-extracted hemolymph was reliable for the identification of ticks at the species level. The proof of concept was achieved using a homemade database made of five *Rickettsia*-free tick species (*Rhipicephalus sanguineus*, *Hyalomma marginatum rufipe*, *Ixodes ricinus*, *Dermacentor marginatus*, and *Dermacentor reticulatus*), one infected tick species (*Amblyomma variegatum* infected by *Rickettsia africae*), and other arthropods, including mosquitoes, lice, triatomines, and fleas [140]. Using a second database enriched in hemolymph from

Rickettsia africae infected ticks, the authors could reliably distinguish noninfected and infected specimens [140–142]. The same group also demonstrated that fleas could be identified based on MS generated from various body parts [143].

3.16 Costs

So far, the price and maintenance expenses of MALDI-TOF MS instruments have been high. Nevertheless, MALDI-TOF MS has been shown to be inexpensive in comparison with phenotypic and genotypic methods of identification. Indeed, the cost for the identification of a given isolate can be less than 1.5 € per identification using MALDI-TOF MS versus 5.9–8.23 € for the VITEK system [28,133]. A prospective costs analysis study compared the MALDI-TOF MS protocol with the standard identification protocol. For this study, the supplementary tests necessary for the identification of some bacteria such as *Streptococcus pneumoniae* and *Shigella*, the additional cost due to repeated MALDI-TOF MS analyses, the instrument maintenance expenses, and microorganism prevalence were included [120]. By integrating all these data, the authors anticipated a 56% of reduction of reagent and labor costs for one year. In another retrospective cost assessment study after MALDI-TOF MS implementation, an overall 89.3% cost saving was obtained [134]. This group also mentioned an important decrease in waste disposal and a reduction in subculture medium expense. As we also reported, MALDI-TOF MS has the potential to reduce the expense due to DNA sequencing for some difficult-to-identify isolates [126], (Table 3.3). In another prospective economical study, we reported that microbial identification was 2.43-fold less expensive with MALDI-TOF MS than with standard methods. However, this ratio varied from 0.70 to 7.0 according to the bacterial species. Some identifications (for example, urinary *Escherichia coli* identified on chromogenic culture media and confirmed by a simple spot indole test) were cheaper with standard identification than with MALDI-TOF MS. The cost savings will thus depend on the epidemiology and the prevalence of each species encountered in the clinical laboratory and the laboratory habits and procedures (Heiniger *et al.*, unpublished data). Using a mathematical model, it was calculated that implementation of MALDI-TOF equipment and identification would be cost-effective if more than 5300 to 8000 strains were identified yearly. With more than 20,000 identifications/year, Tran *et al.* estimated that the initial cost of the instrument would be recovered after about three years [120]. Martiny *et al.* have shown that the implementation of the MALDI-TOF MS technology is an opportunity for the mutualization of processes such as analytical platforms with important cost savings [135]. Further cost savings may occur in the near future with the automation of the preparation of MALDI-TOF target plates.

3.17 Conclusions

MALDI-TOF MS is one of the major revolutions that have occurred in clinical microbiology laboratory in the last decades, nearly rendering obsolete all biochemical identification galleries, because more than 95% of all bacterial identification are nowadays performed using MALDI-TOF MS. With the exception of some body fluids such as

urine, MALDI-TOF MS is still dependent on positive culture (agar plate or blood culture, for instance) because more than 10^4 cells (approximately 10^7 cells/ml) are required to generate a good-quality spectrum suitable for identification [144]. Further developments such as microorganism enrichment by affinity or an increasing in the sensitivity limit of MALDI-TOF MS might make it reliable for identification directly from samples. In the meantime, new approaches such as PCR-MALDI-TOF MS and PCR-ESI MS will extend mass spectrometry applications in clinical diagnostic microbiology as these nucleic-amplification-based methods are suitable for diagnosis directly from clinical samples including blood [50]. MALDI-TOF MS has also proved its reliability for other applications such as antibiotic resistance detection and typing; however, most of the procedures developed so far are still difficult to achieve on a routine basis and are restricted to specialized laboratories. Because MALDI-TOF MS is reliable at determining bacterial biomass, one could expect that in the future this could represents a reliable read-out for rapid AST based on the impact of an antibiotic on bacterial growth. As it is an innovation with increased accuracy and undisputed convenience impacting the identification of rare microorganisms and allowing the discrimination between closely related organisms, MALDI-TOF MS might reveal the real incidence and the pathogenic role of some organisms.

References

1 Osborn, T. M., Phillips, G., Lemeshow, S., Townsend, S., Schorr, C. A. *et al.* 2014. Sepsis severity score: An internationally derived scoring system from the surviving sepsis campaign database. *Critical Care Medicine* 42(9), 1969–1976.

2 Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H. *et al.* 2013. Surviving sepsis campaign: International guidelines for management of severe sepsis and septic shock: 2012. *Critical Care Medicine* 41: 580–637.

3 Leibovici, L., Shraga, I., Drucker, M., Konigsberger, H., Samra, Z., and Pitlik, S. D. 1998. The benefit of appropriate empirical antibiotic treatment in patients with bloodstream infection. *Journal of Internal Medicine* 244: 379–386.

4 Zaragoza, R., Artero, A., Camarena, J. J., Sancho, S., Gonzalez, R., and Nogueira, J. M. 2003. The influence of inadequate empirical antimicrobial treatment on patients with bloodstream infections in an intensive care unit. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 9: 412–418.

5 Wang, Z., Russon, L., Li, L., Roser, D. C., and Long, S. R. 1998. Investigation of spectral reproducibility in direct analysis of bacteria proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM* 12: 456–464.

6 Claydon, M. A., Davey, S. N., Edwards-Jones, V., and Gordon, D. B. 1996. The rapid identification of intact microorganisms using mass spectrometry. *Nature Biotechnology* 14: 1584–1586.

7 Holland, R. D., Wilkes, J. G., Rafii, F., Sutherland, J. B., Persons, C. C. *et al.* 1996. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM* 10: 1227–1232.

8 Anhalt, J. P., and Fenselau, C. 1975. Identification of bacteria using mass spectrometry. *Analytical Chemistry* 47: 219–225.

9 Krishna Murthy, H. M. 1996. Use of multiple-wavelength anomalous diffraction measurements in ab initio phase determination for macromolecular structures. *Methods in Molecular Biology* 56: 127–151.

10 Haag, A. M., Taylor, S. N., Johnston, K. H., and Cole, R. B. 1998. Rapid identification and speciation of *Haemophilus* bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Mass Spectrometry: JMS* 33: 750–756.

11 Suarez, S., Ferroni, A., Lotz, A., Jolley, K. A., Guerin, P. *et al.* 2013. Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. *Journal of Microbiological Methods* 94: 390–396.

12 Suh, M. J., and Limbach, P. A. 2004. Investigation of methods suitable for the matrix-assisted laser desorption/ionization mass spectrometric analysis of proteins from ribonucleoprotein complexes. *European Journal of Mass Spectrometry* 10: 89–99.

13 Richter, S. S., Sercia, L., Branda, J. A., Burnham, C. A., Bythrow, M. *et al.* 2013. Identification of Enterobacteriaceae by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using the VITEK MS system. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 32: 1571–1578.

14 Rychert, J., Burnham, C. A., Bythrow, M., Garner, O. B., Ginocchio, C. C. *et al.* 2013. Multicenter evaluation of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system for identification of Gram-positive aerobic bacteria. *Journal of Clinical Microbiology* 51: 2225–2231.

15 Cherkaoui, A., Hibbs, J., Emonet, S., Tangomo, M., Girard, M. *et al.* 2010. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *Journal of Clinical Microbiology* 48: 1169–1175.

16 Martiny, D., Busson, L., Wybo, I., El Haj, R. A., Dediste, A., and Vandenberg, O. 2012. Comparison of the Microflex LT and Vitek MS systems for routine identification of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 50: 1313–1325.

17 Jamal, W. Y., Ahmad, S., Khan, Z. U., and Rotimi, V. O. 2014. Comparative evaluation of two matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) systems for the identification of clinically significant yeasts. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases* 26: 167–170.

18 Lee, M., Chung, H. S., Moon, H. W., Lee, S. H., and Lee, K. 2015. Comparative evaluation of two matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) systems, Vitek MS and Microflex LT, for the identification of Gram-positive cocci routinely isolated in clinical microbiology laboratories. *Journal of Microbiological Methods* 113: 13–15.

19 Deak, E., Charlton, C. L., Bobenchik, A. M., Miller, S. A., Pollett, S. *et al.* 2015. Comparison of the Vitek MS and Bruker Microflex LT MALDI-TOF MS platforms for routine identification of commonly isolated bacteria and yeast in the clinical microbiology laboratory. *Diagnostic Microbiology and Infectious Disease* 81: 27–33.

20 McElvania TeKippe, E., and Burnham, C. A. 2014. Evaluation of the Bruker Biotyper and VITEK MS MALDI-TOF MS systems for the identification of unusual and/or

difficult-to-identify microorganisms isolated from clinical specimens. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 33: 2163–2171.

21 Jamal, W., Albert, M. J., and Rotimi, V. O. 2014. Real-time comparative evaluation of bioMérieux VITEK MS versus Bruker Microflex MS, two matrix-assisted laser desorption-ionization time-of-flight mass spectrometry systems, for identification of clinically significant bacteria. *BMC Microbiology* 14: 289.

22 Bille, E., Dauphin, B., Leto, J., Bougnoux, M. E., Beretti, J. L. *et al.* 2012. MALDI-TOF MS Andromas strategy for the routine identification of bacteria, mycobacteria, yeasts, *Aspergillus* spp. and positive blood cultures. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 18: 1117–1125.

23 Farfour, E., Leto, J., Barritault, M., Barberis, C., Meyer, J. *et al.* 2012. Evaluation of the Andromas matrix-assisted laser desorption ionization-time of flight mass spectrometry system for identification of aerobically growing Gram-positive bacilli. *Journal of Clinical Microbiology* 50: 2702–2707.

24 Lacroix, C., Gicquel, A., Sendid, B., Meyer, J., Accoceberry, I. *et al.* 2014. Evaluation of two matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) systems for the identification of *Candida* species. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 20: 153–158.

25 Bizzini, A., Durussel, C., Bille, J., Greub, G., and Prod'hom, G. 2010. Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. *Journal of Clinical Microbiology* 48: 1549–1554.

26 Croxatto, A., Prod'hom, G., Faverjon, F., Rochais, Y., and Greub, G. 2016. Laboratory automation in clinical bacteriology: what system to choose? *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 22: 217–235.

27 Croxatto, A., Prod'hom, G., and Greub, G. 2012. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiology Reviews* 36: 380–407.

28 Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P. E. *et al.* 2009. Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 49: 543–551.

29 Wybo, I., Soetens, O., De Bel, A., Echahidi, F., Vancutsem, E. *et al.* 2012. Species identification of clinical Prevotella isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 50: 1415–1418.

30 Schmitt, B. H., Cunningham, S. A., Dailey, A. L., Gustafson, D. R., and Patel, R. 2013. Identification of anaerobic bacteria by Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry with on-plate formic acid preparation. *Journal of Clinical Microbiology* 51: 782–786.

31 Holzmann, T., Frangoulidis, D., Simon, M., Noll, P., Schmoldt, S. *et al.* 2012. Fatal anthrax infection in a heroin user from southern Germany, June 2012. *Euro Surveillance: Bulletin Europeen sur les maladies transmissibles = European Communicable Disease Bulletin* 17(26), pii=20204.

32 Clark, A. E., Kaleta, E. J., Arora, A., and Wolk, D. M. 2013. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: A fundamental shift in the routine practice of clinical microbiology. *Clinical Microbiology Reviews* 26: 547–603.

33 Buckwalter, S. P., Olson, S. L., Connelly, B. J., Lucas, B. C., Rodning, A. A. *et al.* 2016. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of *Mycobacterium* species, *Nocardia* species, and other aerobic actinomycetes. *Journal of Clinical Microbiology* 54: 376–384.

34 Opota, O., Croxatto, A., Prod'hom, G., and Greub, G. 2015. Blood culture-based diagnosis of bacteraemia: State of the art. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 21: 313–322.

35 Moussaoui, W., Jaulhac, B., Hoffmann, A. M., Ludes, B., Kostrzewa, M. *et al.* 2010. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry identifies 90% of bacteria directly from blood culture vials. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 16: 1631–1638.

36 Stevenson, L. G., Drake, S. K., and Murray, P. R. 2010. Rapid identification of bacteria in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 48: 444–447.

37 Kok, J., Thomas, L. C., Olma, T., Chen, S. C., and Iredell, J. R. 2011. Identification of bacteria in blood culture broths using matrix-assisted laser desorption-ionization Sepsityper and time of flight mass spectrometry. *PLoS ONE* 6: e23285.

38 Dubois, D., Segonds, C., Prere, M. F., Marty, N., and Oswald, E. 2013. Identification of clinical Streptococcus pneumoniae isolates among other alpha and nonhemolytic streptococci by use of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system. *Journal of Clinical Microbiology* 51: 1861–1867.

39 Ikryannikova, L. N., Filimonova, A. V., Malakhova, M. V., Savinova, T., Filimonova, O. *et al.* 2013. Discrimination between *Streptococcus pneumoniae* and *Streptococcus mitis* based on sorting of their MALDI mass spectra. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 19: 1066–1071.

40 Sullivan, K. V., Turner, N. N., Roundtree, S. S., Young, S., Brock-Haag, C. A. *et al.* 2013. Rapid detection of Gram-positive organisms by use of the Verigene Gram-positive blood culture nucleic acid test and the BacT/Alert Pediatric FAN system in a multicenter pediatric evaluation. *Journal of Clinical Microbiology* 51: 3579–3584.

41 Khot, P. D., and Fisher, M. A. 2013. Novel approach for differentiating *Shigella* species and *Escherichia coli* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 51: 3711–3716.

42 Schaumann, R., Knoop, N., Genzel, G. H., Losensky, K., Rosenkranz, C. *et al.* 2013. Discrimination of Enterobacteriaceae and non-fermenting Gram negative bacilli by MALDI-TOF mass spectrometry. *The Open Microbiology Journal* 7: 118–122.

43 Deng, J., Fu, L., Wang, R., Yu, N., Ding, X. *et al.* 2014. Comparison of MALDI-TOF MS, gene sequencing and the Vitek 2 for identification of seventy-three clinical isolates of enteropathogens. *Journal of Thoracic Disease* 6: 539–544.

44 de Jong, E., de Jong, A. S., Smidts-van den Berg, N., and Rentenaar, R. J. 2013. Differentiation of *Raoultella ornithinolytica/planticola* and *Klebsiella oxytoca* clinical isolates by matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Diagnostic Microbiology and Infectious Disease* 75: 431–433.

45 De Bel, A., Wybo, I., Vandoorslaer, K., Rosseel, P., Lauwers, S., and Pierard, D. 2011. Acceptance criteria for identification results of Gram-negative rods by mass spectrometry. *Journal of Medical Microbiology* 60: 684–686.

46 Fenselau, C., and Demirev, P. A. 2001. Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrometry Reviews* 20: 157–171.

47 Kim, Y. J., Freas, A., and Fenselau, C. 2001. Analysis of viral glycoproteins by MALDI-TOF mass spectrometry. *Analytical Chemistry* 73: 1544–1548.

48 Evason, D. J., Claydon, M. A., and Gordon, D. B. 2000. Effects of ion mode and matrix additives in the identification of bacteria by intact cell mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM* 14: 669–672.

49 Dubois, D., Grare, M., Prere, M. F., Segonds, C., Marty, N., and Oswald, E. 2012. Performances of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system for rapid identification of bacteria in routine clinical microbiology. *Journal of Clinical Microbiology* 50: 2568–2576.

50 Opota, O., Jaton, K., and Greub, G. 2015. Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 21: 323–331.

51 Idelevich, E. A., Schule, I., Grunastel, B., Wullenweber, J., Peters, G., and Becker, K. 2014. Rapid identification of microorganisms from positive blood cultures by MALDI-TOF mass spectrometry subsequent to very short-term incubation on solid medium. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 20: 1001–1006.

52 Verroken, A., Defourny, L., Lechgar, L., Magnette, A., Delmee, M., and Glupczynski, Y. 2014. Reducing time to identification of positive blood cultures with MALDI-TOF MS analysis after a 5-h subculture. *European Journal of Clinical Microbiology & Infectious Diseases* 34(2): 405–413.

53 Prod'hom, G., Bizzini, A., Durussel, C., Bille, J., and Greub, G. 2010. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for direct bacterial identification from positive blood culture pellets. *Journal of Clinical Microbiology* 48: 1481–1483.

54 Croxatto, A., Prod'hom, G., Durussel, C., and Greub, G. 2015. Preparation of a blood culture pellet for rapid bacterial identification and antibiotic susceptibility testing. *Journal of Visualized Experiments* 92: e51985.

55 Yan, Y., He, Y., Maier, T., Quinn, C., Shi, G. *et al.* 2011. Improved identification of yeast species directly from positive blood culture media by combining Sepsityper specimen processing and Microflex analysis with the matrix-assisted laser desorption ionization Biotyper system. *Journal of Clinical Microbiology* 49: 2528–2532.

56 Buchan, B. W., Riebe, K. M., and Ledeboer, N. A. 2012. Comparison of the MALDI Biotyper system using Sepsityper specimen processing to routine microbiological methods for identification of bacteria from positive blood culture bottles. *Journal of Clinical Microbiology* 50: 346–352.

57 Foster, A. G. 2013. Rapid Identification of microbes in positive blood cultures by use of the vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system. *Journal of Clinical Microbiology* 51: 3717–3719.

58 Prod'hom, G., Durussel, C., and Greub, G. 2013. A simple blood-culture bacterial pellet preparation for faster accurate direct bacterial identification and antibiotic susceptibility testing with the VITEK 2 system. *Journal of Medical Microbiology* 62: 773–777.

59 Hazelton, B., Thomas, L. C., Olma, T., Kok, J., O'Sullivan, M. *et al.* 2014. Rapid and accurate direct antibiotic susceptibility testing of blood culture broths using MALDI Sepsityper combined with the BD Phoenix automated system. *Journal of Medical Microbiology* 63: 1590–1594.

60 Machen, A., Drake, T., and Wang, Y. F. 2014. Same day identification and full panel antimicrobial susceptibility testing of bacteria from positive blood culture bottles made possible by a combined lysis-filtration method with MALDI-TOF VITEK mass spectrometry and the VITEK2 system. *PLoS ONE* 9: e87870.

61 Ferreira, L., Sanchez-Juanes, F., Gonzalez-Avila, M., Cembrero-Fucinos, D., Herrero-Hernandez, A. *et al.* 2010. Direct identification of urinary tract pathogens from urine samples by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 48: 2110–2115.

62 March Rossello, G. A., Gutierrez Rodriguez, M. P., de Lejarazu Leonardo, R. O., Orduna Domingo, A., and Bratos Perez, M. A. 2014. Procedure for microbial identification based on matrix-assisted laser desorption/ionization-time of flight mass spectrometry from screening-positive urine samples. *APMIS: Acta pathologica, microbiologica, et immunologica Scandinavica* 122: 790–795.

63 Kohling, H. L., Bittner, A., Muller, K. D., Buer, J., Becker, M. *et al.* 2012. Direct identification of bacteria in urine samples by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and relevance of defensins as interfering factors. *Journal of Medical Microbiology* 61: 339–344.

64 Wang, X. H., Zhang, G., Fan, Y. Y., Yang, X., Sui, W. J., and Lu, X. X. 2013. Direct identification of bacteria causing urinary tract infections by combining matrix-assisted laser desorption ionization-time of flight mass spectrometry with UF-1000i urine flow cytometry. *Journal of Microbiological Methods* 92: 231–235.

65 Nyvang Hartmeyer, G., Kvistholm Jensen, A., Bocher, S., Damkjaer Bartels, M., Pedersen, M. *et al.* 2010. Mass spectrometry: Pneumococcal meningitis verified and *Brucella* species identified in less than half an hour. *Scandinavian Journal of Infectious Diseases* 42: 716–718.

66 Verroken, A., Janssens, M., Berhin, C., Bogaerts, P., Huang, T. D. *et al.* 2010. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of nocardia species. *Journal of Clinical Microbiology* 48: 4015–4021.

67 Wilen, C. B., McMullen, A. R., and Burnham, C. A. 2015. Comparison of sample preparation methods, instrumentation platforms, and contemporary commercial databases for identification of clinically relevant mycobacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 53: 2308–2315.

68 Kodana, M., Tarumoto, N., Kawamura, T., Saito, T., Ohno, H. *et al.* 2016. Utility of the MALDI-TOF MS method to identify nontuberculous mycobacteria. *Journal of Infection and Chemotherapy: Official Journal of the Japan Society of Chemotherapy* 22: 32–35.

69 Machen, A., Kobayashi, M., Connelly, M. R., and Wang, Y. F. 2013. Comparison of heat inactivation and cell disruption protocols for identification of mycobacteria from solid culture media by use of vitek matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 51: 4226–4229.

70 El Khechine, A., Couderc, C., Flaudrops, C., Raoult, D., and Drancourt, M. 2011. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of mycobacteria in routine clinical practice. *PLoS ONE* 6: e24720.

71 Lotz, A., Ferroni, A., Beretti, J. L., Dauphin, B., Carbonnelle, E. *et al.* 2010. Rapid identification of mycobacterial whole cells in solid and liquid culture media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 48: 4481–4486.

72 Mancini, N., De Carolis, E., Infurnari, L., Vella, A., Clementi, N. *et al.* 2013. Comparative evaluation of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry systems for identification of yeasts of medical importance. *Journal of Clinical Microbiology* 51: 2453–2457.

73 Stevenson, L. G., Drake, S. K., Shea, Y. R., Zelazny, A. M., and Murray, P. R. 2010. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of clinically important yeast species. *Journal of Clinical Microbiology* 48: 3482–3486.

74 Goyer, M., Lucchi, G., Ducoroy, P., Vagner, O., Bonnin, A., and Dalle, F. 2012. Optimization of the preanalytical steps of matrix-assisted laser desorption ionization-time of flight mass spectrometry identification provides a flexible and efficient tool for identification of clinical yeast isolates in medical laboratories. *Journal of Clinical Microbiology* 50: 3066–3068.

75 De Carolis, E., Posteraro, B., Lass-Florl, C., Vella, A., Florio, A. R. *et al.* 2012. Species identification of *Aspergillus*, *Fusarium* and *Mucorales* with direct surface analysis by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 18: 475–484.

76 De Respinis, S., Monnin, V., Girard, V., Welker, M., Arsac, M. *et al.* 2014. Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry using the Vitek MS system for rapid and accurate identification of dermatophytes on solid cultures. *Journal of Clinical Microbiology* 52: 4286–4292.

77 Wang, H., Fan, Y. Y., Kudinha, T., Xu, Z. P., Xiao, M. *et al.* 2016. A comprehensive evaluation of Bruker Biotyper MS and Vitek MS matrix-assisted laser desorption/ionization time-of-flight mass spectrometry systems for the identification of yeasts—part of the National China Hospital Invasive Fungal Surveillance Net (CHIF-NET) Study, 2012–2013. *Journal of Clinical Microbiology.*

78 Kempf, M., Bakour, S., Flaudrops, C., Berrazeg, M., Brunel, J. M. *et al.* 2012. Rapid detection of carbapenem resistance in *Acinetobacter baumannii* using matrix-assisted laser desorption ionization-time of flight mass spectrometry. *PLoS ONE* 7: e31676.

79 Carvalhaes, C. G., Cayo, R., Visconde, M. F., Barone, T., Frigatto, E. A. *et al.* 2014. Detection of carbapenemase activity directly from blood culture vials using MALDI-TOF MS: a quick answer for the right decision. *The Journal of Antimicrobial Chemotherapy* 69: 2132–2136.

80 Vogne, C., Prod'Hom, G., Jaton, K., Decosterd, L., and Greub, G. 2014. A simple, robust and rapid approach to detect carbapenemases in Gram negative isolates by MALDI-TOF mass spectrometry: Validation with triple quadripole tandem mass spectrometry, microarray and PCR. *Clinical Microbiology and Infection* 20(12): 1106–1112.

81 Lu, J. J., Tsai, F. J., Ho, C. M., Liu, Y. C., and Chen, C. J. 2012. Peptide biomarker discovery for identification of methicillin-resistant and vancomycin-intermediate *Staphylococcus aureus* strains by MALDI-TOF. *Analytical Chemistry* 84: 5685–5692.

82 Wolters, M., Rohde, H., Maier, T., Belmar-Campos, C., Franke, G. *et al.* 2011. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. *International Journal of Medical Microbiology: IJMM* 301: 64–68.

83 Edwards-Jones, V., Claydon, M. A., Evason, D. J., Walker, J., Fox, A. J., and Gordon, D. B. 2000. Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. *Journal of Medical Microbiology* 49: 295–300.

84 Griffin, P. M., Price, G. R., Schooneveldt, J. M., Schlebusch, S., Tilse, M. H. *et al.* 2012. Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *Journal of Clinical Microbiology* 50: 2918–2931.

85 Gagnaire, J., Dauwalder, O., Boisset, S., Khau, D., Freydiere, A. M. *et al.* 2012. Detection of *Staphylococcus aureus* delta-toxin production by whole-cell MALDI-TOF mass spectrometry. *PLoS ONE* 7: e40660.

86 Arya, R., and Princy, S. A. 2013. An insight into pleiotropic regulators Agr and Sar: Molecular probes paving the new way for antivirulent therapy. *Future Microbiology* 8: 1339–1353.

87 Traber, K. E., Lee, E., Benson, S., Corrigan, R., Cantera, M. *et al.* 2008. Agr function in clinical *Staphylococcus aureus* isolates. *Microbiology* 154: 2265–2274.

88 Bittar, F., Ouchenane, Z., Smati, F., Raoult, D., and Rolain, J. M. 2009. MALDI-TOF-MS for rapid detection of staphylococcal Panton–Valentine leukocidin. *International Journal of Antimicrobial Agents* 34: 467–470.

89 Szabados, F., Becker, K., von Eiff, C., Kaase, M., and Gatermann, S. 2011. The matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS)-based protein peaks of 4448 and 5302 Da are not associated with the presence of Panton–Valentine leukocidin. *International Journal of Medical Microbiology: IJMM* 301: 58–63.

90 Konrad, R., Berger, A., Huber, I., Boschert, V., Hormansdorfer, S. *et al.* 2010. Matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry as a tool for rapid diagnosis of potentially toxigenic *Corynebacterium* species in the laboratory management of diphtheria-associated bacteria. *Euro Surveillance: Bulletin Europeen sur les maladies transmissibles = European Communicable Disease Bulletin* 15(43), pii=19699.

91 Josten, M., Reif, M., Szekat, C., Al-Sabti, N., Roemer, T. *et al.* 2013. Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of *Staphylococcus aureus* identifies mutations that allow differentiation of the main clonal lineages. *Journal of Clinical Microbiology* 51: 1809–1817.

92 Christner, M., Trusch, M., Rohde, H., Kwiatkowski, M., Schluter, H. *et al.* 2014. Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-Toxigenic *Escherichia coli*. *PLoS ONE* 9: e101924.

93 Stephan, R., Cernela, N., Ziegler, D., Pfluger, V., Tonolla, M. *et al.* 2011. Rapid species specific identification and subtyping of *Yersinia enterocolitica* by MALDI-TOF mass spectrometry. *Journal of Microbiological Methods* 87: 150–153.

94 Dieckmann, R., and Malorny, B. 2011. Rapid screening of epidemiologically important *Salmonella enterica* subsp. *enterica* serovars by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Applied and Environmental Microbiology* 77: 4136–4146.

95 Suzuki, H., Yoshida, S., Yoshida, A., Okuzumi, K., Fukusima, A., and Hishinuma, A. 2015. A novel cluster of *Mycobacterium abscessus* complex revealed by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Diagnostic Microbiology and Infectious Disease* 83: 365–370.

96 Cobo, F. 2013. Application of MALDI-TOF mass spectrometry in clinical virology: A review. *The Open Virology Journal* 7: 84–90.

97 Calderaro, A., Arcangeletti, M. C., Rodighiero, I., Buttrini, M., Gorrini, C. *et al.* 2014. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry applied to virus identification. *Scientific Reports* 4: 6803.

98 La Scola, B., Campocasso, A., N'Dong, R., Fournous, G., Barrassi, L. *et al.* 2010. Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. *Intervirology* 53: 344–353.

99 Chou, T. C., Hsu, W., Wang, C. H., Chen, Y. J., and Fang, J. M. 2011. Rapid and specific influenza virus detection by functionalized magnetic nanoparticles and mass spectrometry. *Journal of Nanobiotechnology* 9: 52.

100 Gabriel, S., and Ziaugra, L. 2004. SNP genotyping using Sequenom MassARRAY 7K platform. *Current protocols in human genetics editorial board, Jonathan L. Haines … [et al.]* Chapter 2: Unit 2 12.

101 von Wintzingerode, F., Bocker, S., Schlotelburg, C., Chiu, N. H., Storm, N. *et al.* 2002. Base-specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: A tool for rapid bacterial identification. *Proceedings of the National Academy of Sciences of the United States of America* 99: 7039–7044.

102 Stanssens, P., Zabeau, M., Meersseman, G., Remes, G., Gansemans, Y. *et al.* 2004. High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Research* 14: 126–133.

103 Ehrich, M., Nelson, M. R., Stanssens, P., Zabeau, M., Liloglou, T. *et al.* 2005. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15785–15790.

104 Sjoholm, M. I., Dillner, J., and Carlson, J. 2008. Multiplex detection of human herpesviruses from archival specimens by using matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 46: 540–545.

105 Piao, J., Jiang, J., Xu, B., Wang, X., Guan, Y. *et al.* 2012. Simultaneous detection and identification of enteric viruses by PCR-mass assay. *PLoS ONE* 7: e42251.

106 Du, H., Yi, J., Wu, R., Belinson, S. E., Qu, X. *et al.* 2011. A new PCR-based mass spectrometry system for high-risk HPV, part II: Clinical trial. *American Journal of Clinical Pathology* 136: 920–923.

107 Yi, X., Zou, J., Xu, J., Liu, T., Liu, T. *et al.* 2014. Development and validation of a new HPV genotyping assay based on next-generation sequencing. *American Journal of Clinical Pathology* 141: 796–804.

108 Zurcher, S., Mooser, C., Luthi, A. U., Muhlemann, K., Barbani, M. T. *et al.* 2012. Sensitive and rapid detection of ganciclovir resistance by PCR based MALDI-TOF analysis. *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology* 54: 359–363.

109 Ecker, D. J., Sampath, R., Li, H., Massire, C., Matthews, H. E. *et al.* 2010. New technology for rapid molecular diagnosis of bloodstream infections. *Expert Review of Molecular Diagnostics* 10: 399–415.

110 Jordana-Lluch, E., Gimenez, M., Quesada, M. D., Ausina, V., and Martro, E. 2014. Improving the diagnosis of bloodstream infections: PCR coupled with mass spectrometry. *BioMed Research International* 2014: 501214.

111 Ecker, D. J., Sampath, R., Massire, C., Blyn, L. B., Hall, T. A. *et al.* 2008. Ibis T5000: A universal biosensor approach for microbiology. *Nature Reviews Microbiology* 6: 553–558.

112 Farrell, J. J., Sampath, R., Ecker, D. J., and Bonomo, R. A. 2013. "Salvage microbiology": detection of bacteria directly from clinical specimens following initiation of antimicrobial treatment. *PLoS ONE* 8: e66349.

113 Nagalingam, S., Lisgaris, M., Rodriguez, B., Jacobs, M. R., Lederman, M. *et al.* 2014. Identification of occult *Fusobacterium nucleatum* central nervous system infection using PCR electrospray ionization mass spectrometry (PCR/ESI-MS). *Journal of Clinical Microbiology* 52: 3462–3464.

114 Farrell, J. J., Tsung, A. J., Flier, L., Martinez, D. L., Beam, S. B. *et al.* 2013. PCR and electrospray ionization mass spectrometry for detection of persistent enterococcus faecalis in cerebrospinal fluid following treatment of postoperative ventriculitis. *Journal of Clinical Microbiology* 51: 3464–3466.

115 Gariani, K., Rougemont, M., Renzi, G., Hibbs, J., Emonet, S., and Schrenzel, J. 2014. Fulminant atypical *Cryptococcus neoformans* pneumonia confirmed by PLEX-ID. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases* 22: 17–18.

116 Jacobs, M. R., Bajaksouzian, S., Bonomo, R. A., Good, C. E., Windau, A. R. *et al.* 2009. Occurrence, distribution, and origins of *Streptococcus pneumoniae* Serotype 6C, a recently recognized serotype. *Journal of Clinical Microbiology* 47: 64–72.

117 Ecker, D. J., Sampath, R., Blyn, L. B., Eshoo, M. W., Ivy, C. *et al.* 2005. Rapid identification and strain-typing of respiratory pathogens for epidemic surveillance. *Proceedings of the National Academy of Sciences of the United States of America* 102: 8012–8017.

118 Bacconi, A., Richmond, G. S., Baroldi, M. A., Laffler, T. G., Blyn, L. B. *et al.* 2014. Improved sensitivity for molecular detection of bacteria and Candida in blood. *Journal of Clinical Microbiology* 52: 3164–3174.

119 Bissonnette, L., and Bergeron, M. G. 2010. Diagnosing infections—current and anticipated technologies for point-of-care diagnostics and home-based testing. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 16: 1044–1053.

120 Tan, K. E., Ellis, B. C., Lee, R., Stamper, P. D., Zhang, S. X., and Carroll, K. C. 2012. Prospective evaluation of a matrix-assisted laser desorption ionization-time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: a bench-by-bench study for assessing the impact on time to identification and cost-effectiveness. *Journal of Clinical Microbiology* 50: 3301–3308.

121 Clerc, O., Prod'hom, G., Vogne, C., Bizzini, A., Calandra, T., and Greub, G. 2013. Impact of matrix-assisted laser desorption ionization time-of-flight mass spectrometry on the clinical management of patients with Gram-negative bacteremia: A prospective observational study. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 56: 1101–1107.

122 Martiny, D., Debaugnies, F., Gateff, D., Gerard, M., Aoun, M. *et al.* 2013. Impact of rapid microbial identification directly from positive blood cultures using

matrix-assisted laser desorption/ionization time-of-flight mass spectrometry on patient management. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 19: E568–81.

123 Seng, P., Abat, C., Rolain, J. M., Colson, P., Lagier, J. C. *et al.* 2013. Identification of rare pathogenic bacteria in a clinical microbiology laboratory: impact of matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of Clinical Microbiology* 51: 2182–2194.

124 Opota, O., Prod'hom, G., Andreutti-Zaugg, C., Dessauges, M., Merz, L. *et al.* 2015. Diagnosis of *Aerococcus urinae* infections: Importance of MALDI-TOF MS and broad-range 16S rDNA PCR. *Clinical Microbiology and Infection* 22(1), e1–e2.

125 Opota, O., Ney, B., Zanetti, G., Jaton, K., Greub, G., and Prod'hom G. 2014. Bacteremia caused by *Comamonas kerstersii* in a patient with diverticulosis. *Journal of Clinical Microbiology* 52: 1009–1012.

126 Bizzini, A., Jaton, K., Romo, D., Bille, J., Prod'hom, G., and Greub, G. 2011. Matrix-assisted laser desorption ionization-time of flight mass spectrometry as an alternative to 16S rRNA gene sequencing for identification of difficult-to-identify bacterial strains. *Journal of Clinical Microbiology* 49: 693–696.

127 Nagy, E., Becker, S., Kostrzewa, M., Barta, N., and Urban, E. 2012. The value of MALDI-TOF MS for the identification of clinically relevant anaerobic bacteria in routine laboratories. *Journal of Medical Microbiology* 61: 1393–1400.

128 La Scola, B., Fournier, P. E., and Raoult, D. 2011. Burden of emerging anaerobes in the MALDI-TOF and 16S rRNA gene sequencing era. *Anaerobe* 17: 106–112.

129 Veloo, A. C., Erhard, M., Welker, M., Welling, G. W., and Degener, J. E. 2011. Identification of Gram-positive anaerobic cocci by MALDI-TOF mass spectrometry. *Systematic and Applied Microbiology* 34: 58–62.

130 Veloo, A. C., Knoester, M., Degener, J. E., and Kuijper, E. J. 2011. Comparison of two matrix-assisted laser desorption ionisation-time of flight mass spectrometry methods for the identification of clinically relevant anaerobic bacteria. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 17: 1501–1506.

131 Jamal, W. Y., Shahin, M., Rotimi, V. O. 2013. Comparison of two matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry methods and API 20AN for identification of clinically relevant anaerobic bacteria. *Journal of Medical Microbiology* 62: 540–544.

132 Garner, O., Mochon, A., Branda, J., Burnham, C. A., Bythrow, M. *et al.* 2014. Multi-centre evaluation of mass spectrometric identification of anaerobic bacteria using the VITEK(R) MS system. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 20: 335–339.

133 Cherkaoui, A., Emonet, S., Fernandez, J., Schorderet, D., and Schrenzel, J. 2011. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of Beta-hemolytic streptococci. *Journal of Clinical Microbiology* 49: 3004–3005.

134 Gaillot, O., Blondiaux, N., Loiez, C., Wallet, F., Lemaitre, N. *et al.* 2011. Cost-effectiveness of switch to matrix-assisted laser desorption ionization-time of flight mass spectrometry for routine bacterial identification. *Journal of Clinical Microbiology* 49: 4412.

135 Martiny, D., Cremagnani, P., Gaillard, A., Miendje Deyi, V. Y., Mascart, G. *et al.* 2014. Feasibility of matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) networking in university hospitals in Brussels. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 33: 745–754.

136 Villegas, E. N., Glassmeyer, S. T., Ware, M. W., Hayes, S. L., and Schaefer, F. W., 3rd. 2006. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry-based analysis of Giardia lamblia and Giardia muris. *The Journal of Eukaryotic Microbiology* 53 Suppl 1: S179–81.

137 Cassagne, C., Pratlong, F., Jeddi, F., Benikhlef, R., Aoun, K. *et al.* 2014. Identification of Leishmania at the species level with matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 20: 551–557.

138 Martiny, D., Bart, A., Vandenberg, O., Verhaar, N., Wentink-Bonnema, E. *et al.* 2014. Subtype determination of Blastocystis isolates by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS). *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 33: 529–536.

139 Calderaro, A., Piergianni, M., Buttrini, M., Montecchini, S., Piccolo, G. *et al.* 2015. MALDI-TOF mass spectrometry for the detection and differentiation of Entamoeba histolytica and Entamoeba dispar. *PLoS ONE* 10: e0122448.

140 Yssouf, A., Almeras, L., Berenger, J. M., Laroche, M., Raoult, D., and Parola, P. 2015. Identification of tick species and disseminate pathogen using hemolymph by MALDI-TOF MS. *Ticks and Tick-Borne Diseases* 6: 579–586.

141 Yssouf, A., Almeras, L., Terras, J., Socolovschi, C., Raoult, D., and Parola, P. 2015. Detection of *Rickettsia* spp. in ticks by MALDI-TOF MS. *PLoS Neglected Tropical Diseases* 9: e0003473.

142 Yssouf, A., Socolovschi, C., Kernif, T., Temmam, S., Lagadec, E. *et al.* 2014. First molecular detection of *Rickettsia africae* in ticks from the Union of the Comoros. *Parasites & Vectors* 7: 444.

143 Yssouf, A., Socolovschi, C., Leulmi, H., Kernif, T., Bitam, I. *et al.* 2014. Identification of flea species using MALDI-TOF/MS. *Comparative Immunology, Microbiology and Infectious Diseases* 37: 153–157.

144 Evason, D. J., Claydon, M. A., and Gordon, D. B. 2001. Exploring the limits of bacterial identification by intact cell-mass spectrometry. *Journal of the American Society for Mass Spectrometry* 12: 49–54.

145 Monteiro, J., Inoue, F. M., Lobo, A. P., Sugawara, E. K., Boaretti, F. M., and Tufik, S. 2015. Fast and reliable bacterial identification direct from positive blood culture using a new TFA sample preparation protocol and the Vitek(R) MS system. *Journal of Microbiological Methods* 109: 157–159.

146 Schieffer, K. M., Tan, K. E., Stamper, P. D., Somogyi, A., Andrea, S. B. *et al.* 2014. Multicenter evaluation of the Sepsityper extraction kit and MALDI-TOF MS for direct identification of positive blood culture isolates using the BD BACTEC FX and VersaTREK((R)) diagnostic blood culture systems. *Journal of Applied Microbiology* 116: 934–941.

147 Ferreira, L., Sanchez-Juanes, F., Munoz-Bellido, J. L., and Gonzalez-Buitrago, J. M. 2011. Rapid method for direct identification of bacteria in urine and blood culture

samples by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: Intact cell vs. extraction method. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 17: 1007–1012.

148 Karpanoja, P., Harju, I., Rantakokko-Jalava, K., Haanpera, M., and Sarkkinen, H. 2014. Evaluation of two matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of viridans group streptococci. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 33: 779–788.

149 Mather, C. A., Rivera, S. F., and Butler-Wu, S. M. 2014. Comparison of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of mycobacteria using simplified protein extraction protocols. *Journal of clinical Microbiology* 52: 130–138.

150 Pulcrano, G., Iula, D. V., Vollaro, A., Tucci, A., Cerullo, M. *et al.* 2013. Rapid and reliable MALDI-TOF mass spectrometry identification of Candida non-albicans isolates from bloodstream infections. *Journal of Microbiological Methods* 94: 262–266.

151 Westblade, L. F., Jennemann, R., Branda, J. A., Bythrow, M., Ferraro, M. J. *et al.* 2013. Multicenter study evaluating the Vitek MS system for identification of medically important yeasts. *Journal of Clinical Microbiology* 51: 2267–2272.

152 Won, E. J., Shin, J. H., Lee, K., Kim, M. N., Lee, H. S. *et al.* 2013. Accuracy of species-level identification of yeast isolates from blood cultures from 10 university hospitals in South Korea by use of the matrix-assisted laser desorption ionization-time of flight mass spectrometry-based Vitek MS system. *Journal of Clinical Microbiology* 51: 3063–3065.

153 Branda, J. A., Rychert, J., Burnham, C. A., Bythrow, M., Garner, O. B. *et al.* 2014. Multicenter validation of the VITEK MS v2.0 MALDI-TOF mass spectrometry system for the identification of fastidious gram-negative bacteria. *Diagnostic Microbiology and Infectious Disease* 78: 129–131.

154 Chao, Q. T., Lee, T. F., Teng, S. H., Peng, L. Y., Chen, P. H. *et al.* 2014. Comparison of the accuracy of two conventional phenotypic methods and two MALDI-TOF MS systems with that of DNA sequencing analysis for correctly identifying clinically encountered yeasts. *PLoS ONE* 9: e109376.

155 Yssouf, A., Almeras, L., Raoult, D., and Parola, P. 2016. Emerging tools for identification of arthropod vectors. *Future Microbiology* 11: 549–566.

156 Yssouf, A., Flaudrops, C., Drali, R., Kernif, T., Socolovschi, C. *et al.* 2013. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of tick vectors. *Journal of Clinical Microbiology* 51: 522–528.

157 Peng, J., Yang, F., Xiong, Z., Guo, J., Du, J. *et al.* 2013. Sensitive and rapid detection of viruses associated with hand foot and mouth disease using multiplexed MALDI-TOF analysis. *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology* 56: 170–174.

158 Jordana-Lluch, E., Carolan, H. E., Gimenez, M., Sampath, R., Ecker, D. J. *et al.* 2013. Rapid diagnosis of bloodstream infections with PCR followed by mass spectrometry. *PLoS ONE* 8:e62108.

4

The Challenges of Identifying Mycobacterium to the Species Level using MALDI-TOF MS

Part 4A Modifications of Standard Bruker Biotyper Method

Graham Rose,[1] Renata Culak,[2] Timothy Chambers,[2] Saheer E. Gharbia[1] and Haroun N. Shah[2,3]

[1] *Genomics Research Unit, Public Health England, London, UK*
[2] *Proteomics Research, Public Health England, London, UK*
[3] *Department of Natural Sciences, Middlesex University, London, UK*

4A.1 Taxonomic Structure of the Genus *Mycobacterium*

A genus of Actinobacteria, mycobacteria are aerobic, rod-shaped bacteria which are characterised by a high GC content and a complex lipid-rich cell wall comprised of mycolic acids which can total up to 60% of the dry weight (Madigan, 2012). It was this physical property of the cell wall that was exploited in 1882 by Koch, who first demonstrated the presence of the obligate human pathogen, *Mycobacterium tuberculosis,* the causative agent of tuberculosis (TB), through staining with alkaline methylene blue (Sakula, 1982). Later in the same year, the Ziehl–Neelsen stain was developed, which used a similar process to identify acid-fast bacteria, and is still widely used to identify mycobacteria.

The genus now comprises more than 174 different species (http://www.bacterio.net/mycobacterium.html). Among these, other pathogenic species include *M. leprae*, which causes leprosy, and *M. bovis*, which principally causes bovine TB. In addition to these pathogenic species, the genus is largely made up of environmental organisms and is therefore widespread throughout nature, ranging from inhabitants of soil and water. However, these non-tuberculosis mycobacteria (NTMs) are important opportunistic human and animal pathogens, with the ability to cause pulmonary diseases similar to tuberculosis, as well as cutaneous and disseminated infections. It was this similarity that led to the group being described as environmental or atypical mycobacteria in an attempt to distinguish them from *M. tuberculosis* infections, but NTM is the favoured designation.

Early classification of mycobacterial species was made on the basis of a variety of phenotypic features, such as growth rate and pigmentation, as well as clinical significance (Stahl and Urbance, 1990). On the basis of the growth rate, a fundamental division can be made, splitting mycobacteria into two major groups: fast and slow growers.

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

The fast growers include mainly opportunistic or non-pathogenic mycobacteria, such as *M. fortuitum* and *M. smegmatis*, which can be cultured from dilute inocula within a week. In contrast, the slow-growing species can take one or more weeks for visible growth from dilute inocula, which includes the species *M. tuberculosis*, *M. bovis* and *M. leprae*, the causative agents of human and bovine TB and leprosy, respectively. Pigmentation separates the genus into three groups: pigmentation when grown in light or photochromogens (includes *M. kansasii*), pigmentation irrespective of light or scotochromogens (includes *M. gordonae*) and finally non-pigmented or nonchromogens (includes *M. tuberculosis* and *M. ulcerans*). Using clinical importance, species can be separated in descending order based on (1) obligate pathogens, (2) facultative intracellular pathogens and (3) saprophytes. Taken together, Timpe and Runyon developed a working classification of all NTMs in the 1950s based on these characteristics, designating NTMs into Runyon groups I–IV (Timpe and Runyon, 1954).

As molecular methods were developed, such as 16S rRNA sequencing, these have now revealed the macro-population structure of mycobacteria (Stahl and Urbance, 1990). The phylogenetic structure of mycobacteria based on this method is shown in Figure 4A.1, and of note is the position of the TB-causing species, collectively known as the *Mycobacterium tuberculosis* complex (MTBC) (Cole *et al.*, 1998), together with the smooth tubercle bacilli, which includes *M. canetti*; it is hypothesised that it was an

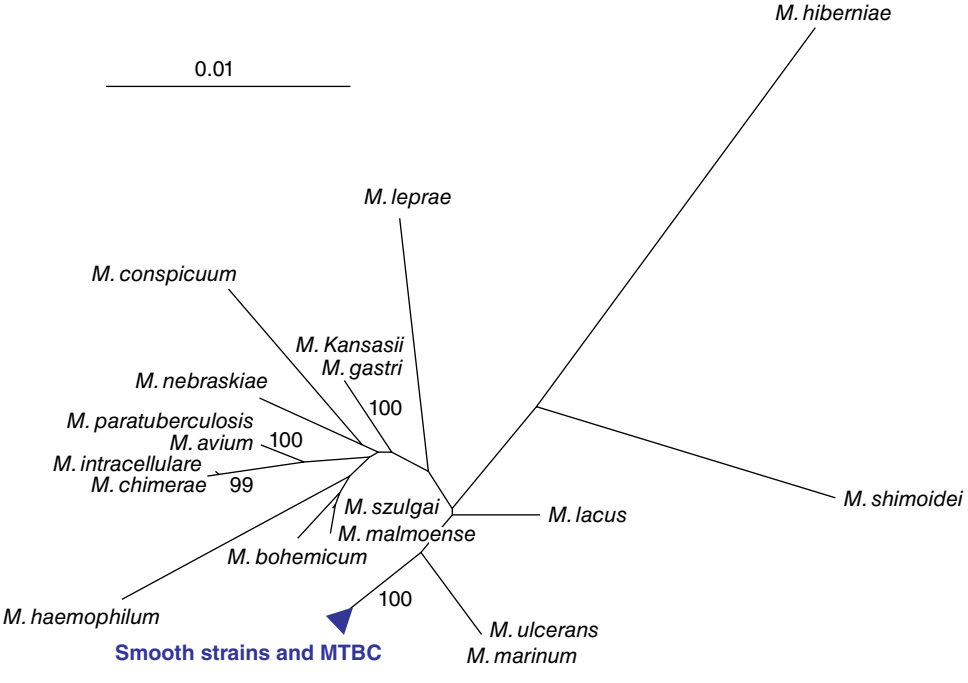


Figure 4A.1 Phylogenetic structure of the genus Mycobacterium. The neighbour-joining tree is based on 16S sequences from 17 smooth mycobacterial and MTBC strains. The blue triangle represents the MTBC strains, which differ by up to one nucleotide. Bootstrap support higher than 90% shown on nodes. Scale bar is pairwise distances after Jukes-Cantor correction. Image reproduced from Gutierrez *et al.* (2005), under the Creative Commons Attribution License (CCAL), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ancestral pool of smooth tubercle-like bacilli from which the MTBC originated (Supply *et al.*, 2013). Lastly, ongoing whole genome sequencing and analysis of different mycobacterial species a enabling the definitive resolution of the phylogenetic relationships of the genus (O'Neill *et al.*, 2015; Tortoli *et al.*, 2015).

4A.2 Tuberculosis-Causing Mycobacteria

Tuberculosis (TB) is caused by several closely related species known as the MTBC, and as previously described, the infamous member of the MTBC is the human-adapted pathogen *M. tuberculosis*, the aetiologic agent of human TB along with *M. africanum*, a phylogenetic variant limited to West Africa (de Jong *et al.*, 2010). Together these species are regarded as human-adapted MTBC members. Early sequencing of MTBC species showed that they share more than 99.9% sequence identity (Sreevatsan *et al.*, 1997), as demonstrated by the collapsed branches within Figure 4A.1 for the MTBC members. However, despite this close relatedness, members of the MTBC display different phenotypic characteristics and mammalian host ranges. The MTBC includes several other species and subspecies that are adapted to various hosts, including both wild and domestic animal species; these bacterial variants have been referred to as 'ecotypes' (Smith *et al.*, 2006). Here, an ecotype is used as the definition of a set of strains using the same or similar ecological resources. The host of *M. bovis* is largely cattle, which is of considerable agricultural significance owing to the associated cost of bovine TB, estimated globally at $3 billion per year (Garnier *et al.*, 2003). *M. bovis* can also cause TB in humans through the consumption of unpasteurized milk (de la Rua-Domenech, 2006), but modern food practices have effectively stopped this transmission route, and person-to-person transmission of *M. bovis* is rare (Evans *et al.*, 2007).

Other MTBC pathogens include *M. microti* (infects voles), *M. caprae* (infects sheep and goats) and *M. pinnipedii* (infects seals and sea lions). An MTBC pathogen of Dassies, or Rock Hyrax, has been isolated in South Africa and named the Dassie bacillus (Parsons *et al.*, 2008), while more recently an MTBC pathogen of banded mongooses has been identified in Botswana named *M. mungi* (Alexander *et al.*, 2010). A special member of the MTBC is *M. canetti*, a rare tubercle bacillus with an unusual smooth colony phenotype, unlike the classical rough appearance of other MTBC members (Van Soolingen *et al.*, 1997). *M. canetti* and the other smooth TB bacilli harbour greater genetic diversity compared with the rest of the MTBC, and are more distantly related to the remaining MTBC than any two other MTBC strains are to each other (Gutierrez *et al.*, 2005). *M. canetti* is subsequently a common choice as an outgroup in phylogenetic analysis (Bentley *et al.*, 2012; Comas *et al.*, 2010). It is anticipated that MTBC members of other ecotypes will likely be identified in future studies.

4A.3 Non-tuberculosis Mycobacteria

In the developed world, NTM incidence is now greater than TB infections (Johnson and Odell, 2014), and these species are recognized as emerging pathogens with significant impact on human health. Incidence is increasing largely owing to longer life expectancy,

immunosuppressive chemotherapy and within immunocompromised patients due to the AIDs pandemic. Owing to their saprophytic lifestyle, NTMs are ubiquitous in nature, and have been isolated from a wide range of habitats, including soil, water, food and artificial environments (Falkinham, 2009).

The most notable member of the NTMs are the species and subspecies that make up the *M. avium* complex (MAC), which are the most common cause of NTM disease in humans, and include many species and subspecies. They are found in both fresh and saltwater sources, as well as various animals, household dust and cigarettes (Eaton *et al.*, 1995; Johnson and Odell, 2014). Within this complex, the species *M. avium* and *M. intracellulare* are the most common NTM infection in patients with AIDS. Another MAC species, M. *chimaera*, has recently been implicated in endocarditis infections through contaminated water within equipment used during heart surgery (Sax *et al.*, 2015), while a MAC subspecies, *M. avium* subspecies *paratuberculosis*, has been associated with Crohn's disease (Behr and Kapur, 2008; Hermon-Taylor, 2009). Moving to non-MAC species, the second most common NTM to cause respiratory disease within patients with AIDS is *M. kansasii,* which was first described in 1953 (Buhler and Pollak, 1953). *M. abscessus,* a fast-growing NTM, is the third most common cause of respiratory disease (Johnson and Odell, 2014), and is also the cause of skin and soft tissue infections, along with *M. fortuitum* and *M. chelonae*.

1. Tackling TB is one of the key priorities for Public Health England ('Tuberculosis in the UK', Public Health England, 2014 Report), as it remains an infectious disease with one of the highest mortality rates. Research has focused largely on elucidating pathogenic factors of the causative organism, *M. tuberculosis*; however, to date, there are no obvious virulence factors such as toxins even among 'superspreader' strains. The sequencing of whole genomes has opened up new vistas for clinical microbiology, making it possible to decipher potential virulence factors from their encoding genes. Tuberculosis is a major threat to the United Kingdom, particularly with the upsurge in antimicrobial resistance and the increasing number of latent and asymptomatic carriers in the population (Van der Werf, 2014). Reliable and rapid detection of pathogens, as well as other opportunistic mycobacteria, are of major consequence because the number of species reported to be causing other pulmonary infections is escalating. New vaccines and drugs are needed to stem the worldwide epidemic of TB that kills over two million people annually ('Global Tuberculosis Control', WHO Geneva 2012 Report). To rationally develop new anti-tubercular agents in the midst of the declining efficacy of BCG and a concomitant increase in antibiotic resistance, it is essential to employ the most powerful technologies to dissect out detailed mechanisms of pathogenicity of *M. tuberculosis* and related mycobacteria to uncover potential new target molecules. It is equally important to understand clearly the host response to ascertain how mycobacteria circumvent host defences and cause disease.

Among the plethora of potential virulence determinants, the mycobacterial cell wall components have been shown to have profound immune modulatory activity. This is partly attributed to the interaction of the surface components with macrophages during phagocytosis and involves pattern recognition receptors (e.g. Toll-like receptors). Despite a century of work on *M. tuberculosis,* the full spectrum of factors that contribute to their virulence remains only partly understood. This is largely due to lack of clearly defined pathogenic determinants that are produced by other actinobacteria such

as the exotoxin of *Corynebacterium diphtheriae.* Current views are that several potential virulence factors contribute to the capacity of this species to adapt, infect, persist and cause transmissible pathology in its host species. Evidence today from whole genome analysis indicates that components of their unique cell wall play a crucial role in the underlying mechanisms of pathogenicity and are potentially the most likely site of new vaccine and antimicrobial drug targets. Detailed understanding of the complex cell envelope is therefore a necessary precursor to current approaches to studying the aetiology of tuberculosis. One of the striking features of the *M. tuberculosis* genome is the large number of genes (517 = 15.5% of total coding capacity) involved in cell wall and cell processes. Recent genome analysis revealed that over 225 annotated genes encode enzymes for the metabolism of fatty acids, of which approximately 100 are predicted to function in the β-oxidation of fatty acids (Cole *et al.*, 1998). This preponderance of enzymes involved in lipid metabolism is most likely related to its ability to utilize fatty acids as a major carbon, a trait retained from its original environmental niche. It was long known that *M. tuberculosis* can shift its metabolism from carbohydrates during in vitro growth to fatty acids when grown in infected host cells, a hypothesis that was subsequently supported by more recent genome analysis (van Els *et al.*, 2014). Various alkanes may be used by mycobacteria as a carbon source, their uptake being dependent on the species, the length of the alkane, its availability and the environment they colonize. Interestingly, they secrete glycolipid surfactants, which allow for the cellular uptake and usage of these compounds. Degradation of alkanes with two or more carbons starts with the oxidation of the terminal methyl group by alkane hydroxylases, yielding the corresponding primary alcohol, which in turn is oxidized by alcohol dehydrogenase to an aldehyde, which is then eventually converted to a fatty acid to be used as an energy source or for cell wall synthesis.

Research focused on the proteome and transcriptome of mycobacteria have been aimed at elucidating detailed mechanisms of adaptation and virulence by unravelling global information on gene expression by measuring the levels of proteins during different growth conditions such as iron starvation (Calder and Horwitz, 1998) and macrophage infections (Lee and Horwitz, 1995). Despite considerable work on the proteome of this taxon, this is perhaps the area that is most poorly understood, and yet they are the most highly represented in its genome. For example, proteins of unknown function and conserved hypothetical function can account for >1500 genes (~30% of the coding capacity), in addition to the unusual presence of the unrelated PE and PPE families of acidic, glycine-rich proteins. The function of many proteins localized in the cell wall and cell membrane is still unknown, and the vast majority of the proteome remains designated 'unknown function'.

Clear and unambiguous definition of a species and subspecies is crucial to monitoring disease trends from the environment and controlling infections in the future. Pathogenic species were once in the environment; it is unlikely that this trend will cease. Instead, gene decay or gene acquisition, for example through horizontal gene transfer, is likely to drive more species into the human ecosystem. The dDDH data based on WGS is revealing that among some mycobacteria, the centre of variation of 'human species' appears to overlap with some environmental taxa and emphasizes the fluidity of this transmission. Even for such a well-studied genus as *Mycobacterium*, which consists of 170 species, many are poorly circumscribed and extremely difficult to reliably delineate. *M. tuberculosis* is a high-risk pathogen, and thus safe and sustainable healthcare

services are essential to combat disease. Traditional methods of identifying these species are being replaced by NGS technologies, but this still needs considerable development to have a direct impact on a patient's well-being. Proteomics, in the form of MALDI-TOF mass spectrometry, is poised to supersede existing methods, but in the case of mycobacteria it is not optimized and needs a programme of focused research and development. As the pilot programme to implement WGS into tuberculosis patient care and information management pathways develops, a key gap remains in supporting rapid diagnostics. Genomics require 3–5 days to process and analyze DNA sequencing before results are reported; by contrast, species identification by MALDI-TOF mass spectrometry (MS) takes minutes and, if implemented, will aid selection of intervention and treatment choice. Building the required database will provide a safe, inexpensive and rapid method of identification because most laboratories today have access to MALDI-TOF MS technology. Once established, this would lead to an unprecedented advance for infection treatment and control because same-day identification of *Mycobacterium* species, resistance and relatedness to particular clades would be available. This technology paves the way for the application of the next generation of more automated and advanced forms of MS. The development of specialized mycobacterial MALDI-TOF MS databases began a process of transformation.

4A.4 MALDI-TOF MS Mycobacteria Library and Parameters for Identification

The Mycobacteria Library was created as a stand-alone database to assemble spectra of reference and clinical strains generated using a modified method that was applied to that of other prokaryotes, to optimize the mass spectral profiles of Mycobacterium species. Factors such as their robust cell wall, dense mycolic acid layers and relatively lower concentration of ribosomal proteins compared to other microorganisms exacerbate the problem of obtaining high-confidence identification scores. Therefore, a bespoke approach was undertaken by Bruker Daltonics to create a database (Version 2) of 256 MSPs of 131 species of Culture Collection strains and 44 clinical isolates of mycobacteria.

The identification of most bacteria by MALDI-TOF MS is assessed by the following parameters: a score of 2–3 provides identification at a high level of confidence; a score 1.7–1.9, identification at a low confidence level while a score below 1.7 indicates unreliable identification. However, because of the inherent difficulties associated with identification of mycobacteria, a decrease in log (score) value to 1.8 for high-confident species and to 1.6 for low-confident species identification is generally accepted by most laboratories. In addition, manual acquisition appears to improve the quality of the spectra and therefore increases species identification, particularly at the lower score. This could be due to the low biomass of the starting material and manual adjustment to enhance the acquisition of the lower intensity of peaks obtained for mycobacterial samples (Pranada *et al.*, 2014).

Previously published procedures for the extraction of mycobacterial proteins have been technically complex and involved extensive washing and centrifugation steps leading to loss of biological material (Khéchine *et al.*, 2011; Machen *et al.*, 2013). To protect handlers of suspected pathogens, Bruker recommends a heat inactivation step of 30 min

at 95 °C (Bruker's SOP: 'Inactivation and sample preparation procedure for mycobacteria sp.'), but others advise 60 min at 95 °C. To assess its potency, we tested several methodologies and concluded that 30 min at 95 °C in a water bath, rather than a heating block, leads to adequate inactivation. We further tested the method described by Khéchine *et al.* (2011), which is recommended for specialist laboratories and includes treatment with Tween 20 in the lysis step to aid cell disruption. However, it was necessary to wash off the Tween 20 before mass spectral analysis, which added substantially to the time and a potential loss of biological material. MALDI-TOF MS identification requires about 10^5 cells to provide reliable identification.

The methods reported here are based on the use of a vortex and the FastPrep cell disruption system in the presence of glass or silica beads and removal or depletion of the lipid layer prior to cell disruption for identification with the Bruker Biotyper instrument (Part 4A), and in Part 4B, a new system, designated ASTA, with its own mycobacterial database, is explored.

4A.5 Methods for Extraction

To assess the effect of disruption, viable, intact cells of *M. fortuitum* were compared with cells treated with dichloromethane in an attempt to remove or deplete their lipid content. Initially, cells were inactivated by incubation for 30 min at 95 °C, after which mechanical lysis with 0.5 mm zirconium beads was also used. Preparations, visualized by electron microscopy, revealed significant differences in cell shapes and aggregation as shown in Figure 4A.2.

4A.5.1 Method: Bruker's Protocol

Cells were suspended in 300 µl of water, and three different concentrations of cells were tested: 10 µl loop, half of that amount (5 µl) and 1 µl loop. The cells were inactivated for 30 min at 95 °C followed by addition of 900 µl of 100% ethanol. After centrifugation at 14,000 g for 2 min and drying of the pellet, 50 µl of 100% acetonitrile was added, and the pellet was re-suspended. A small spatula of silica/zirconium beads (0.1 mm) was added, and the samples were vortexed for 1 min. Fifty microlitres of 70% TFA was added. The samples were mixed well by vortexing, and centrifuged. Each sample was analyzed on four spots: the first two had 1 µl of supernatant applied followed by 1 µl of matrix, and the following two had 1.5 µl of sample followed by 1.5 µl of matrix.

4A.5.2 The Methods of Khéchine *et al.*, 2011

The same amounts of cells were used as in Bruker's protocol, and the cells were suspended in 500 µl of water containing 0.5% Tween 20. The samples were inactivated for 1 h at 95 °C, after which the pellets were washed twice with 500 µl of water; each wash being followed by 10 min centrifugation at 13,000 g. The pellets were re-suspended in 500 µl of water, and a small spatula of silica/zirconium beads (0.1 mm) was added. The samples were batched as follows:

A) Three tubes were placed in the FastPrep for 3 min at a max setting (level 6).
B) Three tubes were vortexed for 1 min.

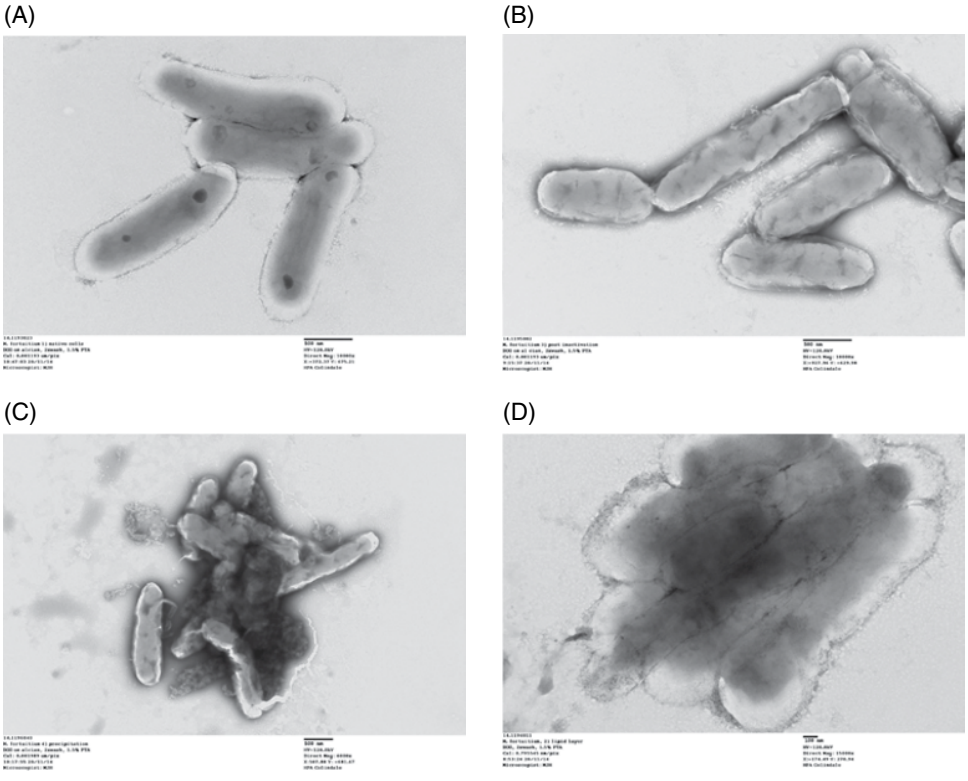


Figure 4A.2 Electron microscopy images of *Mycobacterium fortuitum* cells: (A) Native cells (cell suspension in water diluted 1:1 in a fixing solution), (B) cells after inactivation (30 min at 95°C in a water bath), (C) pellet of inactivated cells after mechanical lysis with 0.5 mm silica beads, (D) cells after lipid extraction with dichloromethane (before treatment with ethanol and bead beating); loss of polarity and clumping of cells were evident.

After centrifugation for 1 min at 11,000 g, the supernatants were spotted on the target plate, 1.5 µl of each sample followed by 1.5 µl of matrix (as suggested in Khéchine *et al.*). One microlitre of each supernatant was placed in duplicate onto an MSP 96 target polished steel (Bruker Daltonics) and air-dried.

From both methods, the spots were overlaid with 1 µL of α-Cyano-4-hydroxycinnamic acid (CHCA) (saturated solution of CHCA in 50 % acetonitrile/2.5 % trifluoroacetic acid) and air-dried to co-crystallize the sample and matrix. The target plate was then processed using a Microflex LT mass spectrometer (Bruker Daltonics, UK, software version 3.4). Data collection was done in an automatic mode by collecting 240 laser shots from six different positions within the spot. MALDI-TOF measurements were recorded in a positive linear mode within the mass range 2–20 kDa.

This method took 45 min in addition to the time to complete Bruker's protocol. The samples prepared using the FastPrep had the highest scores (all above 2.3), but the vortexed samples were also satisfactory with identification scores of between 2.0 and 2.29.

4A.5.3 Silica/Zirconium Bead Variation

In this study, we tested using cells harvested from MGIT broth and Löwenstein–Jensen (LJ) slopes as a starting material because both are routinely used for culture and isolation of clinical isolates of suspected mycobacteria. Two sizes of silica/zirconium beads were tested. The spectra obtained using 0.5 mm beads were cleaner and yielded higher scores (see Figure 4A.3).

For MGIT broth samples, 2 ml of broth was used with 0.1 mm silica beads and with acetonitrile to solubilize the pellet followed by the addition of 70 % formic acid.

For the LJ slope samples, 0.1 mm silica beads were used with acetonitrile to solubilize the pellet followed by 70% formic acid. LJ slope samples with 0.5 mm silica beads were also used, and 70% formic acid was used for the suspension of the pellet instead of the acetonitrile. LJ-grown cultures gave better and more consistent results compared to those from liquid media (MGIT), confirming a previous observation by Suzuki *et al.* (2015). The results obtained using 0.5 mm zirconium beads rather than 0.1 mm were noticeably superior.

4A.5.4 Results and Recommendations

Initially, the method we developed was tested using *M. fortuitum* strains, and results were outstanding with 98% of samples scoring over 2.000. This was used to establish wider profile acquisitions from all available strains, but lower scores were obtained for some species. Because the method by Khéchine *et al.* (2011) required more time and extra washes due to Tween 20 interference with the final result, it was decided to change some of the extraction parameters. After testing several experimental protocols, we found the following robust method.

A quarter of a loopful of cells (about 5 µl volume) is thoroughly suspended in 300 µl water and placed in a boiling bath for 30 min. The samples are then cooled, and 900 µl of 100% ethanol is added to each tube; each sample vortexed for 10 s. The samples are centrifuged at 14,000g for 2 min, and the supernatant is decanted. The centrifugation step is repeated, and the remaining supernatant is removed using a pipette. The pellet is suspended in 50 µl of 70% formic acid, mixed well by pipetting, and 0.5 mm zirconium beads is added (a third of the volume of the sample). Each tube is mixed well by vortexing for 1 min. Fifty µl of 100% acetonitrile is added to the sample and mixed well by vortexing for 1 min. The sample is centrifuged at 14,000g for 2 min, and 1 µl of each supernatant is placed in duplicate onto MSP 96 target polished steel (Bruker Daltonics) and air-dried. The spots are overlaid with 1 µl of α-Cyano-4-hydroxycinnamic acid (CHCA) (saturated solution of CHCA in 50% acetonitrile/2.5% trifluoroacetic acid) and air-dried to co-crystallize the sample and matrix. The target plate is then processed using a Microflex LT mass spectrometer (Bruker Daltonics, UK, software version 3.4). Data collection is done in an automatic mode by collecting 240 laser shots from six different positions within the spot. MALDI-TOF MS measurements are recorded in a positive linear mode within the mass range 2–20 kDa.

The results listed in Table 4A.1 show that mechanical disruption using 0.5 mm zirconium beads provided high scores. Some strains proved to be more difficult to extract; for example, the highest score achieved with cells of *M. gordonae* was 1.567. Two strains of *M. selandiae* and *M. tomidae* have similar mass spectral profiles and were identified as *M. austroafricanum* and *M. chimaera intracellulare*, respectively. This ambiguous identification is likely to be the result of taxonomic issues and was confirmed by DNA–DNA hybridization (Walker T., PhD thesis).

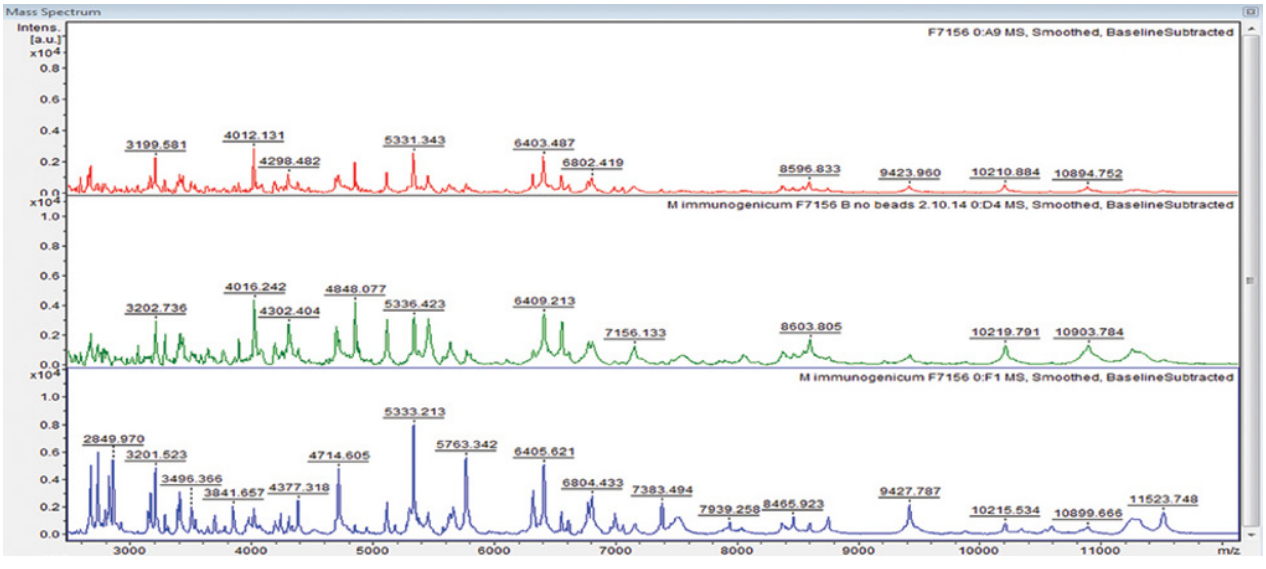


Figure 4A.3 Influence of growth and methods on identification scores. 1: (Top) MGIT broth, 0.1 mm silica beads, elution in 100% acetonitrile and 70% formic acid. Score: 1.881 2: (Middle) LJ culture, 0.1 mm silica beads, elution in 100% acetonitrile and 70% formic acid. Score: 1.782 3: (Bottom) As in A above, LJ culture, 0.5 mm beads. Score: 2.066

Table 4A.1 List of mycobacterial strains used blindly for MALDI-TOF MS identification.

Mycobacterium Strain	MALDI ID Identification	Score
M. abscessus subsp. massiliense JCM15300	*M. abscessus*	2.398
M. algericum DSMZ45454	*M. algericum*	2.054
M. aurum NCTC10437	*M. aurum*	2.294
M. avium NCTC 13034	*M. avium*	2.253
M. chubuense NCTC 10819	*M. chubuense*	2.072
M. celatum JCM12373	*M. celatum*	2.402
M. chimaera JCM14737	*M. chimaera intracellulare* group	2.309
M. chelonae X29303	*M. chelonae*	2.108
M. farcinogens NCTC10955	*M. farcinogenes senegalense* group	2.450
M. fortuitum NCTC10894	*M. fortuitum*	2.349
M. immunogenicum F7156	*M. mucogenicum phocaicum* group	2.093
M. kansasii NCTC10268	*M. kansasii*	2.219
M. kumamotonense JCM13453	*M. kumamotonense*	2.047
M. perigrinum NCTC10264	*M. perigrinum*	2.521
M. phlei NCTC8151	*M. phlei*	2.412
M. rhodesiae NCTC10779	*M. rhodesiae*	2.050
M. scrofulaceum NCTC10803	*M. scrofulaceum*	2.259
M. smegmatis NCTC00333	*M. smegmatis*	2.311
M. smegmatis NCTC 08159	*M. smegmatis*	2.443
M. thermoresistibile NCTC 10409	*M. thermoresistibile*	2.164
M. vaccae NCTC 10916	*M. vaccae*	2.437
M. vulneris JCM18115	*M. vulneris*	2.063
M. xenopi NCTC 10042	*M. xenopi*	2.202
M. agri JCM6377	*M. agri*	1.818
M. kyorinense JCM15038	*M. kyorinense*	1.819
M. shimoidei JCM12376	*M. shimoidei*	1.661
M. terrae JCM12143	*M. terrae*	1.68
M. gordonae NCTC9822	*M. gordonae*	1.567
M. hafniae NCTC11056	Not reliable ID	
M. triplex JCM14744	Not reliable ID	
M. selandiae NCTC 11054	*M. austroafricanum*	2.171
M. tomidae NCTC 10428	*M chimaera_intracellulare group*	2.109

With the exception of *M. chelonae* X29303 and *M. immunogenicum* F7156, which were clinical isolates, others were from Culture Collections. All results were obtained by using the extraction method with a cell inactivation of 30 min at 95 °C, followed by cell disruption with 0.5 mm zirconium beads as described in 4A.5.4.

4A.6 Protein Profiling of Cell Extracts using SELDI-TOF MS

The challenge of analyzing *Mycobacterium* spp. by MALDI-TOF MS requires further work and potentially more thorough disruption of cells; perhaps another form of MALDI-TOF MS would in the future prove more appropriate (see review, Shah *et al.*, 2005). We have investigated SELDI-TOF MS for a range of human pathogens and, in general, have found it particularly useful for distinguishing species that are closely related, such as subgroups of *Propionibacterium acnes* (see Dekio *et al.*, 2015, Chapter 5). SELDI-TOF MS is unique in that it uses chromatographic surfaces to retain proteins based on their physio-chemical characteristics (see Chapter 1, Section 1.3). Proteins that are bound and retained on the surfaces are analyzed by MALDI-TOF MS. Here, the samples prepared for analysis by MALDI-TOF MS were used, and 1 µl of each sample was applied to each spot of the ProteinChip® array (for details, Shah *et al.*, 2005). The chip was dried, and 0.5 µl of sinapinic acid applied twice onto each spot. The ProteinChip® arrays were analyzed in a mass spectrometer (Ciphergen BioSystems, Model, PBS II) according to an automated data collection protocol. The spectra were generated at a laser intensity 220, high mass 50 kDa, detector sensitivity 10 and focus mass 25 kDa. The instrument was operated in positive ion mode, and a nitrogen laser emitting at 337 nm was used. Figure 4A.4 shows a partial mass spectrum of the extended mass range that SELDI-TOF MS achieves and the unique spectral profile for each *Mycobacterium* spp. tested. A comprehensive study of the parameters is required to optimize this approach further, but preliminary evidence suggests that it may prove a valuable adjunct to the existing MALDI-TOF MS method for species that are poorly resolved by virtue of the additional mass ions derived.

4A.7 Conclusion

The impetus to utilize MALDI-TOF MS for species identification has been widely reported, and new, improved methods are frequently being investigated (Lotz, *et al.*, 2010; Mäkinen *et al.*, 2006, Mather *et al.*, 2014; Wang *et al.*, 2012). However, the identification of *Mycobacterium* to the species level has been fraught with difficulties, largely because of the poor growth, low cell yield and long mean generation times of some species. Added to this, the group comprises immense species diversity in which many taxa are described using poorly defined characters. Even the use of whole genome sequencing is often unable to resolve differences between some closely related taxa, and specific SNPs are being used to define subgroups. However, for species that are better defined, MALDI-TOF MS has shown enormous potential. Key to successful implementation depends greatly on sample preparation. It appears to be difficult to establish one reproducible method for the extraction of mycobacterial proteins. Mechanical disruption is essential; depending on the laboratory setup, vortex or beat beating using FastPrep or Biospec were applied with high-confidence identification scores here (see, e.g. Table 4A.1). It was suggested by several MALDI-TOF users that a local, site-specific database should be established to aid identification of geographically significant species and strains (Khéchine *et al.*, 2011; Mather *et al.*, 2014). Many authors agree that this technique allows for faster and easier diagnosis of mycobacteria compared with conventional phenotypic identification methods.

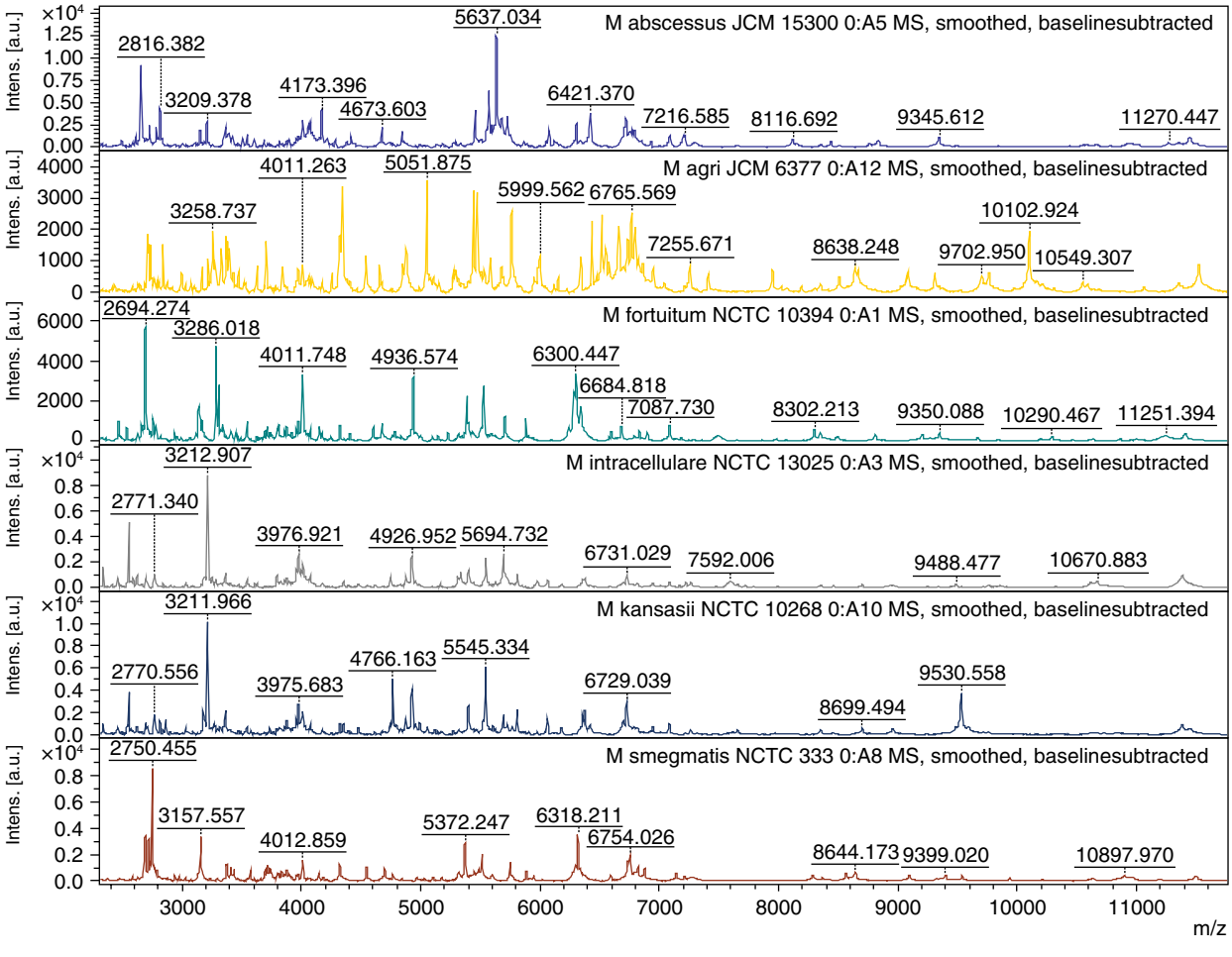


Figure 4A.4 Partial mass spectrum using SELDI-TOF MS of *Mycobacterium* species. From top to bottom; *M. abscessus, M. agri. M. fortuitum, M. intracellulare, M. kansasii* and *M. smegmatis.* SELDI-TOF MS has the capacity to generate mass ions in excess of 100,000 kDA (see Shah *et al.*, 2005) and clearly has potential application for identification of *Mycobacterium* species.

At scientific conferences, very comprehensive studies have been reported in symposia devoted entirely to problems relating to the characterization of *Mycobacterium* species. In one such study (ECCMID 2015), the general consensus reached was that for the identification of *Mycobacterium* species using the Bruker Biotyper, the confidence threshold should be lowered to between 1.6 and 1.9 for this group. However, the genus *Mycobacterium* comprises many environmental species that are poorly defined and excluded from MALDI-TOF MS mycobacterial databases. It is our view that a general rule should not be applied across the whole genus. Here we demonstrated that many species give an acceptable score of >2 (see Table 4A.1). However, other species such as *M. gordonae* yield poor confidence scores of around 1.6. In the long term, it might be necessary to develop a sliding scale for various groups of mycobacterial species. This will highlight the poorly resolved taxa which may be investigated further while emphasizing those that can be identified with confidence. Currently, the high scores reported for many species indicate that MALDI-TOF MS has considerable potential for identification of *Mycobacterium* species, and more in-depth analytical methods would continue to improve the methodology, particularly to separate closely related taxa of the *Mycobacterium tuberculosis* complex of species that are poorly differentiated by MALDI-TOF MS.

References

Alexander, K. A., Laver, P. N., Michel, A. L., Williams, M., van Helden, P. D., Warren, R. M. and Gey van Pittius, N. C., 2010. Novel *Mycobacterium tuberculosis* complex pathogen, M. mungi. *Emerg. Infect. Dis.* 16, 1296–1299. doi:10.3201/eid1608.100314

Behr, M. A. and Kapur, V., 2008. The evidence for *Mycobacterium paratuberculosis* in Crohn's disease. *Curr. Opin. Gastroenterol.* 24, 17–21. doi:10.1097/MOG.0b013e3282f1dcc4

Bentley, S. D., Comas, I., Bryant, J. M., Walker, D., Smith, N. H., Harris, S. R., Thurston, S., Gagneux, S., Wood, J., Antonio, M., Quail, M. A., Gehre, F., Adegbola, R. A., Parkhill, J. and de Jong, B. C., 2012. The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the M. tuberculosis complex. *PLoS Negl. Trop. Dis.* 6, e1552. doi:10.1371/journal.pntd.0001552

Buhler, V. B. and Pollak, A., 1953. Human infection with atypical acid-fast organisms; report of two cases with pathologic findings. *Am. J. Clin. Pathol.* 23, 363–374.

Calder, K. M. and Horwitz, M. A. (1998). Identification of iron-regulated proteins of *Mycobacterium tuberculosis* and cloning of tandem genes encoding a low iron-induced protein and a metal transporting ATPase with similarities to two-component metal transport systems. *Microb. Pathog.* 24, 133–143.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V, Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. and Barrell, B. G., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544. doi:10.1038/31159

Comas, I., Chakravartti, J., Small, P. M., Galagan, J., Niemann, S., Kremer, K., Ernst, J. D. and Gagneux, S., 2010. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* 42, 498–503. doi:10.1038/ng.590

de Jong, B. C., Antonio, M. and Gagneux, S., 2010. *Mycobacterium africanum*–review of an important cause of human tuberculosis in West Africa. *PLoS Negl. Trop. Dis.* 4, e744. doi:10.1371/journal.pntd.0000744

de la Rua-Domenech, R., 2006. Human *Mycobacterium bovis* infection in the United Kingdom: Incidence, risks, control measures and review of the zoonotic aspects of bovine tuberculosis. *Tuberculosis (Edinb).* 86, 77–109. doi:10.1016/j.tube.2005.05.002

Eaton, T., Falkinham, J. O. and Vonreyn, C. F., 1995. Recovery of *Mycobacterium avium* from cigarettes. *J. Clin. Microbiol.* 33, 2757–2758.

Evans, J. T., Smith, E. G., Banerjee, A., Smith, R. M. M., Dale, J., Innes, J. A., Hunt, D., Tweddell, A., Wood, A., Anderson, C., Hewinson, R. G., Smith, N. H., Hawkey, P. M. and Sonnenberg, P., 2007. Cluster of human tuberculosis caused by *Mycobacterium bovis*: Evidence for person-to-person transmission in the UK. *Lancet* (London, England) 369, 1270–1276. doi:10.1016/S0140-6736(07)60598-4

Falkinham, J. O., 2009. Surrounded by mycobacteria: Nontuberculous mycobacteria in the human environment. *J. Appl. Microbiol.* 107, 356–67. doi:10.1111/j.1365-2672.2009.04161.x

Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P. R., Parkhill, J., Barrell, B. G., Cole, S. T., Gordon, S. V. and Hewinson, R. G., 2003. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7877–82. doi:10.1073/pnas.1130426100

Global Tuberculosis Control. WHO Geneva 2012 Report.

Gutierrez, M. C., Brisse, S., Brosch, R., Fabre, M., Omaïs, B., Marmiesse, M., Supply, P. and Vincent, V., 2005. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathog.* 1, e5. doi:10.1371/journal.ppat.0010005

Hermon-Taylor, J., 2009. *Mycobacterium avium* subspecies *paratuberculosis*, Crohn's disease and the Doomsday scenario. *Gut Pathog.* 1, 15. doi:10.1186/1757-4749-1-15

Johnson, M. M. and Odell, J. A., 2014. Nontuberculous mycobacterial pulmonary infections. *J. Thorac. Dis.* 6, 210–20. doi:10.3978/j.issn.2072-1439.2013.12.24

Khéchine, A., Couderc, C., Flaudrops, C., Raoult, D. and Drancourt, M. 2011. Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight Mass Spectrometry identification of Mycobacteria in routine clinical practice. *PLoS ONE* 6(9), e24720.

Lee, B-Y. and Horwitz, M. A. (1995). Identification of macrophage and stress-induced proteins of *Mycobacterium tuberculosis*. *J. Clin. Investig.* 96,245–249.

Machen, A., Kobayashi, M., Connelly, M. R. and Wang, Y. F. (Wayne). 2013. Comparison of heat inactivation and cell disruption protocols for identification of Mycobacteria from solid culture media by use of Vitek Matrix-Assisted Laser Desorption Ionization—Time of Flight Mass Spectrometry. *J. Clin. Microbiol.*51, 4226–4229.

Madigan, M., 2012. *Brock Biology of Microorganisms*, 13th edn, International Microbiology. doi:10.1016/B978-1-4832-3136-5.50010-3

Mather, C. A., Rivera, S. F. and Butler-Wu, S. M. 2014. Comparison of the BrukerBiotyper and Vitek MS Matrix-Assisted Laser desorption Ionisation—Time of Flight Mass Spectrometry systems for identification of Mycobacteria using simplified protein extraction protocols. *J. Clin. Microbiol.* 52, 130–138.

O'Neill, M. B., Mortimer, T. D. and Pepperell, C. S., 2015. Diversity of *Mycobacterium tuberculosis* across Evolutionary Scales. *PLoS Pathog.* 11, e1005257. doi:10.1371/journal.ppat.1005257

Parsons, S., Smith, S. G. D., Martins, Q., Horsnell, W. G. C., Gous, T. A., Streicher, E. M., Warren, R. M., van Helden, P. D. and Gey van Pittius, N. C., 2008. Pulmonary infection due to the dassie bacillus (*Mycobacterium tuberculosis* complex sp.) in a free-living dassie (rock hyrax-Procavia capensis) from South Africa. *Tuberculosis (Edinb).* 88, 80–83. doi:10.1016/j.tube.2007.08.012

Pranada, A. B., Timke, M., Witt, E. and Kostrzewa, M. 2014. Evaluation of lower threshold values for MALDI Biotyper analysis of Nontuberculous Mycobacteria. Poster ICAAC 2014, D-190.

Public Health England, Tubeculosis in the UK. 2014 Report.

Sakula, A., 1982. Robert Koch: Centenary of the discovery of the tubercle bacillus, 1882. *Thorax* 37, 246–251. doi:10.1136/thx.37.4.246

Sax, H., Bloemberg, G., Hasse, B., Sommerstein, R., Kohler, P., Achermann, Y., Rössle, M., Falk, V., Kuster, S. P., Böttger, E. C. and Weber, R., 2015. Prolonged outbreak of *Mycobacterium chimaera* infection after open-chest heart surgery. *Clin. Infect. Dis.* 61, 67–75. doi:10.1093/cid/civ198

Shah, H. N., Encheva, V., Schmid, O., Nasir, P. *et al.* (2005). Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS): A potentially powerful tool for rapid characterisation of microorganisms. In M. J. Miller (Ed.), *Encyclopedia of Rapid Microbiological Methods* (Vol. 3, pp. 57–96). DHI Publishing, IL, USA.

Smith, N. H., Kremer, K., Inwald, J., Dale, J., Driscoll, J. R., Gordon, S. V., van Soolingen, D., Hewinson, R. G. and Smith, J. M., 2006. Ecotypes of the *Mycobacterium tuberculosis* complex. *J. Theor. Biol.* 239, 220–225. doi:10.1016/j.jtbi.2005.08.036

Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S. and Musser, J. M., 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9869–9874. doi:10.1073/pnas.94.18.9869

Stahl, D. A. and Urbance, J. W., 1990. The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. *J. Bacteriol.* 172, 116–124.

Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., Majlessi, L., Criscuolo, A., Tap, J., Pawlik, A., Fiette, L., Orgeur, M., Fabre, M., Parmentier, C., Frigui, W., Simeone, R., Boritsch, E. C., Debrie, A.-S., Willery, E., Walker, D., Quail, M. A., Ma, L., Bouchier, C., Salvignol, G., Sayes, F., Cascioferro, A., Seemann, T., Barbe, V., Locht, C., Gutierrez, M.-C., Leclerc, C., Bentley, S. D., Stinear, T. P., Brisse, S., Médigue, C., Parkhill, J., Cruveiller, S. and Brosch, R., 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. *Nat. Genet.* 45, 172–179. doi:10.1038/ng.2517

Suzuki, H., Yoshida, S., Yoshida, A., Okuzumi, K., Fukusima, A. and Hishinuma, A. 2015. A novel cluster of *Mycobacterium abscessus* complex revealed by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF-MS).*Diagn. Microbiol..Infect. Dis.* 83, 365–370. doi:10.1016/jdiagmicrobio.2015.08.011

Timpe, A. and Runyon, E. H., 1954. The relationship of atypical acid-fast bacteria to human disease; a preliminary report. *J. Lab. Clin. Med.* 44, 202–209.

Tortoli, E., Fedrizzi, T., Pecorari, M., Giacobazzi, E., De Sanctis, V., Bertorelli, R., Grottola, A., Fabio, A., Ferretti, P., Di Leva, F., Serpini, G. F., Tagliazucchi, S., Rumpianesi, F., Jousson, O. and Segata, N., 2015. The new phylogenesis of the genus *Mycobacterium*. *Int. J. Mycobacteriol.* 4, 77. doi:10.1016/j.ijmyco.2014.10.017

Tuberculosis in the UK, Public Health England, 2014 Report.

van Els, C. A., Corbière, V., Kaat Smits, K. and van Gaans-van den Brink, J.. A. M. (Epub 2014 Aug 11). Toward understanding the essence of post-translational modifications for the *Mycobacterium tuberculosis* immunoproteome. *Front. Immunol.* 5, 361.

Van der Werf, M., Ködmön, C., Katalinić-Jankowić, V., Kummik, T., Soini, H., Richter, E., Papaventsis, D., Tortoli, E., Perrin, M., van Sooligen, D., Žolnir-Dović, M. and Ostergaard Thomsen, V. (2014). Inventory of non-tuberculous mycobacteria in the European Union. *BMC Infect. Dis.* 14, 62.

Van Soolingen, D., Hoogenboezem, T., De Haas, P. E. W., Hermans, P. W. M., Koedam, M. A., Teppema, K. S., Brennan, P. J., Besra, G. S., Portaels, F., Top, J., Schouls, L. M. and Van Embden, J. D. A., 1997. A novel pathogenic taxon of the Mycobacterium tuberculosis Complex, Canetti: Characterization of an exceptional isolate from Africa. *Int. J. Syst. Bacteriol.* 47, 1236–1245. doi:10.1099/00207713-47-4-1236

Part 4B ASTA's MicroID System and Its MycoMp Database for Mycobacteria

Yangsun Kim[1] and Jae-Seok Kim[2]

[1] *ASTA Inc, 11 Floor, Bldg A, AICT, 145, Gwanggyo-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, Korea*
[2] *Department of Laboratory Medicine, Hallym University College of Medicine, Seoul, Korea*

4B.1 Introduction

4B.1.1 The Genus *Mycobacterium*, Disease and MALDI-TOF Mass Spectrometry

Although the most important species of the genus are members of the *Mycobacterium tuberculosis* complex (MTB), which are responsible for 1.5 million deaths annually (World Health Organization, 2015), infections caused by nontuberculous mycobacteria (NTM) can also be serious, especially in patients who are immunosuppressed or have an underlying chronic pulmonary disease (Hoefsloot *et al.*, 2013; Johnson and Odell, 2014).

Identification and differentiation of the species of the genus Mycobacterium is complex and is described in Part 4A. However, regarding the diagnosis, treatment, and prevention of MTB and NTM disease, it is recommended that clinically significant NTM isolates be identified to the species level, and MTB be differentiated from NTM (Griffith *et al.*, 2011; Griffith, 2015). *Mycobacterium* spp. identification has traditionally relied upon phenotypic traits (e.g., growth rate and pigmentation) and biochemical analyses (e.g., nitrate reduction and semi-quantitative catalase activity) (Griffith *et al.*, 2011; Griffith, 2015). Unfortunately, many of these methods are not only time consuming but also have poor resolution. As a result, species identification by molecular-based methods has been augmented and includes the use of commercially available DNA probes and DNA sequencing of target genes (e.g., 16S rRNA and *rpoB*) (Kim *et al.*, 1999; Lee *et al.*, 2000; Woo *et al.*, 2008) (see Figure 4A1, Part 4A). Although these methods are routinely used in developed countries, they are expensive to perform and require specific technical expertise and facilities.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) provides a rapid alternative for microorganism identification based upon differences in protein profiles (Claydon *et al.*, 1996; Hettick *et al.*, 2004; Seng *et al.*, 2009). MALDI-TOF MS spectra can be obtained by direct cell analysis which involves spotting a colony from a culture plate onto a MALDI target without extensive pre-treatment. The spectrum is then compared with protein profiles in the database for

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

identification. The US Food and Drug Administration (FDA) has approved the Bruker Biotyper and VITEK MALDI-TOF MS for species-level identification of several gram-positive and gram-negative bacteria with this simple sample treatment method. However, MALDI-TOF-MS-based identification of mycobacteria has not yet been approved by the FDA, because methods of optimization are still in progress (see Part 4A).

The identification of mycobacteria using MALDI-TOF MS first started in 1996 with the analysis of an *M. smegmatis* strain (Claydon *et al.*, 1996). Since then, several studies (Balážová *et al.*, 2014; El Khéchine *et al.*, 2011; Kim *et al.*, 2015; Pignone *et al.*, 2006; Saleeb *et al.*, 2011) have evaluated the utility of MALDI-TOF MS for the identification of mycobacterial species. It is now well established that prior to analysis of mycobacteria, extensive protein extraction procedures are required for biosafety reasons, and to facilitate the removal or rupture of the dense lipid layer on the surface of the cell (see Part 4A). Commercially available protein extraction methods have been compared, and new simplified sample preparation methods were suggested (Day *et al.*, 2014).

In this section, recent progress in the use of MALDI-TOF MS for the identification of mycobacteria from clinical samples is described using new approaches for faster, safer, and more efficient sample treatment, and algorithms and databases for more efficient identification.

4B.2 MycoMp Database for Mycobacterium: The ASTA Mycobacterial Database

A new commercial MALDI-TOF MS analytical system for microbiology, designated the MicroID system (ASTA, Suwon), was launched recently. MycoMP is a database focussed specifically on mycobacterial identification. The MicroID system comprises a linear MALDI-TOF MS system (Tinkerbell LT), MicroID software, a MicroID database and a disposable kit for sample treatment (Figure 4B.1).

The linear MALDI-TOF MS system Tinkerbell LT is designed to be as simple as possible for use by general laboratory personnel who do not have specialized mass spectrometry experience. The majority of the data acquisition operates automatically. Minimum maintenance, reduced maintenance costs and time are achieved with a long-life Nd YLF laser and vacuum pump with modular components. Furthermore, the system is also designed to be extended with a mass selector which offers high resolution and sensitivity in the selected mass range. The selection of positive or negative ions can also be used for the future development of small molecule pattern databases for drug resistance or sensitivity. Figure 4B.2 shows MS spectra obtained by Tinkerbell LT from four different *Mycobacterium* species. Different mass spectral profiles for each species are clearly evident; in another example, similarities between peak patterns of *M. tuberculosis* and Bacillus Calmette–Guérin (BCG) spectra are shown in Figure 4B.3.

4B.3 MicroID Software

The identification software MicroID™ uses a unique mass picking and matching method. Here, masses are picked by a machine-learning algorithm to determine the maximum independence of each peak in intra- and inter-classification clusters. The mass and intensity profiles of test samples pass through a cluster-by-cluster matching process with

Figure 4B.1 The Micro ID system comprises the ASTA Tinkerbell LT MALDI-TOF spectrometer, Micro ID software and database, and disposable kit with disposable MALDI plates, matrix, standards solution.

a pre-built reference database (DB), and the results are shown as probabilities of matching extent to potential species or strains. The software employs norm (distance) value discrimination to compute cross-correlation or matching similarity of a sample spectrum to a reference dataset. Test data are matched to the reference data from each species in the reference DB, and probabilities of matching extent are computed. Results are shown as a percentage for each specific species or strain. The algorithm employs a deep-learning technique to reduce computation time and find the optimum intensity factor and number of mass peaks. Data variations from different users are covered with this algorithm. Automatic processing and reporting functions allow for fast and convenient identification of 96 samples in a single run. After MALDI-TOF MS analysis, users can reprocess data to reconfirm results whenever it is necessary. The MicroID™ user interface is composed of three panels with three different user interfaces that are intuitive for beginners.

4B.4 Database

The MicroID™ database includes strains from the Korean Collection of Type Culture (KCTC) as well as clinical isolates from hospitals in Korea. It currently comprises 2591 species and 3905 strains and is being extended. The database is divided into specific DBs according to the application. The structure of the MicroID database is shown in

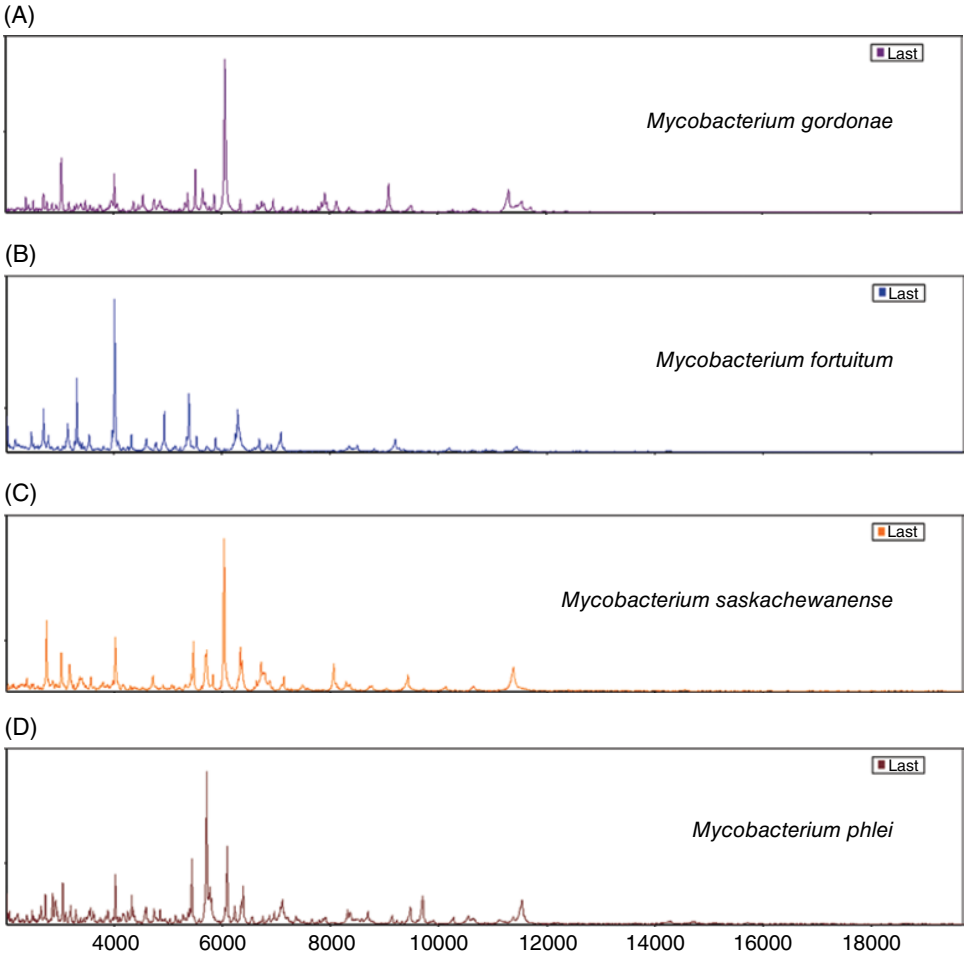


Figure 4B.2 Typical spectra obtained by MALDI-TOF-MS (Tinkerbell LT, Suwon, Korea) of *Mycobacterium* species: (A) *Mycobacterium gordonae*, (B) *Mycobacterium fortuitum*, (C) *Mycobacterium saskachewanense* and (D) *Mycobacterium phlei*.

Figure 4B.4. Each DB was created for a specific application; for example, the Core DB is for general bacteria, the Food DB for food-borne bacteria, the Clin DB for clinical applications, the Agri DB for agriculture and animals, the Environ DB for environmental applications and the MycoMP DB is for *M. tuberculosis* and NTM, which can be used independently or combined with the Clin DB.

4B.5 MycoMP Database for Mycobacteria

Table 4B.1 list the strains that comprise the MycoMP database from the KCTC (Taejeon, Korea) and reference strains obtained from clinical isolates in Korea (Table 4B.2). All clinical isolates were verified by sequencing the 16SrRNA, *rpo*B and tuf genes. Reference strains and clinical isolates were frozen at −80°C. Strains were

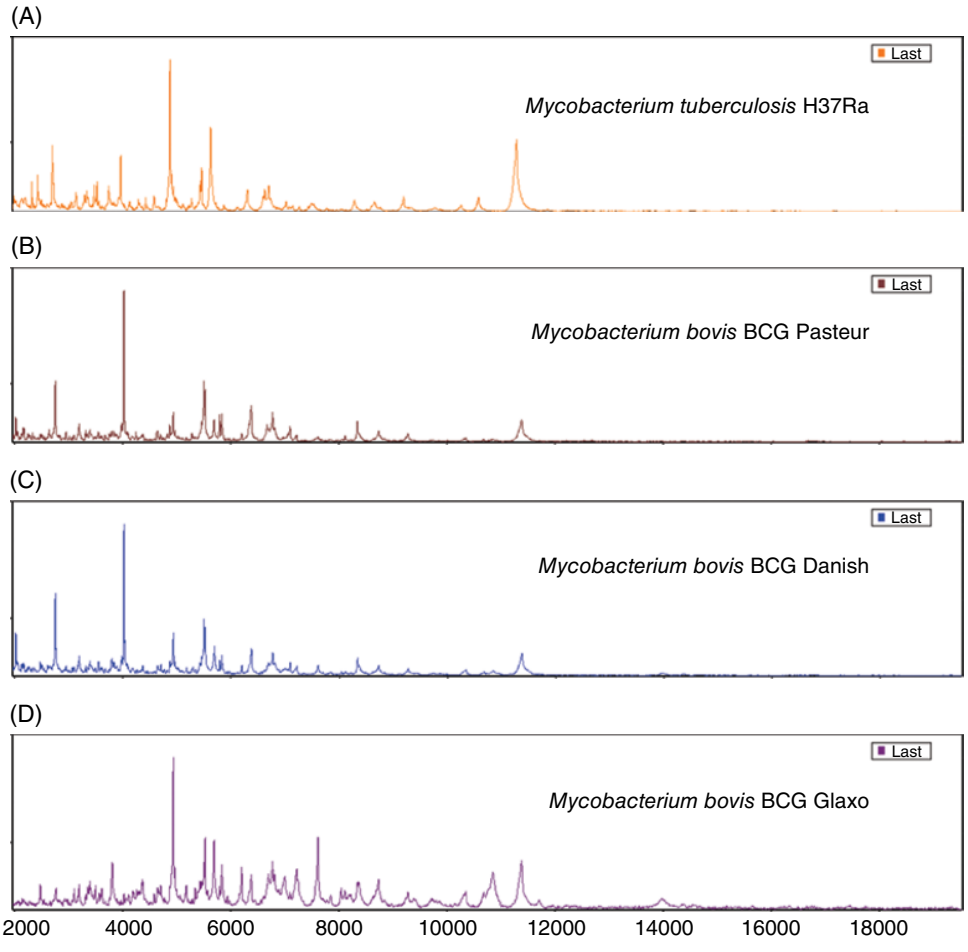


Figure 4B.3 MALDI spectra representing the *Mycobacterum tuberculosis* complex: (A) *Mycobacterium tuberculosis*, (B), (C), (D) *M. bovis* BCG.

thawed at room temperature and cultured in 2% or 3% Ogawa, Löwenstein-Jensen or Mycobacteria Growth Indicator Tubes (MGITs) to compensate for variation in the culture medium. However, differences in media composition did not have an effect on the mass spectral pattern of mycobacteria. A flowchart of the developing Myco MP DB is shown in Figure 4B.5.

Apart from the dense lipid layer that complicates protein extraction from cells and prolongs the methodology, cell inactivation is also recommended prior to MALDI-TOF MS analysis to minimize exposure of laboratory personnel to the *Mycobacterium tuberculosis* complex. The ASTA's mycobacterial protein extraction protocol was derived on the basis of extensive experiments for optimizing many variables. Using microtubes pre-filled with 1 mm glass beads, the protein extraction efficiency is maximized, and steps for protein extraction are minimized (Figure 4B.6). Briefly, the optimization procedure was as follows: one loopful of mycobacterial biomass was collected in a 2 ml screw-cap microcentrifuge tube containing 20 pills of 1 mm glass beads.

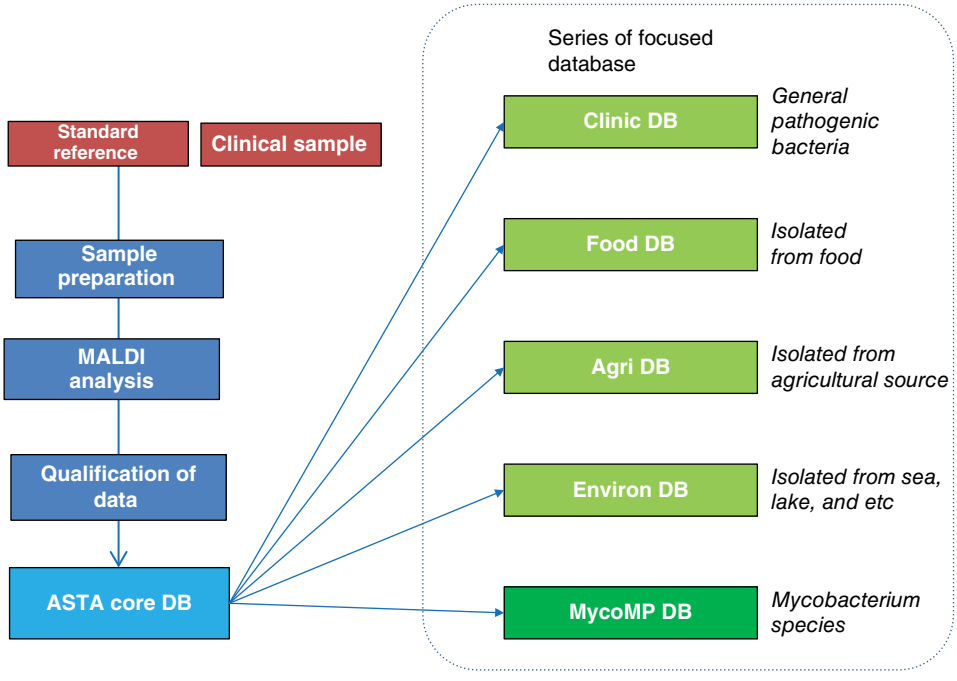


Figure 4B.4 Structure of MicroID database. Each database (DB) is dedicated to a specific application.

After the addition of 0.3 ml distilled water (Merk Millipore), the sample was boiled for 30 min to inactivate the mycobacterial cells. Following addition of a further 0.7 ml ethanol (Sigma-Aldrich), the sample was vortexed for 5 min, and then centrifuged for 5 min. The supernatant was discarded. The pellet was re-suspended in 30 µl 100% formic acid (Sigma-Aldrich) and then subjected to bead beating (FastPrep-24; MP Biomedicals) for 1 min (max speed: 6 m/s). After the addition of 30 µl 95% acetonitrile (Sigma-Aldrich), the sample was vortexed for 5 s, and then centrifuged at 3000 rpm for 3 s. Then 1µl supernatant was transferred to a MALDI target plate (ASTA Inc. Suwon, Korea).

MALDI-TOF MS spectra were obtained with 1 µl of supernatant from each test sample pipetted onto a spot on the disposable MALDI target plate, and 1 µl of the calibration standard (*E. coli* ASTA standard) onto a separate spot. The samples were treated with 2 µl of the MALDI matrix (a freshly prepared saturated solution of α-Cyano-4-hydroxycinnamic acid in 50% ACN and 2.5% trifluoroacetic acid). Spectra were acquired over a mass range from 2,000 to 20,000 Da, and 1,000 laser shots over 20 sites on each sample were summed.

The correlation of *Mycobacterium* spp. in the MycoMP DB was analyzed using MALDI-TOF mass spectra in the MycoMP DB (Figure 4B.7). The MycoMP DB was used to evaluate 82 mycobacterial isolates and showed 92.7% matching accuracy. *M. tuberculosis* and BCG are clearly distinguished from others. This DB is currently being tested and expanded with clinical samples. Strains from ATCC are being used for validation of the database. For more validation, comparison with other databases such as Biotyper 3.0 DB is required.

Table 4B.1 The type strains of *Mycobacteriium* spp. from KCTC (Korean Collection of Type Cultures) used for the MycoMP database.

No.	Species	KCTC strain no.	No.	Species	KCTC strain no.	No.	Species	KCTC strain no.
1	*M. abscessus*	19621	26	*M. farcinogenes*	19647	51	*M. parascrofulaceum*	9979
2	*M. acapulcensis*	9501	27	*M. fortuitum subsp. fortuitum*	9510	52	*M. paraseoulense*	19145
3	*M. agri*	9502	28	*M. fortuitum subsp. fortuitum*	1122	53	*M. paraterrae*	19556
4	*M. alvei*	19709	29	*M. fortuitum subsp. fortuitum*	29797	54	*M. phlei*	2192
5	*M. asiaticum*	9503	30	*M. frederiksbergense*	19100	55	*M. phlei*	3037
6	*M. aubagnense*	19639	31	*M. gallinarum*	9511	56	*M. phlei*	9689
7	*M. aubagnense*	29645	32	*M. gilvum*	19423	57	*M. porcinum*	9517
8	*M. austroafricanum*	9504	33	*M. gordonae*	9513	58	*M. pulveris*	9518
9	*M. abscessus sp. bolletii*	19281	34	*M. heckeshornense*	19648	59	*M. rufum*	29649
10	*M. botniense*	19646	35	*M. holsaticum*	19650	60	*M. rufum*	29650
11	*M. brisbanense*	19641	36	*M. immunogenum*	19643	61	*M. sakatchewanense*	9978
12	*M. brumae*	19711	37	*M. interjectum*	19649	62	*M. salmoniphilum*	29801
13	*M. canariasense*	19644	38	*M. intracellulare*	9514	63	*M. sediminis*	19999
14	*M. celatum*	19714	39	*M. kansasii*	9515	64	*M. senuense*	19147
15	*M. chelonae subsp. chelonae*	9505	40	*M. komossense*	29798	65	*M. seoulense*	19146
16	*M. chelonae*	29796	41	*M. koreense*	19819	66	*M. septicum*	29802
17	*M. chlorophenolicum*	19089	42	*M. manitobense*	9977	67	*M. smegmatis*	9108
18	*M. chubuense*	19712	43	*M. massiliense*	19086	68	*M. sp*	1466
19	*M. conceptionense*	19640	44	*M. moriokaense*	9516	69	*M. sp*	1829

20	*M. conceptionense*	39499	45	*M. moriokaense*	29799	70	*M. szulgai*	9520
21	*M. cookii*	19715	46	*M. mucogenicum*	19088	71	*M. terrae*	9614
22	*M. cosmeticum*	19713	47	*M. neworleansense*	29800	72	*M. vaccae*	19087
23	*M. diernhoferi*	9844	48	*M. neoaurum*	19096	73	*M. vanbaalenii*	9966
24	*M. diernhoferi*	9506	49	*M. obuense*	19097	74	*M. yongonense*	19555
25	*M. fallax*	9508	50	*M. parakoreense*	19818	74	*M. tuberculosis*	H37Ra

Table 4B.2 103 Clinical strains of Mycobacteria belonging to 25 *Mycobacteriium* species.

No.	Species		No.	Species		No.	Species	
1	*M. abscessus*	13	10	*M. gordonae*	4	19	*M. mucogenicum*	2
	M. abscessus ssp. bolletii	8	11	*M. indicuspranii*	1	20	*M. neoaurum*	3
2	*M. algericum*	1	12	*M. intracellulare*	12,1	21	*M. paragordonae*	1
3	*M. alvei*	1	13	*M. iranicum*	1	22	*M. peregrinum*	5
4	*M. avium*	13	14	*M. kansasii*	6,1		*M. peregrinum*	2
5	*M. chelonae*	3	15	*M. kumamotonense*	1	23	*M. phocaicum*	2
6	*M. chimaera*	6	16	*M. kyorinense*	1,1	24	*M. porcinum*	2
7	*M. colombiense*	2	17	*M. lentiflavum*		25	*M. yongonense*	1
8	*M. conceptionense*	2	18	*M. massiliense*	1		Total	121
9	*M. fortuitum*	5						

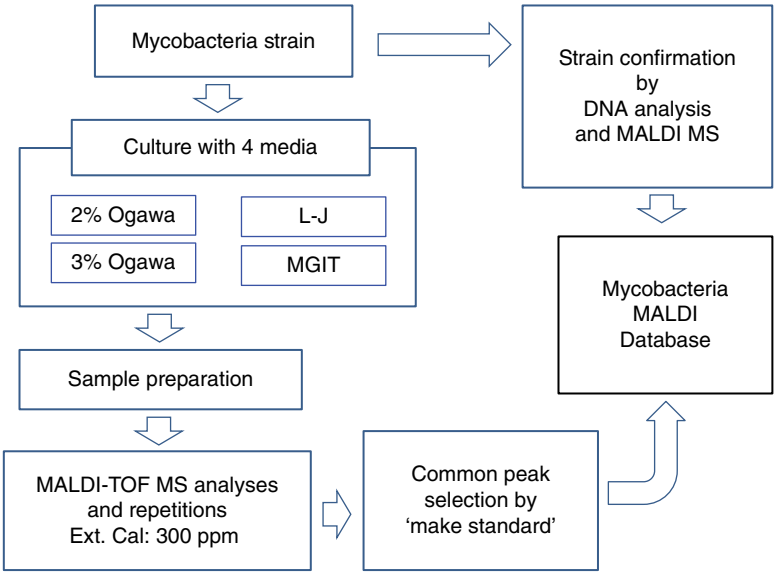


Figure 4B.5 The workflow for developing a mycobacterial MALDI-TOF MS database.

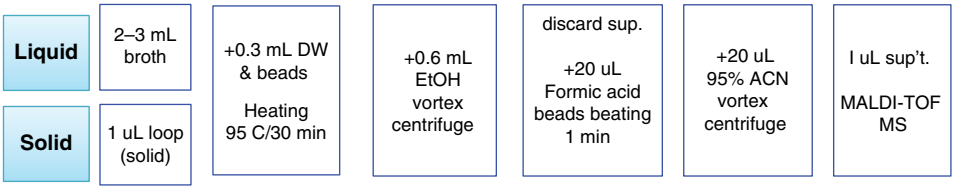


Figure 4B.6 Stages in the preparation of a mycobacterial sample for MALDI-TOF MS analysis.

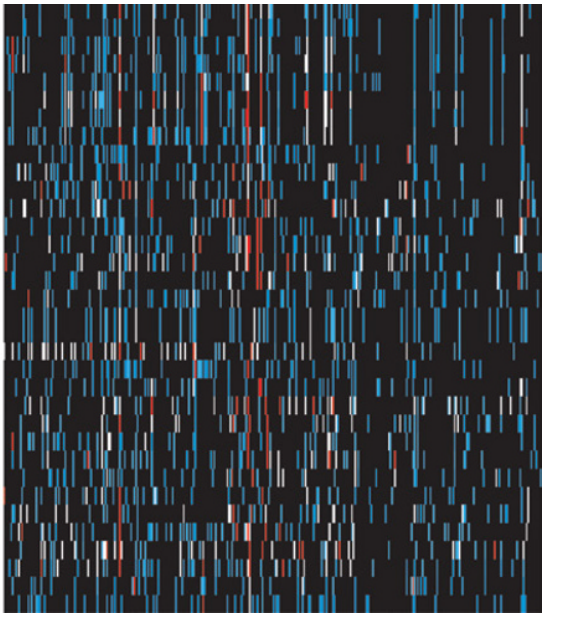


Figure 4B.7 Cluster analysis of selected mass spectra of *Mycobacterium* spp. from the MycoMP database using Perseus (Computational Systems Biochemistry, Germany). The inter-relatedness of the *Mycobacteriium* species is illustrated and *M. tuberculosis* is distinguished from other mycobacteria.

4B.6 Conclusion

We have established that this new system has the capacity to identify most mycobacterial strains to the species level using MALDI-TOF MS, while the software can differentiate the intrinsic similarities from small differences in mass profiles among species. Due to tremendous advances in the development of identification software technology, improvements in MALDI-TOF MS resolution and the simplicity and speed of sample preparation, it is anticipated that MALDI-TOF MS will be used routinely for mycobacterial identification in most clinical laboratories in the near future.

References

Balážová, T., Makovcová, J., Šedo, O., Slaný, M., Faldyna, M. and Zdráhal, Z. (2014). The influence of culture conditions on the identification of *Mycobacterium* species by MALDI-TOF MS profiling. *FEMS Microbiol. Lett.* 353, 77–84.

Claydon, M. A., Davey, S. N., Edwards-Jones, V. and Gordon, D. B. (1996). The rapid identification of intact microorganisms using mass spectrometry. *Nat. Biotechnol.* 14, 1584–1586.

Day, C. L., Moshi, N. D. Abrahams, D. A., van Rooyen, M., O'rie, T., de Kock, M., and Hanekom, W. A. (2014). Patients with tuberculosis disease have *Mycobacterium tuberculosis*-specific CD8 T cells with a pro-apoptotic phenotype and impaired proliferative capacity, which is not restored following treatment. *PLoS ONE* (online journal). April 16. https://doi.org/10.1371/journal.pone.0094949

El Khéchine, A., Couderc, C., Flaudrops, C., Raoult, D. and Drancourt, M. (2011). Matrix-assisted laser desorption/ionization time-of-flight practice. *PLoS ONE* (online journal) 6(9), e24720. doi:10.1371/journal.pone.0024720.

Griffith, D. E. (2015). *Mycobacterium*: General characteristics, laboratory detection, and staining procedures. In Jorgensen, J. H., Pfaller, Carroll, K. C., Funke, G., Landry, M. L., Richer, S. S. and Warnock, D. W. (Eds.), *Manual of Clinical Microbiology* (11th ed., pp. 536–569). ASM Press, Washington, DC.

Griffith, D. E., Aksamit, T., Brown-Elliott, B. A., Catanzaro, A., Daley, C., Gordin, F., Holland, S. M., Horsburgh, R., Huitt, G., Iademarco, M. F., Iseman, M., Olivier, K., Ruoss, S., von Reyn, C. F., Wallace, R. J., Jr. and Winthrop, K. (2011). ATS Mycobacterial Diseases Subcommittee, American Thoracic Society, Infectious: Diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am. J. Respir. Crit. Care Med.* 175, 367–416.

Hettick, J. M., Kashon, M. L., Simpson, J. P., Siegel, P. D., Mazurek, G. H. and Weissman, D. N. (2004). Proteomic profiling of intact mycobacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.* 76, 5769–5776.

Hoefsloot, W., van Ingen, J., Andrejak, C., Angeby, K., Bauriaud, R., Bemer, P. *et al*. (2013). The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: An NTM-NETcollaborative study. *Eur. Respir. J.* 42, 1604–1613.[https://en.wikipedia.org/wiki/Mycobacterium]

Johnson, M. M. and Odell, J. A. (2014). Nontuberculous mycobacterial pulmonary infections. *J. Thorac. Dis.* 6, 210–220.

Kim, B. J., Lee, S. H., Lyu, M. A., Kim, S. J., Bai, G. H., Chae, G. T., Kim, E. C., Cha, C. Y. and Kook, Y. H. (1999). Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB). *J. Clin. Microbiol.* 37, 1714–1720.

Kim, T. S., Lee, K., Hong, Y. J., Hwang, S. M., Park, J. S., Park, K. U., Song, J. and Kim, E. C. (2015). MALDI-TOF MS: Its application in the clinical laboratory and a paradigm shift in clinical microbiology. *Lab. Med. Online.* 5, 176–187.

Lee, H., Park, H. J., Cho, S. N., Bai, G. H. and Kim, S. J. (2000). Species identification of mycobacteria by PCR-restriction fragment length polymorphism of the rpoB gene. *J. Clin. Microbiol.* 38, 2966–2971.

Lotz, A., Ferroni, A., Beretti, J.-L., Dauphin, B., Carbonnelle, E., Guet-Revillet, H., Veziris, N., Heym, H., Jarlier, V., Gaillard, J.-L., Pierre-Audigier, C., Frapy, E., Berche, P., Nassif, X. and Bille, E. (2010). Rapid identification of mycobacterial whole cells in solid and liquid culture media by matrix-assisted laser desorption ionization–time of flight mass spectrometry. *J. Clin. Microbiol.* 48, 4481–4486.

Mäkinen, J., Marjamaki, M., Marttila, H. J. and Soini, H. (2006). Evaluation of a novel strip test, GenoType Mycobacterium CM/AS, for species identification of mycobacterial cultures. *Clin. Microbiol. Infect.* 12, 481–483.

Mather, C. A., Rivera, S. F. and Butler-Wu, S. M. (2014). Comparison of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of mycobacteria using simplified protein extraction protocols. *J. Clin. Microbiol.* 52, 130–138.

Pignone, M., Greth, K. M., Cooper, J., Emerson, D. and Tang, J. (2006). Identification of mycobacteria by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. *J. Clin. Microbiol.* 44, 1963–1970.

Saleeb, P. G., Drake, S. K., Murray, P. R. and Zelazny, A. M. (2011). Identification of mycobacteria in solid-culture media by matrix assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 49, 1790–1794.

Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P. E., Rolain, J. M. *et al.* (2009). Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Infect. Dis.* 49, 543–551.

Wang, J., Chen, W. F. and Li, Q. X. (2012). Rapid identification and classification of *Mycobacterium* spp. using whole-cell protein barcodes with matrix assisted laser desorption ionization time of flight mass spectrometry in comparison with multigene phylogenetic analysis. *Anal. Chim. Acta* 716, 133–137.

Woo, P. C., Lau, S. K., Teng, J. L., Tse, H. and Yuen, K. Y. (2008). Then and now: Use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin. Microbiol. Infect.* 14, 908–934.

World Health Organization (WHO) (2015). Global Tuberculosis Report 2015. WHO Press, Geneva, Switzerland. WHO/HTM/TB/2015.

5

Transformation of Anaerobic Microbiology since the Arrival of MALDI-TOF Mass Spectrometry

Elisabeth Nagy,[1] Mariann Ábrók,[1] Edith Urbán,[1] A.C.M. Veloo,[2] Arie Jan van Winkelhoff ,[2] Itaru Dekio,[5] Saheer E. Gharbia[4] and Haroun N. Shah[3]

[1] *Institute of Clinical Microbiology, Faculty of Medicine, University of Szeged, Szeged, Hungary*
[2] *University Medical Center, Groningen, Department of Medical Microbiology, Groningen, The Netherlands*
[3] *Proteomics Research, Public Health England, London, UK*
[4] *Genomics Research Unit, Public Health England, London, UK*
[5] *Medical Centre East, Tokyo Women's Medical University, Tokyo, Japan*

5.1 Introduction

Interest in anaerobic bacteria over the last century have been recurrent, driven very much by new technologies and restricted to a few highly specialized laboratories. However, anaerobic species have been reported since the late nineteenth century (see, e.g. Veillon and Zuber, 1898) and have been an integral part of *Bergey's Manual* from its inception (Bergey, 1923). In particular, considerable attention has been given to gram-positive clostridial spore-formers because of the severity of diseases they cause. By contrast, gram-negative species, which have long been regarded as the dominant taxa of the colon, have been largely overlooked until the 1970s. Prior to this, the most exhaustive studies were undertake at the Pasteur Institute, where Prévot and colleagues carried out detailed analysis of pathogenicity of strains while simultaneously employing standard organic chemistry methods to determine their products of metabolism to facilitate characterization of species (see, e.g. Prévot, 1938). The arrival of gas chromatography (GC) in the 1970s simplified these methods, and microbiological laboratories began acquiring these en masse. Their introduction in clinical laboratories had an enormous impact on interest in anaerobes, and centres such as the Virginia Polytechnic Institute and State University, USA, were among the first to employ GCs to profile the volatile fatty acids of anaerobes to enable their routine identification (Holdeman *et al.*, 1977). This, together with the arrival of anaerobic cabinets for large-scale culture, triggered an enormous upsurge of interest in anaerobes, and several official societies were established to promote work on anaerobes. The United Kingdom's 'Anaerobic Discussion Group' was established in 1978 and later became the 'Society for Anaerobic Microbiology'.

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

Its conference proceedings were already being published by 1979 (Shah and Hardie, 1979), after which biennial conferences were held to stimulate further work on anaerobes. In 1992, two of the authors (HNS, SEG) relocated to Dalhousie University, Halifax, Canada, and founded the 'Canadian Society of Anaerobic Microbiology' in February 1992. This was established with strong links to Professor Hebe M. Bianchini's laboratory in Argentina (Laboratorio Microbiología, Centro de Educación Médica e Investigaciónes Clínicas, Buenos Aires). A chance visit to Halifax by Professor Sydney Finegold around the same period swiftly led to its extension with the United States and South America to form the 'Anaerobic Society of the Americas' the same year. These societies vigorously promoted anaerobic microbiology. In particular, increasing resistance to metronidazole, the antibiotic reserved for treatment of anaerobic infections, spurred on much work on resistance mechanisms. At the same time, new technologies in mass spectrometry such as pyrolysis mass spectrometry (MS), electron impact and fast atomic bombardment MS were applied to study the systematics of gram-negative taxa such as the *Bacteroidaceae* which eventually led to substantial taxonomic revisions in the 1980s (see e.g., Shah and Collins, 1980, 1983, 1988–1990; Magee *et al.*, 1989; Tavana *et al.*, 1998).

Many of these taxonomic proposals were confirmed with the arrival of 16S rRNA (see e.g., Lawson *et al.*, 1989, 1991) and which in subsequent years continued to shape the systematics of this group. However, by the late 1990s to early 2000, interest in anaerobes declined considerably because much of the equipment used to identify anaerobes routinely such as gas chromatographs, mass spectrometers and spectrophotometers were no longer in general use in microbiology laboratories. The almost total reliance on comparative 16S rRNA analysis presented a formidable task for clinical laboratories that depended on phenotypic tests to characterize clinical isolates. This was exacerbated by the absence of reliable characters for defining species, especially because many of the newly proposed taxa such as *P. gingivalis, P. endodontalis and P. asaccharolytica* were non-fermentative (see, e.g. Shah and Collins, 1988).

The arrival of MALDI-TOF MS had an immediate, significant impact on anaerobic microbiology largely because among the first microorganisms to be studied were members of the fastidious, poorly characterized and obscure members of the genus *Porphyromoas* (see Chapter 1; Shah *et al.*, 2002). These were analyzed to assess the resolution of MALDI-TOF MS and to establish proof of concept of this new and novel technology. Soon the basic parameters for a reproducible method were published (Shah *et al.*, 2000), and assembly of the first comprehensive database of MALDI-TOF mass spectral profiles reported (Keys *et al.*, 2004) and proof of concept established (Shah, 2005). In the ensuing years, many laboratories reported extensive studies on the characterization of anaerobic species using MALDI-TOF MS, and its implementation in clinical laboratories has been transformative. Whereas a few years ago, meetings on anaerobes were poorly attended, the arrival of MALDI-TOF MS has led to significant rekindling of interest in anaerobes.

This chapter reports work on the use of MALDI-TOF MS for the characterization of clinically important anaerobes to the species level from clinical samples, efforts to improve the coverage of anaerobic species by expanding the database and the potential

of the technique to type specific human pathogens. Specific reference is made here to the identification of *Bacteroides* and other clinically relevant anaerobes, where the impact has been considerable.

5.2 Identification in the Clinical Laboratory

Since the development of commercially available equipment for oxygen-free culture of bacteria from clinical specimens, there has been a significant increase in information relating to infections caused by non-spore-forming anaerobes. Apart from the traditionally studied sites, where an abundant anaerobic flora is present on mucosal surfaces, such as in intraabdominal or pelvic infections and infections in connection with the oral cavity, a great variety of other infections were proven to be caused by strict anaerobic bacteria alone or in combination with other microaerophilic or facultative anaerobes. The main problems for routine laboratories to deal with the diagnosis of anaerobic infections are the difficulties to identify these microorganisms in real time, due to their slow growth even in good anaerobic environment and the need to confirm their strict anaerobic nature by subculturing (Jousimies-Somer *et al.*, 2002). Furthermore, quite a few clinically important anaerobic bacteria are biochemically inactive. Accordingly, their identification by commercially available kits, which need a 24–48-h incubation in an anaerobic environment, such as the API20A (bioMérieux, Mercy-l'Etoile, France), or detection of preformed enzymes within 4–6 h, such as RapID ANA II (Remel, KS, USA), Rapid 32A (bioMérieux) and VITEK ANI card (bioMérieux), is not always possible. On the other hand, these identification kits have relatively narrow spectrum databases for anaerobes and usually require large amounts of inoculum, which is difficult to obtain with anaerobes producing very small colonies even after a long incubation time.

With the arrival of MALDI-TOF MS into the routine of clinical microbiology laboratories, it became evident that it will be much easier to identify a great variety of bacteria and yeasts at the species level by comparing the mass spectra of the unknown isolates with reference spectra of known species included in the database. Routine identification of the most frequently isolated anaerobes, such as clostridia and *Bacteroides* strains, also became available very early, based mainly on the mass spectra of the conserved ribosomal proteins. In early studies (Shah *et al.*, 2002; Stingu *et al.*, 2008), it was already seen that even members of the genera *Prevotella* (*P. nigrescens* and *P. intermedia*) or *Clostridium* (*C. septicum* and *C. chauvei*), which are difficult to differentiate, could be discerned by their MALDI-TOF mass spectra with an efficacy similar to those of the DNA-based techniques.

Two systems are available today for routine identification of clinical isolates, the MALDI Biotyper (Bruker Daltonik GmbH, Bremen, Germany), which utilizes the Microflex or Autoflex MS with the MALDI Biotyper database (versions: research use only, RUO; and in-vitro diagnostic, IVD), and the VITEK MS system (bioMérieux, Marcy-l'Etoile, France), which uses the Shimadzu Axima MS with an IVD database and SARAMIS as the RUO database. The two systems differ in the requirements for sample preparation, in the interpretation of the measured mass spectra and the databases of the microorganisms. The performance of these equipments and their databases were tested by many reference and routine laboratories dealing with everyday testing of anaerobic specimens.

5.3 Pre-analytical Requirements Influence Species Identification of Anaerobic Bacteria

Despite the simple usage of MALDI-TOF MS instruments developed for routine identification of clinical isolates, the pre-analytical steps influence the quality of the mass spectra and the correct identification. Most studies using the MALDI Biotyper agree that the composition of the culture media does not affect the identification of anaerobes (Grosse-Herrenthey *et al.*, 2008; Federko *et al.*, 2012; Hsu and Burnham, 2014). However, at least in the case of *Clostridium* spp., the extended incubation time may lead to deviation in the fingerprint patterns due to extensive sporulation (Grosse-Herrenthey *et al.*, 2008). Application of different complete or selective media used by routine laboratories to isolate anaerobic bacteria, such as sheep blood agar, colistin nalidixic acid agar, *Brucella* laked blood agar with kanamycin and vancomycin and *Bacteroides* bile esculin agar, did not influence the identification of 28 anaerobic clinical isolates representing 16 species (Hsu and Burnham, 2014).

The performance of MALDI-TOF MS not only depends on the quality of the database of the system but also on the quality of the spectrum obtained for a given isolate. The proteins detected with the current method for bacterial identification using MALDI-TOF MS are predominantly ribosomal proteins (Arnold and Reilly, 1999). Consequently, the quality of a spectrum is optimal when bacterial cells are in the exponential phase of growth and is influenced by the extraction method used. The on-target extraction by 70% formic acid and full extraction by formic acid and acetonitrile both aim to disrupt the bacterial cell envelope to release the proteins. Consequently, the amount of bacteria spotted on the target and the homogeneity of the smear across the surface will also affect the reproducibility of the spectrum. Insufficient amount of bacteria results in poorly resolved mass ions, whereas too large a mass of cells yields only the prominent peaks and saturation at the detector (De Bruyne *et al.*, 2011). Studies have shown that a heavy smear combined with an on-target extraction by 70% formic acid provides the best results (Williams *et al.*, 2003; Ford and Burnham, 2013; McElvania *et al.*, 2013). However, "a heavy smear" is difficult to define. Veloo *et al.* (2014) assessed some of the parameters needed for high-quality spectra of anaerobic bacteria. Compared to aerobic bacteria, many anaerobic species show slower growth, and for some species colonies are small or rough, which make it difficult to obtain a good-quality, homogeneous preparation. The best results are obtained when the MALDI-TOF MS measurement is performed after 48 h of incubation in an anaerobic environment and when bacterial cells are spotted on the same day and are not exposed to oxygen for more than 24 h (see Table 5.1). The influence of oxygen exposure on MALDI-TOF MS performance was also investigated by Hsu and Burnham (2014). After 1 d and 5 d of exposure to oxygen, 71% and 80%, respectively, of the tested anaerobic bacteria were correctly identified. However, the identification score obtained after 1 d was higher than after 5 d of exposure to oxygen. Veloo *et al.* (2014) observed that for *Fusobacterium necrophorum* a mass spectrum was not obtained after 24 h of exposure to oxygen nor from *Prevotella intermedia* after 48 h of exposure. They attributed this to cell wall damage caused by oxygen exposure, causing the leakage of proteins out of the cell. For all other tested gram-negative and gram-positive anaerobic bacteria, a good-quality spectrum was still obtained after 48 h of exposure to oxygen, with no decrease in identification score.

Table 5.1 The influence of exposure to oxygen on the quality of the MALDI-TOF MS spectrum.

	Exposure to oxygen (h)				
Species	**0**	**1**	**6**	**24**	**48**
Gram-negative bacteria					
Bacteroides thetaiotaomicon	2.14	2.24	2.10	2.17	2.35
Bacteroides stercoris	2.19	2.14	2.18	2.25	2.27
Parabacteroides johnsonii	2.23	2.32	2.29	2.31	2.28
Fusobacterium necrophorum	2.08	2.30	2.19	<1.7[a]	<1.7[a]
Fusobacterium nucleatum[b]	2.08	2.27	2.32	2.19	2.08
Prevotella intermedia	1.92	1.93	1.86	1.95[c]	<1.7[a]
Prevotella oris	2.08	2.17	2.01	2.35	2.29
Alistipes onderdonkii	2.21	2.24	2.29	2.29	2.21
Veillonella parvula	2.25	2.13	2.2	2.20	2.16
Gram-positive bacteria					
Finegoldia magna	2.24	2.05[c]	2.34	2.52	2.50
Peptoniphilus harei	2.15	2.00	2.13	2.23[b]	2.19
Peptoniphilus ivorii	<1.7[d]	1.81[e]	<1.7[d]	1.76[e]	1.80[e]
Clostridium hathewayii	2.36	2.23	2.29	2.07	2.19
Clostridium ramosum	2.03	2.19	2.27	2.06	2.17
Actinomyces graevenitzii	2.05[c]	2.03[e]	2.11	2.06[e]	<1.7[d]
Actinomyces meyeri	2.03	2.27	2.35	2.19	2.25
Bifidobacterium longum	2.00[c]	2.12[c]	1.99[c]	2.15	2.17
Propionibacterium acnes	2.18	2.14[c]	2.12	2.13	2.13
Eggerthella lenta	2.11	2.30	2.27	2.16	2.02[e]
Collinsella aerofaciens	2.26[c]	2.16	2.23[c]	2.32[c]	2.28[c]

[a] With direct spotting and on-target extraction, no reliable identification or peaks were obtained.
[b] The MALDI-TOF MS system identifies *F. nucleatum* as either *F. nucleatum* or *F. naviforme*.
[c] On-target extraction with 70% formic acid.
[d] With all three sample preparations methods, no reliable identification or peaks were obtained.
[e] Full extraction.
Adapted from: Veloo *et al.* (2014). The influence of incubation time, sample preparation and exposure to oxygen on the quality of the MALDI-TOF MS spectrum of anaerobic bacteria. *Clin. Microbiol. Infect.* **20**: O1091–O1097.

Among the different techniques, the on-target extraction by 70% formic acid was found the most successful for gram-positive anaerobic bacteria (Hsu and Burnham, 2014; Veloo *et al.*, 2014). In general, the method of spotting the cells has a significant impact on the quality of the spectrum obtained and, as expected, this is influenced by the technical experience of the staff and also the nature of bacterial colony being spotted. In Table 5.2, it is shown that an experienced scientist produces better and more consistent results (Veloo *et al.*, 2014). However, the inherent nature of the sample also

Table 5.2 The range of log scores and identification results of spotting the same strain ten times by two different examiners.

	Less experienced examiner					Experienced examiner				
Strain	**Range**	**SD**	**No reliable ID (n)[a]**	**Genus ID (n)**	**Species ID (n)**	**Range**	**SD**	**No reliable ID (n)a**	**Genus ID (n)**	**species ID (n)**
Gram-negative bacteria										
B. thetaiotaomicron	2.031–2.198	0.052	0	10	10	2.080–2.257	0.052	0	10	10
P. intermedia	1.958–2.121	0.050	0	10	8	1.976–2.092	0.042	0	10	8
F. necrophorum	2.263–2.387	0.050	0	10	10	2.187–2.418	0.072	0	10	10
F. nucleatum	1.596–2.108	0.152	1	9	4	1.192–2.107	0.269	1	9	5
C. ureolyticus	0.962–2.012	0.859	8	2	1	1.700–2.043	0.826	2	8	2
V. parvula	2.306–2.402	0.039	0	10	10	2.182–2.416	0.071	0	10	10
Gram-positive bacteria										
P. micra	1.687–2.338	0.233	1	9	6	2.202–2.399	0.058	0	10	10
F. magna	0.959–2.116	0.344	1	9	5	1.831–2.108	0.075	0	10	7
P. ivorii[b]	1.255–1.743	0.162	7	3	0	1.162–1.637	0.137	10	0	0
A. minutum[b]	1.831–2.528	1.244	4	6	5	2.183–2.481	0.105	0	10	10
C. butyricum	2.054–2.244	0.070	0	10	10	2.009–2.265	0.086	0	10	10
A. israellii[b]	1.114–1.314	0.639	10	0	0	1.968–2.006	0.838	8	2	1
A. graevenitzii[b]	1.338–2.191	0.926	5	5	2	1.851–2.250	1.082	4	6	4
A. meyeri[b]	1.946–2.239	1.116	6	4	3	1.308–2.217	0.961	6	4	2
B. dentium[b]	1.912–2.304	0.908	2	8	7	2.163–2.388	0.971	2	8	8
B. longum[b]	1.780–2.229	0.150	0	10	6	2.018–2.218	0.071	0	10	10
P. acnes[b]	2.014–2.177	0.713	1	8	8	2.058–2.316	0.088	0	10	10
E. lenta[b]	1.263–1.851	0.833	8	2	0	2.104–2.311	0.929	2	8	8

[a] A log score <1.7 or no peaks.
[b] An on-target extraction was performed using 70% formic acid.
Adapted from: Veloo *et al.* (2014). The influence of incubation time, sample preparation and exposure to oxygen on the quality of the MALDI-TOF MS spectrum of anaerobic bacteria. *Clin. Microbiol. Infect.* **20**: O1091–O1097.

affects the mass spectrum. For example, species that produce tiny colonies such as *Campylobacter ureolyticus* or species that characteristically yield a rough, dry colony morphology such as *Actinomyces israellii* often produce poor mass spectra.

5.4 Recent Database Developments for Anaerobes

In routine laboratories, the most frequent anaerobic isolates, besides clostridia and a great variety of gram-positive anaerobic cocci (GPACs), are the members of the gram-negative anaerobic genera *Bacteroides*, *Parabacteroides, Prevotella* and *Porphyromonas*. Recent taxonomic changes in these genera are not fully concordant with the commercially available phenotypic identification kits and, accordingly, certain species, such as *Bacteroides nordii, Bacteroides dorei, Parabacteroides gordonii, Parabacteroides johnsonii, Prevotella heparinolytica* and *Prevotella zoogleoformans*, cannot be identified by them (Watanabe *et al.*, 2010; Könönen *et al.*, 2015). Attention switched to the potential use of MALDI-TOF MS. In some cases, database developments were performed in collaboration with a company or carried out by reference laboratories by incorporating the mass spectra of the newly characterized species into the original database.

A large collection of *Bacteroides/Parabacteroides* clinical isolates was involved in an antibiotic surveillance study during 2008–2009 (Nagy *et al.*, 2011a). The strains were collected from different countries of Europe to determine species-specific resistance differences among the isolates. Besides using phenotypic methods, species determination was also carried out using the MALDI Biotyper (version 3.0) (Nagy *et al.*, 2009). Applying the full formic acid/acetonitrile extraction for sample preparation, 97.5% of the 277 strains were correctly identified to the species level (log score $\geq$ 2.000). This result indicated a rather advanced database for the genus *Bacteroides/Parabacteroides* even at that time. In case of discrepant results, 16S rRNA gene sequencing was used, which confirmed the MALDI-TOF MS results in each case. There were only seven isolates which could not be identified by the MALDI Biotyper, and the sequencing data showed that they were missing from the database. After the inclusion of a reference spectrum of one of the four *Parabacteriodes distasonis* isolates into the database, the three other isolates were also identified with high log scores (>2.500). Furthermore, spectra of *Bacteroides eggerthii*, *Bacteroides goldsteinii* and *Bacteroides intestinalis* isolates determined by 16S rRNA sequencing were also added to the Biotyper database (Nagy *et al.*, 2009).

Prevotella is another important gram-negative anaerobic genus. After its description and the transfer of 16 species from the genus *Bacteroides* (Shah and Collins, 1990), several new species (up to 44) were added to it. Most of these species have been described since 2004 (Shah *et al.*, 2009), and a majority of them may be involved in a variety of human infections. Species-level identification of *Prevotella* strains by the classical methods is challenging, time consuming and often leads to erroneous results. Using the MALDI Biotyper (Reference Library 3.2.1.0), which included the spectra of only 20 *Prevotella* species at that time, Wybo *et al.* (2012) could identify only 62.7% of 102 *Prevotella* clinical isolates at the species level and 73.5% at the genus level. The commercial database was extended in-house with the spectra of 23 further *Prevotella* reference strains, increasing the number of the identifiable species to 33. This improved the species- and genus-level identification of the 102 clinical isolates from 62.7% to 83.3%

and from 73.5% to 89.2%, respectively. Following the addition of a sequenced clinical isolate of *Prevotella heparinolytica*, the other two isolates of the collection could also be identified (Wybo *et al.*, 2012).

Among gram-positive bacteria, *Clostridium* species are clinically the most important anaerobes. Although some species can be easily identified, determination of others may be difficult (Jousimies-Somer *et al.*, 2002). During early development of the MALDI Biotyper database, 64 *Clostridium* isolates belonging to 31 species (most of them originating from different type culture collections) were used. A further 28 clinical and environmental *Clostridium* isolates, identified by biochemical tests and sequencing, were added to improve the discriminatory power of MALDI-TOF MS for this genus (Grosse-Herrenthey *et al.*, 2008). Using this developed database, 25 environmental *Clostridium* isolates could be correctly identified by the MS method compared with sequencing or biochemical tests. Chean *et al.* (2014) recently compared VITEK MS and MALDI Biotyper for the identification of clinically relevant *Clostridium* isolates evaluating also the impact of the sample preparation and the completeness of the databases. They found that out of 52 blood culture isolates belonging to 10 *Clostridium* spp. identified by sequencing, VITEK MS identified 47 (90.4%) isolates at the species level using the 'direct transfer' sample preparation method, whereas MALDI Biotyper identified all 52 isolates using the 'extended direct transfer method', adding formic acid to the samples on the target plate. In another study, there was only 1 of 66 clinical *Clostridium* isolates representing 12 species which could not be identified at the species level by MALDI Biotyper. The 16S rRNA sequence of this isolate showed a 99.9% identity to that of *Clostridium hathewayi*, which was not present in the database of the Biotyper. After constructing the main spectra using the full chemical extraction of this strain, it was included to the database of the MALDI Biotyper (Nagy *et al.*, 2012).

GPACs including 13 genera and at least 33 species are among those anaerobic species which are difficult to identify by conventional methods. They are usually inactive in biochemical tests, and several recent taxonomic changes make their identification even more difficult. As early as 2000, MALDI-TOF MS was explored to study the taxonomy of the group (see Murdoch and Shah, 1999; Murdoch *et al.*, 2000; Rajendram, 2003). Database development for these bacteria was carried out using the AXIMA (Shimadzu) MALDI-TOF MS equipment (Veloo *et al.*, 2011a). In this study, 12 sequenced reference strains representing 12 species and six genera and 77 clinical isolates identified by classical methods and sequencing were involved. The SARAMIS database was developed for better identification of GPACs by adding the species-specific identifying spectra, the so-called SuperSpectra, of all these isolates. The performance of the constructed database was tested using 107 clinical isolates. Ninety-six of them (90%) could be identified to the species level. However, 3 of 32 *Finegoldia magna*, 1 of 3 *Peptoniphilus ivorii* and 1 of 2 *Anaerococcus vaginalis* could not be determined when direct transfer sample preparation was applied.

Actinomyces spp. are components of the normal flora and frequently isolated on anaerobically incubated media. Identification of these gram-positive anaerobic and aero-tolerant bacteria is difficult, and MS may be a great help for routine laboratories. The number of described species has increased during recent years, and about 20 of them are relevant for human medicine. Recently, a database for MALDI Biotyper was developed in-house by Stingu *et al.* (2015). Mass spectra obtained from 11 reference strains and 140 sequenced clinical isolates were used to create the reference database

representing 14 *Actinomyces* species. A cross-validation of this reference database yielded correct identification for all species, which were represented by more than two strains in the database. This database was challenged by 574 unknown *Actinomyces* isolates. All of them could be identified belonging to the 12 most frequent species. For *Actinomyces naeslundii* and *Actinomyces johnsonii* as well as *Actinomyces meyeri* and *Actinomyces odontolyticus,* similar to the 16S rRNA sequences, the mass spectra were not discriminatory. Only tentative identification could be achieved for species which were not sufficiently represented in the collection, such as *Actinomyces graevenitzii*, *Actinomyces europaeus* or *Actinomyces radicidentis* (Stingu *et al.*, 2015).

5.5 Application of the MALDI-TOF MS Method for Routine Identification of Anaerobes in the Clinical Practice

Early evaluations of MS for identification of anaerobic bacteria in routine clinical microbiological laboratories started in 2010 using recent or earlier collected clinical isolates, but only a narrow spectrum of genera and species was tested in these first studies (Culebras *et al.*, 2011; Knoester *et al.*, 2012). Further studies summarizing the experience of the performance of MALDI Biotyper or VITEK MS in routine clinical microbiological practice differed in the number of anaerobic isolates tested (<100 to >1000), as well as in the number and composition of the genera and the species involved (between 11 and 39, and 26 and 102, respectively), mainly depending on the clinical background of the laboratories (Nagy 2014) (Table 5.3). In most cases, results obtained by the MS method were compared with those obtained by phenotypic identification (La Scola *et al.*, 2011; Federko *et al.*, 2012; Nagy *et al.*, 2012; Barreau *et al.*, 2013; Coltella *et al.*, 2013; Kierzkowska *et al.*, 2013; Barba *et al.*, 2014; Garner *et al.*, 2014; Li *et al.*, 2014; Lee *et al.*, 2015). If discrepant results were found, 16S rRNA sequencing was regularly applied to confirm the MALDI-TOF MS identification. In a few studies, 16S rRNA sequencing was used as a 'gold standard' for the identification of anaerobes if MALDI-TOF MS failed to give a species-level identification (Justesen *et al.*, 2011; Barreau *et al.*, 2013).

During the routine use of MS, several rare and/or recently described anaerobic species were identified by the MALDI Biotyper, such as *Anaerotruncus colihominis, Anaerococcus murdochii, Anaerococcus tetradius, Dialister micraerophilus, Porphyromonas gulae, Bacteroides heparinolyticus, Bacteroides salyersiae, Bacteroides tecticus, Prevotella nanceiensis, Prevotella baroniae* or *Turicibacter sanguinis* (Barreau *et al.*, 2013), showing the rapid development of the databases and the applicability of this method for routine identification of clinically important anaerobic bacteria. The performance of the MS-based identification of anaerobic bacteria in daily routine work is highly influenced by the composition of the database at the time of the study and the identification criteria (i.e. cut-off values for species- or genus-level identification) provided by the manufacturers or selected by the laboratories. In these studies, the percentage of the species- and genus-level identification of strict anaerobic bacteria varied between 70.8% and 93.8% and 88% and 98%, respectively (Table 5.3). A highly diverse set of anaerobic clinical isolates belonging to 39 genera and 102 species including numerous less frequent anaerobes was tested by Schmitt *et al.* (2013), which may explain the relatively low level of correct species identification (70.8%) using the cut-off

Table 5.3 Identification of anaerobic bacteria by MALDI-TOF MS in clinical microbiology laboratories (evaluation according to the manufacturer's instructions).

Time of collection of strains	MALDI-TOF MS systems used (version of the database)	Number of isolates tested (genus/species)	Correct identification at species level	Correct identification only at genus level	Not reliable identification	Reference
2004-06	Bruker Biotyper (2.0)	193 (2/13)	181 (93.8%)	11 (5.7%)	1 (0.5%)	Culebras *et al.*, 2011
2010	Bruker Biotyper (NF)	296 (6/NF)	164 (55%)	85 (29%)	47 (15.8%)	Knoester *et al.*, 2012
<2011	Bruker Biotyper (2.0.4.)	152 (24/75)	125 (82%)	12 (17%)	15 (9.8%)	Fedorko *et al.*, 2012
2010–11	Bruker Biotyper (3.0)	283 (15/58)	218 (77%)	31 (11%)	34 (12%)	Nagy *et al.*, 2012
2011	Bruker Biotyper (3.0)	238 (13/34)	185 (77.7%)	33 (14%)	20 (8.4%)	Fournier *et al.*, 2012
2010–11	Bruker Biotyper (3.0)	484 (18/51)	387 (80%)	87 (18%)	8 (2%)	Coltella *et al.*, 2013
2011–12	Bruker Biotyper (3.0)	253 (39/102)	179 (70.8%)	53 (20.9)	20 (7.9%)	Schmitt *et al.*, 2013
2012	Bruker Biotyper (3.0)	1325 (32/95)	1030 (77.7%)	281 (21%)	14 (1%)	Barreau *et al.*, 2013
2012	VITEK MS IVD (2.0)	651 (11/26)	594 (91.2%)	8 (1.2%)	49 (7.5%)	Garner *et al.*, 2013
2014	VITEK MS IVD (2.0)	50 (10/14)	46 (92%)	1 (2%)	4 (8%)	Li *et al.*, 2014
2011	VITEK MS IVD (1.1)	249 (12/27)	209 (83.9%)	18 (7.2%)	22 (8.8%)	Lee *et al.*, 2015

NF: not found.

≥2.000 log score. In some publications, lower log scores (≥1.700 to ≥1.900) were accepted for species-level identification (La Scola *et al.*, 2011; Barreau *et al.*, 2013; Schmitt *et al.*, 2013), and significantly higher percentages of correct identification could be achieved for anaerobic species. Two database versions (DB1 including 4111 entries and DB2 including 4613 entries) of the Bruker system were evaluated in the Anaerobes Reference Unit (Public Health Wales, Cardiff, UK) on 1195 clinical isolates (representing 200 different anaerobic species and 60 anaerobic genera) identified also by 16S rRNA gene sequencing (Copsey *et al.*, 2013). The more developed DB2 identified 71% of the isolates compared to the 63% of correct identifications using the DB1. In this study, many rare isolates of gram-negative taxa (e.g. *Anaerospirillum, Desulfovibrio, Selenomonas, Sneathia, Sutterella*) and gram-positive genera (e.g. *Abiotrophia, Anaerostipes, Catabacter, Collinsella, Caprobacillus, Flavonifractor, Moryella, Parascardovia, Propioniferax, Solobacterium, Tissierella* and *Turicibacter*) were represented with one only isolate in both databases or appeared to be excluded. The best identification was achieved for *Bacteroides*, *Clostridium*, *Propionibacterium acnes* and *Finegoldia magna* by the updated database, which correctly identified 91%, 88%, 71% and 94% of the isolates of these genera and species, respectively.

Despite the fact that there are fewer reports on the VITEK MS system and its IVD database compared to the MALDI Biotyper, there are no significant differences in the performance of the two systems for the common clinical isolates of anaerobic species (Justesen *et al.*, 2011; Veloo *et al.*, 2011b; Martiny *et al.*, 2012; Jamal *et al.*, 2013; Garner *et al.*, 2014; Lee *et al.*, 2015).

The biodiversity of some anaerobic species is well known. It is well recognized that the expansion of the reference spectrum libraries and the inclusion of well-characterized (16S rRNA/protein gene sequenced) clinical isolates into the databases is important to improve the performance of the MS systems for correct identification of anaerobic bacteria. In those cases where certain species, such as *Anaerococcus vaginalis*, *Campylobacter (Bacteroides) ureolyticus*, *Finegoldia magna* or *Clostridium hathewayi*, were represented in the Biotyper database only by one or two entries, they obviously did not cover their natural variability (Veloo *et al.*, 2011a; Nagy *et al.*, 2012). Out of the frequently isolated genera, clinical isolates of *Fusobacterium* and *Actinomyces* are still difficult to differentiate at the species level by MALDI-TOF MS, which may be explained by the wide heterogeneity of some species (such as *Fusobacterium nucleatum*) due to potential horizontal gene transfer (Claypool *et al.*, 2010). Sample preparation may also be important to improve species-level identification of some anaerobic species (see Section 5.3 above).

However, despite the rapid expansion of the databases, there are still examples where both MALDI-TOF MS systems failed to yield species-level identification, such as in the cases of *Bacteroides vulgatus* and the closely related, newly described *Bacteroides dorei* or *Bacteroides ovatus* and *Bacteroides xylanisolvens*. The Shimadzu/SARAMIS system gave a double result (i.e. *B. vulgatus/B. dorei* and *B. ovatus/B. xylanisolvens*), whereas the MALDI Biotyper misidentified these species (Justesen *et al.*, 2011). Mass spectra of these species pairs do not distinguish them, indicating the possible limitation of this technology or rather the possibility that the two species are identical. A similar observation was reported by Chean *et al.* (2014), who identified *Clostridium barati* variably as *Clostridium barati* or *Clostridium paraperfringens* using the MALDI Biotyper. In fact, *Clostridium. barati* and *Clostridium paraperfringens* are considered to be a single

species based on morphology, biochemical tests and DNA-DNA homology; however, both were registered separately in the MALDI Biotyper database at the time of the investigation.

5.6 The European Network for the Rapid Identification of Anaerobes (ENRIA) Project

Some anaerobic bacterial species are represented in the MALDI Biotyper database by only the type strains. Because of the intrinsic diversity of many species discussed above and the lack of a good representative type strain of some species, reliable identification of an unknown strain may sometimes be challenging. This can be partly resolved by the inclusion of additional reference spectra to the MALDI Biotyper database, obtained from genotypically well-characterized clinical isolates of a given species. Seng *et al.* (2009) noted that 'the statistical significance of the correlation between the precision in MALDI-TOF MS identification and the number of reference spectra increased from ≥5 reference spectra to ≥10 reference spectra in the database'. This suggests that it is essential that the number of reference spectra present in the MALDI Biotyper database should be ≥5 in order to correctly identify an unknown strain.

Ribosomal proteins (Arnold and Reilly, 1999) utilized for identification using MALDI-TOF MS produce mass ions in the range 2–20 kDa; thus, the topology of the dendrogram created from the spectra may resemble that of a phylogenetic tree derived from 16S rRNA gene sequences (Stackebrandt and Ebers, 2006). However, a phylogenetic tree is based on genotypic information, whereas a MALDI-TOF MS dendrogram is based on proteins that may have undergone post-translational modifications. This supports the need to have sufficient reference spectra present in the database to allow for eventual intraspecies variation. It is anticipated that the intraspecies variation will fluctuate among different taxa. Analysis of composite spectra and validation of the optimized database should reveal which species require more than five reference spectra to be present in the database to identify wild type strains.

For the optimization of the MALDI Biotyper database for the identification of anaerobic bacteria, a large collection of strains representing all clinically relevant species is needed. To achieve this goal, ENRIA was established through a collaboration between the ESCMID study groups for anaerobic infections (ESGAI) and epidemiological markers (ESGEM). The ENRIA study team consists of a group of core laboratories that are specialized in anaerobic bacteriology and represents different European countries in collaboration with the manufacturer of the Microflex MALDI-TOF MS system, Bruker Daltoniks, Bremen, Germany. The goal of ENRIA is to optimize the MALDI Biotyper database by collecting strains of anaerobic bacterial species not yet included or underrepresented in the database. To achieve this objective, a minimum of five reference spectra of each species is being collected and will be added to the database. Validation of the identity is achieved through 16S rRNA gene sequencing of each strain using 98.7% as a cut-off value for a reliable species identification, as recommended by Stackebrandt and Ebers (2006). To gain insight in the performance of each collaborating laboratory regarding the identification of anaerobic bacteria, a "ring test" was organized. Characterized strains were sent out blindly to each laboratory, and identification scores

were collected and discussed at closed meetings of the group. The results of this quality control assessment of anaerobic species identification by the MS-based method will be published at the end of the project.

So far more than 600 strains have been collected and characterized. Among these strains are species regularly encountered in clinical specimens and which were underrepresented in the MALDI Biotyper database, and newly described species, whose clinical relevance is not established yet. The addition of reference spectra of hitherto undescribed species will assist in clarifying their clinical relevance. Furthermore, species that are phenotypically difficult to identify can be reliably identified with the use of MS, and their clinical relevance can be assessed.

5.7 Subspecies-Level Typing of Anaerobic Bacteria Based on Differences in Mass Spectra

At present, there are only a few clinical situations where typing of anaerobic bacteria is needed for epidemiological or other reasons. *Clostridium difficile* causes severe diarrhoea in hospital settings or in the community and is one of those species where typing may be important to follow the spread of virulent, antibiotic-resistant subtypes. Among the many different typing methods, PCR ribotyping is accepted worldwide for *Clostridium difficile*. This method is time consuming, expensive and generally performed in specialized reference laboratories. With the availability of MS, possible use of protein-profile-based typing for *Clostridium difficile* was tested. Using a standard collection of isolates belonging to 25 different *Clostridium difficile* PCR ribotypes, a database was constructed and recorded in the SARAMIS software (Reil *et al.*, 2011). The database was validated with 355 *Clostridium difficile* clinical isolates belonging to 29 different PCR ribotypes. In this study, the most frequent ribotypes were type 001 (70%), 027 (4.8%) and 078/126 (4.7%). The Shimadzu MALDI-TOF MS system with the specially developed SARAMIS database could differentiate all three frequent ribotypes and allowed a very rapid, effective discrimination of these strains from the others. As this is the only report using MALDI-TOF typing for this purpose, it is still unclear if MS-based typing will have wide application and give results comparable with PCR ribotyping for *Clostridium difficile*.

Using various molecular typing methods, such as arbitrary primed PCR, ribotyping, multilocus enzyme electrophoresis and sequencing of the *recA* and *glnA* genes, it was clearly demonstrated that *Bacteroides fragilis* can be divided into Division I and Division II (Gutacker *et al.*, 2000). It has been shown that *Bacteroides fragilis* strains, which harbour the *cfiA* gene (responsible for carbapenemase production), belong to Division II, whereas those which do not have the *cfiA* gene form Division I (Nagy *et al.*, 2011b; Wybo *et al.*, 2011). Rapid selection of isolates harbouring the *cfi*A gene may cause the ineffectiveness of carbapenem treatment of such infection. Two studies using MALDI-TOF MS identification for anaerobic bacteria revealed that the *Bacteroides fragilis* strains belonging to Divisions I and II can be differentiated more rapidly by MALDI-TOF MS than by DNA-based methods (Nagy *et al.*, 2011b; Wybo *et al.*, 2011). The two studies followed different approaches. Nagy *et al.* (2011b) measured the MS of well-defined *cfiA*-positive and *cfiA*-negative *Bacteroides fragilis* strains and evaluated them using ClinProTools 2.2 software (Bruker Daltonik). The software searched for group-specific

peaks and peak shifts. It was found that MS peaks (4826, 9375, 9649 m/z) are characteristic of Division II isolates, and the software also searched for these peaks in a randomly selected group of *Bacteroides fragilis* strains identified earlier by MALDI-TOF MS with a high log score (>2.000). Nine of 38 *Bacteroides fragilis* strains were found to belong to Division II, and the presence of the *cfiA* gene was confirmed by specific PCR (Nagy *et al.*, 2011b). This approach was used during a six-month period in a Hungarian clinical microbiology laboratory when routine identification of *Bacteroides* isolates was done by MALDI-TOF MS. Out of 60 *Bacteroides fragilis* strains, five (8.3%) belonged to Division II. This corresponds to or is slightly higher than the prevalence of these strains (5.7%) observed previously in this country (Fenyvesi *et al.*, 2014). Wybo *et al.* applied a different method at the same time. The composite correlation index tool of the MALDI Biotyper and a dendogram calculated of all tested *Bacteroides fragilis* isolates (248) clearly separated those which harboured the *cfiA* gene. The dendogram created by MALDI-TOF spectra analysis also grouped together the carbapenem-resistant *Bacteroides fragilis* strains in another study (Trevino *et al.*, 2012). Using this approach, it was possible to report carbapenemase-positive *Bacteroides fragilis* isolates directly identified from positive blood cultures using the MALDI Biotyper OC software and a dedicated library of *cfi*A-negative and *cfi*A-positive main spectral profile (MSP) (Johannson *et al.*, 2014a). Getting the *cfi*A-positive MSP as the first best match and with a log score difference of >0.3 to the second best match was sufficient to consider the isolate *cfi*A-positive. Detection of the *cfi*A gene by PCR confirmed this approach.

5.8 Impact of MALDI-TOF MS on Subspecies Classification of *Propionibacterium acnes*: Insights into Protein Expression using ESI-MS-MS

Propionibacterium acnes is another anaerobic species where typing was explored to distinguish isolates belonging to the normal skin flora from those involved in superficial infection (acne) or deep infections (e.g. prosthetic joint infections, endocarditis or osteomyelitis). *Propionibacterium acnes* can be differentiated into a number of distinct phylotypes by MLST or other typing methods, which are known as types IA_1, IA_2, IB, IC, II and III (McDowell *et al.*, 2012). It has been shown that there are some associations of these *Propionibacterium acnes* phylotypes with certain pathological conditions. In a recent study (Nagy *et al.*, 2013), MS-based typing was tested for the resolution of these genetic subgroups of *Propionibacterium acnes* after routine identification by MALDI Biotyper. The ClinProTools 2.2 and the FlexAnalysis 3.3 software (Bruker) were used to analyze the mass spectra of reference strains belonging to types IA, IB, IC, II and III. Peak variations were identified visually and also by FlexAnalysis 3.3 software. A differentiating library was created and used to type clinical isolates of *Propionibacterium acnes*. The MALDI-TOF MS typing results were comparable with those obtained blindly by MLST for the clinical isolates; however, the MS-based typing could not differentiate subtypes in the IA phylogroup (Nagy *et al.*, 2013). Consequent usage of this typing method directly after the MALDI-TOF MS identification may help to find further correlation between the main types of *Propionibacterium acnes* and their involvement in specific infection processes.

In a separate series of studies by Dekio *et al.* (2012, 2013, 2015), various forms of mass spectrometry (MALDI and SELDI-TOF MS and electrospray ionization tandem mass spectrometry) were used to study the relative pathogenic potential and distribution of the three main phylotypes of *Propionibacterium acnes*. A total of 75 clinical isolates from human skin, blood product contaminants and eye infections, along with several Culture Collection strains of *Propionibacterium acnes* and other *Propionibacterium* species, were studied. Plate-cultured cells were lysed and the cellular proteins analyzed using a Bruker MALDI-TOF MS and a Ciphergen SELDI-TOF MS instrument as described previously (Lancashire *et al.*, 2005; Shah *et al.*, 2005). The latter extended the mass range of the ions more than twofold. Both MALDI- and SELDI-TOF MS readily differentiated strains into two distinct mass spectral subgroups (Figure 5.1). Whole genome sequencing was then used to undertake comparative in silico DNA/DNA hybridization (DDH) based on BLAST analysis as described by Dekio *et al.* (2015). The results showed good congruence between MS types and DDH values. Thus, between 91% and 100% were found among genomes of type I strains and 96% among type II strains. However, values between type I and type II were between 74% and 78.5%, whereas those between types I/II and type III were between 72% and73%, suggesting that type III strains were more distantly related. These results are in agreement with an acceptable sequence homology to retain them into a single species, but delineate isolates into two subspecies. Extensive electron microscopy (Figure 5.2), biochemical tests and in-depth proteome analysis were used to characterize strains belonging to these phylotypes, and two centres of variation were clearly evident, for which two new subspecies were proposed. Type I and type II, which included the original type strain, were assigned to *Propionibacterium acnes* subspecies *acnes,* whereas type III strains, which were characterized by more elongated cells, were assigned to *Propionibacterium acnes* subspecies *elongatum* (Dekio *et al.*, 2015).

Representative strains that exhibited growth under aerobic, microaerophilic and anaerobic atmospheres were cultured under these growth conditions, and their cellular proteins were subjected to SDS-PAGE analysis. Figure 5.3 shows an area of the gel where there were striking differences in protein expression for one strain (type IB, strain K115) which had the capacity to grow in air, microaerophically and anaerobically. These areas were excised, fractions digested with trypsin and analyzed by electrospray ionization tandem MS using a Thermo Fisher LTQ Orbitrap. Type I strains overexpressed the virulence CAMP factor when cultured anaerobically/microaerophically, whereas the enzyme methyl malonyl CoA mutase was also shown to be upregulated. The former may have a direct impact on hair follicle plugging and consequently inflammatory acne. By contrast, when grown under aerobic conditions, no such overexpression was observed. The major end product of the metabolism of *Propionibacterium acnes* is propionic acid (which therefore gives the genus its name) and is produced during normal growth under anaerobic conditions. The overexpression of enzymes involved in metabolism such as methylmalonyl-CoA mutase must confer an ecological advantage for type I strains. This is an adenosyl-cobalamin-dependent enzyme that catalyzes the structural rearrangement of succinyl-CoA into methylmalonyl-CoA as part of the propionic acid fermentation pathway in *Propionibacterium* species. It is essential for growth and energy conservation of this species in anaerobic environments and must confer a selective advantage over competing isolates.

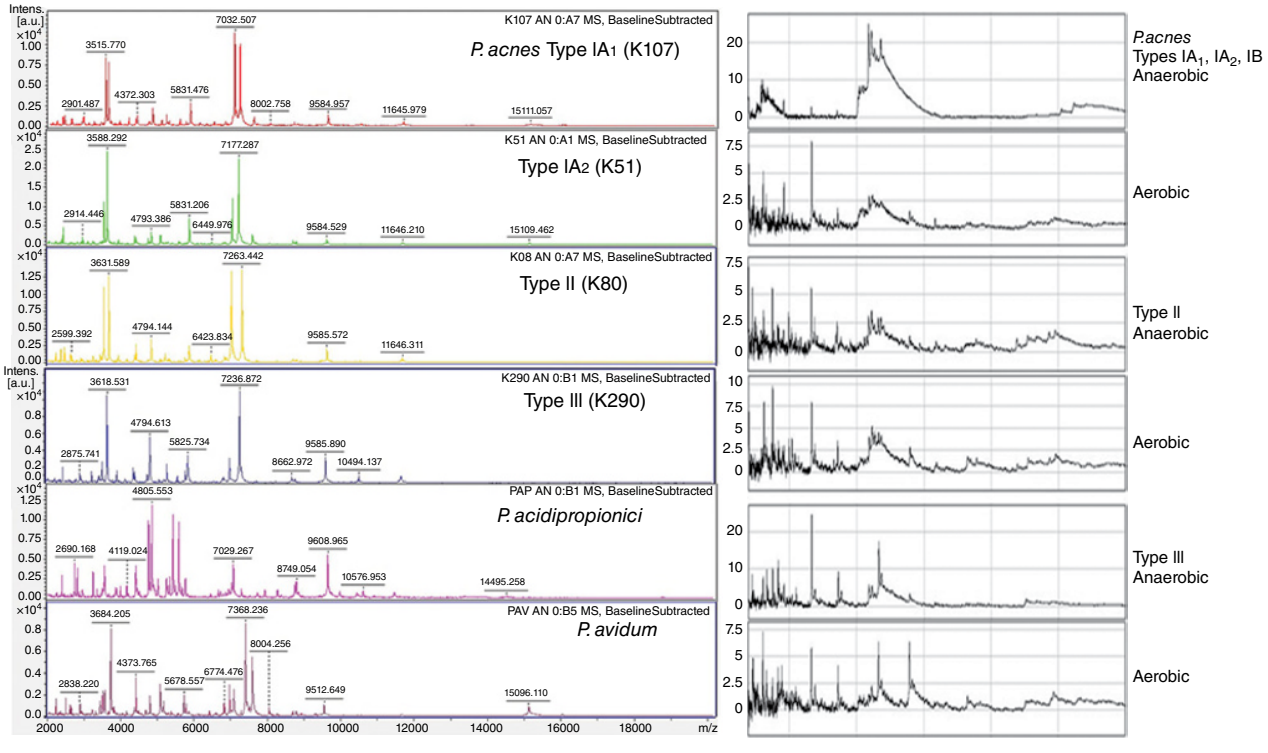


Figure 5.1 (Left): Partial MALDI-TOF MS profiles of anaerobic culture cell extracts of *P. acnes* types and other species. Types I and II share more common mass ions compared to type III. (Right): SELDI profiles of aerobic and anaerobic cultured cell extracts *P. acnes* types covering a mass range between 5 kDa to 30 kDa,

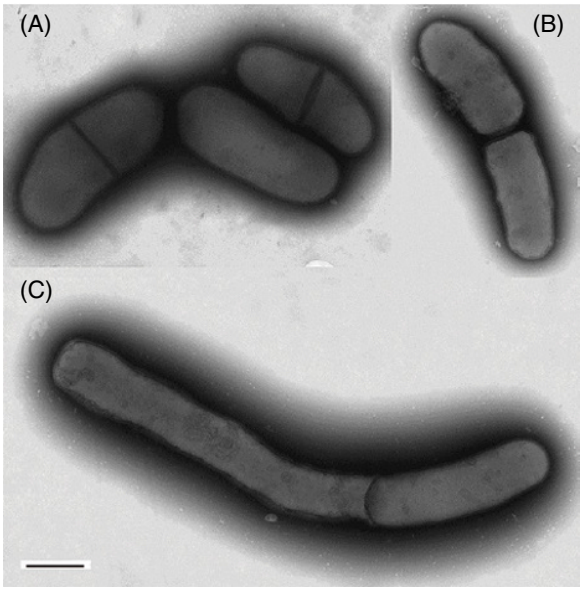


Figure 5.2 Cell morphologies of three types revealed by electron microscopy. A: Type IB (strain K115), B: Type II (strain B23), C: Type III (strain B12) showing an elongated cell structure (0.4–0.7 µm and up to 15 µm long).

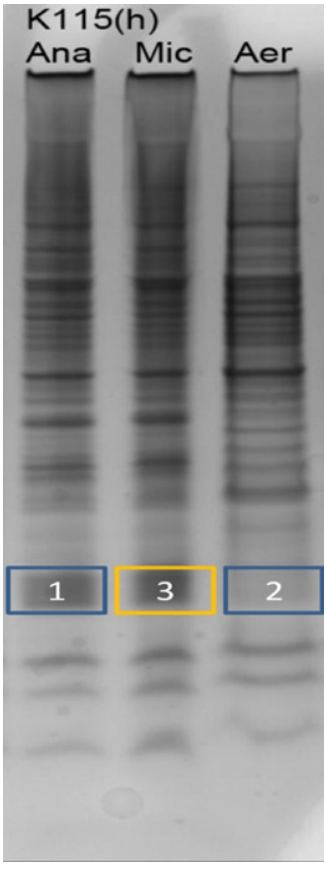


Figure 5.3 1D SDS-PAGE gel analysis of type IB (strain K115). The anaerobic and microaerophilic culture lysates showed dense bands at 12–15 kDa (1 and 3) compared with same section of aerobic lysate band (2). These were excised, trypsin-digested and analyzed using ESI-MS/MS (see text for details).

5.9 Direct Identification of Anaerobic Bacteria from Positive Blood Cultures

Direct identification of bacteria (including anaerobes) and fungi from positive blood cultures by MALDI-TOF MS (La Scola and Raoult, 2009; Moussaoui *et al.*, 2010; Meex *et al.*, 2012; Leli *et al.*, 2013) using different sample preparation methods, such as the Sepsityper kit (Bruker) or in-house methods, is a new approach. So far, only a few cases have been described where anaerobic bacteria could be identified directly from positive anaerobic blood cultures. Using an in-house sample preparation method, Leli *et al.* (2013) reported the species-level identification of 85.5% of gram-positive and 96.9% of gram-negative isolates directly from positive blood cultures, and all seven anaerobic species (three *Bacteroides fragilis*, one *Bacteroides thetaiotaomicron*, one *Clostridium paraputrificum*, one *Parvimonas micra* and one *Actinomyces odontolyticus*) were correctly identified if $\geq$1.7 log score was accepted as a cut-off for species-level identification. In another study, 91% of gram-negative and 89% of gram-positive bacteria in positive blood cultures were correctly identified to the species level, including 5 of 13 anaerobic isolates (all *Bacteroides fragilis*) (Moussaoui *et al.*, 2010). However, one *Bacteroides fragilis*, one *Bacteroides ovatus*, four *Propionibacterium acnes* and two *Clostridium* spp. could not be identified even at the genus level.

Applying artificial spiking of anaerobic blood culture bottles by different *Bacteroides fragilis* isolates, species-level identification and detection of isolates belonging to Division II were achieved with high log scores in all cases after positive signal, using an in-house sample preparation (Johannson *et al.*, 2014b). Further studies are needed to establish the applicability of this method for rapid identification on a wide range of different anaerobic species obtained directly from positive blood cultures or from the culture of any other originally sterile body site. Another possible approach is the subculturing of the positive blood cultures on solid media for 5 h and performing MALDI-TOF MS identification immediately when colonies appear on the surface of the media (Verroken *et al.*, 2014). Using this technique, only 1 *Actinomyces* sp. and a *Clostridium perfringens* were identified from the 10 different anaerobes involved (Verroken *et al.*, 2014). As anaerobic bacteria grow rather slowly even under strict anaerobic conditions, the 5 h incubation time to obtain colonies from the positive blood cultures may not be sufficient.

References

Arnold, R. J. and Reilly, J. P. (1999). Observation of *Escherichia coli* ribosomal proteins and their posttranslational modifications by mass spectrometry. *Anal. Biochem.* **269**, 105–112.

Barba, M. J., Fernández, A., Oviano, M., Fernández, B., Velasco, D. and Bou, G. (2014). Evaluation of MALDI-TOF mass spectrometry for identification of anaerobic bacteria. *Anaerobe* **30**, 126–128.

Barreau, M., Pagnier, I. and La Scola, B. (2013). Improving the identification of anaerobes in the clinical microbiology laboratories through MALDI-TOF mass spectrometry. *Anaerobe* **22**, 123–125.

Bergey, D. H. (1923). *Bergey's Manual of Determinative Bacteriology*. Williams & Wilkins Co, Baltimore.

Chean, R., Kotsanas, D., Francis, M. J., Palombo, E. A., Jadhav, S. R., Awad, M. M., Lyras, D., Korman, T. M. and Jenkin, G. A. (2014). Comparing the identification of *Clostridium* spp. by two matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry platforms to 16S rRNA PCR sequencing as a reference standard: A detailed analysis of age of culture and sample preparation. *Anaerobe* **30**, 85–89.

Claypool, B. M., Yoder, S. C., Citron, D. M., Finegold, S. M., Goldstein, E. J. and Haake, S. K. (2010). Mobilization and prevalence of a Fusobacterial plasmid. *Plasmid* **63**, 11–19.

Coltella, L., Mancinelli, L., Onori, M., Lucignano, B., Menichella, D., Sorge, R., Raponi, M., Mancini, R. and Russo, C. (2013). Advancement in the routine identification of anaerobic bacteria by MALDI-TOF mass spectrometry. *Eur. J. Clin. Microbiol. Infect. Dis.* **32**, 1183–1192.

Copsey, S., Scotford, S., Jones, S., Wootton, M., Hall, V. and Howe, R. A. (2013). Evaluation of MALDI-TOF MS for the identification of anaerobes. *23rd ECCMID Berlin*, Germany P-1715.

Culebras, E., Rodriguez-Avial, I., Betriu, C., Gomez, M. and Picazo, J. J. (2011). Rapid identification of clinical isolates of *Bacteroides* species by matrix assisted laser-desorption/ionization time-of-flight mass spectrometry. *Anaerobe* **18**, 163–165.

De Bruyne, K., Slabbinck, B., Waegeman, W., Vauterin, P., De Beats, B. and VanDamme, P. (2011). Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Syst. Appl. Microbiol.* **34**, 20–29.

Dekio, I., Culak, R., Fang, M., Ball, G., Gharbia, S. E. and Shah, H. N. (2013). Correlation between phylogroups and intracellular proteomes of *Propionibacterium acnes* and differences in the protein expression profiles between anaerobically and aerobically grown cells. *Biomed. Res. Int., Art*. 151797.

Dekio, I., Culak, R., Misra, R., Gaulton, T., Fang, M., Sakamoto, M., Ohkuma, M., Oshima, K., Hattori, M., Klenk, H.-P., Rajendram, D., Gharbia, S. E. and Shah, H. N. (2015). Dissecting the taxonomic heterogeneity within *Propionibacterium acnes*; proposal for *Propionibacterium acnes* subsp. *acnes* subsp. nov. and *Propionibacterium acnes* subsp. *elongatum* subsp. nov. *Int. J. Syst. Evol. Bacteriol.* **60**, 4776–4787.

Dekio, I., Rajendram, D., Morita, E., Gharbia, S. E. and Shah, H. N. (2012). Genetic diversity of *Propionibacterium acnes* strains isolated from human skin in Japan and comparison with their distribution in Europe. *J. Med. Microbiol.* **61**, 622–630.

Federko, D. P., Drake, S. L., Stock, F. and Murray, P. R. (2012). Identification of clinical isolates of anaerobic bacteria using matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Eur. J. Clin. Microbiol. Infect. Dis.* **31**, 2257–2262.

Fenyvesi, V. S., Urbán, E., Bartha, N., Ábrók, M., Kostrzewa, M., Nagy, E., Minárovics, J. and Sóki, J. (2014). Use of MALDI-TOF/MS for routine detection of *cfi*A-positive *Bacteroides fragilis* strains. *Intern. J. Antimicrob. Agents* **44**, 469–476.

Ford, B. A. and Burnham, C. A. D. (2013). Optimization of routine identification of clinically relevant gram-negative bacteria by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry and the Bruker. *Biotyper. J. Clin. Microbiol.* **51**, 1412–1420.

Garner, O., Mochon, A., Branda, J., Burnham, C. A., Bythrow, M., Ferraro, M., Ginocchio, C., Jennemann, R., Manji, R., Procop, G. W., Richter, S., Rychert, J., Sercia, L., Westblade, L. and Lewinski, M. (2014). Multi-centre evaluation of mass spectrometric identification of anaerobic bacteria using VITEK MS system. *Clin. Microbiol. Infect.* **20**, 335–339.

Grosse-Herrenthey, A., Maier, T., Gessler, F., Schaumann, R., Böhnel, H., Kostrzewa, M. and Krüger, M. (2008). Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization time-of-flight mass spectrometry. *Anaerobe* **14**, 242–249.

Gutacker, M., Valsangiacomo, C. and Piffaretti, J. C. (2000). Identification of two genetic groups in *Bacteroides fragilis* by multilocus enzyme electrophoresis: Distribution of antibiotic resistance (*cfi*A, *cep*A) and enterotoxin (*bft*) encoding genes. *Microbiology* **146**, 1241–1254.

Holdeman, L. V., Cato, E. P. and Moore, W. E. C. (Eds.). (1977). *Anaerobe Laboratory Manual*, 4th edition. Virginia Polytechnic Institute and State University, Blacksburg.

Hsu, Y. S. and Burnham, C. D. (2014). MALDI-TOF MS identification of anaerobic bacteria: assessment of pre-analytical variables and specimen preparation techniques. *Diagn. Microbiol. Infect. Dis.* **79**, 144–148.

Jamal, W. Y., Shahin, M. and Rotimi, V. O. (2013). Comparison of two matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry methods and API20AN for identification of clinically relevant anaerobic bacteria. *J. Med. Microbiol.* **62**, 540–544.

Johannson, A., Nagy, E., Sóki, J. on behalf of ESGAI. (2014a). Detection of carbapenemase activities of *Bacteroides fragilis* strains with matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS). *Anaerobe* **26**, 49–52.

Johannson, A., Nagy, E. and Sóki, J. (2014b). Instant screening and verification of carbapenemase activity in *Bacteroides fragilis* in positive blood culture, using matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Med. Microbiol.* **63**, 1105–1110.

Jousimies-Somer, H. R., Summanen, P., Citron, D. M., Baron, E. J., Wexler, H. M. and Finegold, S. M. (Eds.). (2002). *Wadsworth-KTL Anaerobic Bacteriology Manual*, 6th edition. Star Publishing Company, California, USA.

Justesen, U. S., Holm, A., Knudsen, E., Andersen, L. B., Jensen, T. G., Kemp, M., Skov, M. N., Gahrn-Hansen, B. and Møller, J. K. (2011). Species identification of clinical isolates of anaerobic bacteria: A comparison of two matrix-assisted laser desorption ionization – time of flight mass spectrometry systems. *J. Clin. Microbiol.* **49**, 4314–4318.

Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M. and Shah, H. N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect. Genetics Evol.* **4**, 221–242.

Kierzkowska, M., Majewska, A., Kuthan, R. T., Sawicka-Grzelak, A. and Mlynarczyk, G. (2013). A comparison of Api20A vs MALDI-TOF MS for routine identification of clinically significant anaerobic bacterial strains to the species level. *J. Microbiol. Methods* **92**, 209–212.

Könönen, E., Conrads, G. and Nagy, E. (2015). *Bacteroides, Prophyromonas, Prevotella, Fusobacterium* and other anaerobic Gram-negative rods. In Jorgensen, J. H., Pfaller, M. A., Carroll, K. C., Funke, G., Landry, M. L., Richter, S. S. and Warnock, D. W. *Manual of Clinical Microbiology*, 11th edition. ASM Press. pp. 967–993.

Knoester, M., van Veen, S. Q., Vlaas, E. C. J. and Kuijper, E. J. (2012). Routine identification of clinical isolates of anaerobic bacteria: Matrix-assisted laser desorption ionization-time of flight mass spectrometry performs better than conventional identification methods. *J. Clin. Microbiol.* **50**, 1504.

Lancashire, L., Schmid, O., Shah, H. N. and Ball, G. (2005). Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. *Bioinformatics* **21**, 2191–2199.

Lawson, P. A., Gharbia, S. E., Shah, H. N. and Clark, D. R. (1989). Recognition of *Fusobacterium nucleatum* subgroups Fn-1, Fn-2 and Fn-3 by 16S Ribosomal RNA gene restriction patterns. *FEMS Microbiol. Lett.* **62**, 41–45.

Lawson, P. A., Gharbia, S. E., Shah, H. N., Clark, D. R. and Collins, M. D. (1991). Intrageneric relationships of members of the genus *Fusobacterium* as determined by reverse transcriptase sequencing of the small-subunit rRNA. *Int. J. Syst. Bacteriol.* **41**, 347–354.

La Scola, B. and Raoult, D. (2009). Direct identification of bacteria in positive blood culture bottles by matrix assisted laser desorption ionisation time-of-flight mass spectrometry. *PLoS ONE* **4**, E8041.

La Scola, B., Fournier, P.-E. and Raoult, D. (2011). Burden of emerging anaerobes in the MALDI-TOF and 16S rRNA gene sequencing era. *Anaerobe* **17**, 106–112.

Lee, W., Kim, M., Yong, D., Jeong, S. H., Lee, K. and Chog, Y. (2015). Evaluation of VITEK mass spectrometry (MS), a matrix-assisted laser desorption ionization time-of-flight MS system for identification of anaerobic bacteria. *Ann. Lab. Med.* **35**, 69–75.

Leli, C., Cenci, E., Cardaccia, A., Moretti, A., D'Alò, F., Pagliochini, R., Barcaccia, M., Farinelli, S., Vento, S., Bistoni, F. and Mencacci, A. (2013). Rapid identification of bacterial and fungal pathogens from positive blood cultures by MALDI-TOF MS. *Int. J. Med. Microbiol.* **303**, 205–209.

Li, Y., Gu, B., Liu, G., Xia, W., Fan, K., Mai, Y., Huang, P. and Pan, S. (2014). MALDI-TOF MS versus VITEK 2ANC card for identification of anaerobic bacteria. *J. Thorac. Dis.* **6**, 534–538.

Martiny, D., Busson, L., Wybo, I., Haj, R. A. E., Dediste, A. and Vandenberg, O. (2012). Comparison of the Microflex LT and Vitek MS systems for the routine identification of bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. *J. Clin. Microbiol.* **50**, 1313–1325.

Meex, C., Neuville, F., Descy, J., Huynen, P., Hayette. M-P., De Mal, P. and Melin, P. (2012). Direct identification of bacteria from BacT/ALERT anaerobic positive blood cultures by MALDI-TOF MS: MALDI Sepsityper kit versus an in-house saponin method for bacterial extraction. *J. Med. Microbiol.* **61**, 1511–1516.

Magee, J. T., Hindmarch, J. M., Bennett, K. W., Duerden, B. I. and RAries, R. E. (1989). A pyrolysis mass spectrometry study of fusobacteria. *J. Med. Microbiol.* **28**, 227–236.

McDowell, A., Barnard, E., Nagy, I., Gao, A., Tomida, S., Li, H., Eady, A., Cove, J., Nord, C. E. and Patrick, S. (2012). An expanded multilocus sequence typing scheme for *Propionibacterium acnes*: Investigation of 'pathogenic', 'commensal' and antibiotic resistant strains. *PLoS ONE* **7**, e41480.

McElvania Tekippe, E., Shuey, S., Winkler, D. W., Butler, M. A. and Burnham, C. A. D. (2013). Optimizing identification of clinically relevant gram-positive organisms by use of the Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system. *J. Clin. Microbiol.* **51**, 1421–1427.

Moussaoui, W., Jaulhac, B., Hoffmann, A. M., Ludes, B., Kostrzewa, M., Riegel, P. and Prévost, G. (2010). Matrix assisted laser desorption ionization time-of-flight mass spectrometry identifies 90% of bacteria from blood culture vials. *Clin. Microbiol. Infect.* **16**, 1631–1638.

Murdoch, D. A., Gharbia, S.E. and Shah, H. N. (2000). Proposal to restrict the genus *Peptostreptococcus* (Kluyver & van Niel 1936) to *Peptostreptococcus anaerobius*. *Anaerobe* **6**, 257–260.

Murdoch, D. A. and Shah, H. N. (1999). Reclassification of *Peptostreptococcus magnus* (Prévot 1933) Holdeman and Moore 1972 as *Finegoldia magna* comb. nov. and *Peptostreptococcus micros* (Prévot 1933) Smith 1957 as *Micromonas micros* comb. nov. *Anaerobe* **5**, 555–559.

Nagy, E. (2014). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry: a new possibility for the identification and typing of anaerobic bacteria. *Future Microbiol.* **9**, 217–133.

Nagy, E., Becker, S., Kostrzewa, M., Barta, N. and Urbán, E. (2012). The value of MALDI-TOF MS for the identification of clinically relevant anaerobic bacteria in routine laboratories. *J. Med. Microbiol.* **61**, 1393–1400.

Nagy, E., Becker, S., Sóki, J., Urbán, E. and Kostrzewa, M. (2011b). Differentiation of division I (*cfiA*-negative) and division II (*cfiA*-positive) *Bacteroides fragilis* strains by matrix assisted laser desorption ionization- time of flight mass spectrometry. *J. Med. Microbiol.* **60**, 1584–1590.

Nagy, E., Maier, T., Urbán, E., Terhes, G. and Kostrzewa, M. (2009). Species identification of clinical isolates of *Bacteroides* by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. *Clin. Microbiol. Infect.* **15**, 796–802.

Nagy, E., Urbán, E., Becker, S., Kostrzewa, M., Vörös, A., Hunyadkürti, J. and Nagy, I. (2013). MALDI-TOF MS fingerprinting facilitates rapid discrimination of phylotypes I, II and III of *Propionibacterium acnes*. *Anaerobe* **20**, 20–26.

Nagy, E., Urbán, E. and Nord, C.-E. (2011a). Antimicrobial susceptibility of *Bacteroides fragilis* group isolates in Europe: 20 years of experience. *Clin. Microbiol. Infect.* **17**, 371–379.

Prévot, A. R. (1938). *Annales de Institut Pasteur (Paris)* **60**, 285–307.

Rajendram, D. (2003). PhD Thesis; 'Dissecting the Diversity of the Genus *Peptostreptococcus* using Genomics and Proteomic Analyses' (University of East London). Supervisors: H. N. Shah and S. E. Gharbia.

Reil, M., Erhard, M., Kuijper, E. J., Kist, M., Zaiss, H., Witte, W., Gruber, H. and Borgmann, S. (2011). Recognition of *C. difficile* PCR-ribotypes 001, 027, and 126/078 using an extended MALDI-TOF MS system. *Eur. J. Clin. Microbiol. Infect. Dis.* **30**, 1431–1436.

Schmitt, B. H., Cunningham, S. A., Dailey, A. L., Gustafson, D. R. and Patel, R. (2013). Identification of anaerobic bacteria by Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry with on-plate formic acid preparation. *J. Clin. Microbiol.* **51**, 782–786.

Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P. E., Rolain, J. M. *et al*. (2009). Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Infect. Dis.* **49**, 543–551.

Shah, H. N. (2005). MALDI-TOF-mass spectrometry: Hypothesis to proof of concept for diagnostic microbiology. *Clin. Lab. Internat.* **29**, 35–38.

Shah, H. N. and Collins, M. D. (1980). Fatty acid and isoprenoid quinone composition in the classification of *Bacteroides melaninogenicus* and related taxa. *J. Appl. Bacteriol.* **48**, 75–84.

Shah, H. N. and Collins, M. D. (1983). Genus *Bacteroides* a chemotaxonomical perspective. *J. Appl. Bacteriol.* **55**, 403–416.

Shah, H. N. and Collins, M. D. (1988). Proposal for reclassification of *Bacteroides asaccharolyticus, Bacteroides gingivalis* and *Bacteroides endodontalis* in a new genus, *Porphyromonas. Int.J. Syst. Bacteriol.* **38**, 128–131.

Shah, H. N. and Collins, M. D. (1989). Proposal to restrict the genus *Bacteroides* (Castellani and Chalmers) to *Bacteroides fragilis* and closely related species. *Int. J. Syst. Bacteriol.* **39**, 85–97.

Shah, H. N. and Collins, D. M. (1990). *Prevotella*, a new genus to include *Bacteroides melaninogenicus* and related species formally classified in the genus *Bacteroides*. *Int. J. Syst. Bacteriol.* **40**, 205–208.

Shah, H. N., Keys, C., Gharbia, S. E., Ralphson. K., Trundle, F., Brookhouse, I. and Claydon, M. (2000). The application of MALDI-TOF mass spectrometry to profile the surface of intact bacterial cells. *Microbial. Ecol. Health. Dis.* **12**, 241–246.

Shah, H. N., Keys. C. J., Schmid, O. and Gharbia, S. E. (2002). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and proteomics: A new era in anaerobic microbiology. *Clin. Infect. Dis.* **35**, 58–64.

Shah, H. N., Encheva, V., Schmid, O., Nasir, P. *et al*. (2005). Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS): A potentially powerful tool for rapid characterisation of microorganisms. In *Encyclopedia of Rapid Microbiological Methods*, Vol. **3**, Ed. M. J. Miller; DHI Publishing, LLC, River Grove, IL, USA. pp. 57–96.

Shah, H. N. and Hardie, J. M. (1979). Taxonomic studies on *Bacteroides melaninogenicus, Bacteroides oralis, Bacteroides ruminicola and related organisms. Res. Clin. Forums* **1**, 51–53.

Shah, H. N., Olsen, I., Bernard, K., Finegold, S. M., Gharbia, S. E. and Gupta, R. S. (2009). Approaches to the study of the systematics of anaerobic, gram-negative, non-sporeforming rods: Current status and perspectives. *Anaerobe* **15**, 179–194.

Stackebrandt, E. and Ebers, J. (2006). Taxonomic parameters revisited: Tarnished gold standards. *Microbiol. Today* **33**, 152–155.

Stingu, C. S., Borgmann, T., Rodloff, A. C., Vielkind, P., Jentsch, H., Schellenberger, W. and Eschrich, K. (2015). Rapid identification of oral *Actinomyces* species cultivated from subgingival biofilm by MALDI-TOF-MS. *J Oral Microbiol.* **7**, 26110 http://dx.doi.org/10-3402/jom.v7.26110.

Stingu, C. S., Rodloff, A. C., Jentsch, H., Schaumann, R. and Eschrich, K. (2008). Rapid identification of oral anaerobic bacteria cultivated from subgingival biofilm by MALDI-TOF MS. *Oral Microbiol. Immun.* **23**, 372–376.

Tavana, A. M., Drucker, D. B. and Boote, V. (1998). Phospholipid molecular species distribution of *Porphyromonas asaccharolytica* ATCC 25260T: Effects of temperature, culture age and pH. *J. Appl. Microbiol.* **85**, 1029–1035.

Trevino, M., Areses, P., Penalver, M. D., Cortizo, S., Pardo, F., del Molino, M. L., García-Riestra, C., Hernández, M., Llovo, J. and Regueiro, B. J. (2012). Susceptibility trends of *Bacteroides fragilis* group and characterization of carbapenemase producing strains by automated REP-PCR and MALDI-TOF. *Anaerobe* **18**, 37–43.

Veillon, A. and Zuber, M. M. (1898). Recherches sur quelques microbes strictement anaerobies et leur role en pathologie. *Arch. Med. Exp.* **10**, 517–545.

Veloo, A. C. M., Elgersma, P. E., Friedrich, A. W., Nagy, E. and van Winkelhoff, A. J. (2014). The influence of incubation time, sample preparation and exposure to oxygen on the quality of the MALDI-TOF MS spectrum of anaerobic bacteria. *Clin. Microbiol. Infect.* **20**, O1091–O1097.

Veloo, A. C. M., Erhard, M., Welker, M., Welling, G. W. and Degener, J. E. (2011a). Identification of gram-positive anaerobic cocci by MALDI-TOF Mass Spectrometry. *Syst. Appl. Microbiol.* **34**, 58–62.

Veloo, A. C. M., Knoester, M., Degener, J. E. and Kuijper, E. J. (2011b). Comparison of two matrix-assisted laser desorption ionization–time of flight mass spectrometry methods for the identification of clinically relevant anaerobic bacteria. *Clin. Microbiol. Infect.* **17**, 1501–1506.

Verroken, A., Defourny, L., Lechgar, L., Magnette, A., Delmée, M. and Glupczynski, Y. (2014). Reducing time to identification of positive blood cultures with MALDI-TOF MS analysis after a 5-h subculture. *Eur. J. Clin. Microbiol. Infect. Dis.* **34**, 405–413.

Watanabe, Y., Nagai, F., Morotomi, M., Sakon, H. and Tanaka, R. (2010). *Bacteroides clarus* sp. nov., *Bacteroides fluxus* sp. nov. and *Bacteroides oleiciplenus* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **60**, 1864–1869.

Williams, T. L., Andrzejewski, D., Lay, J. O. Jr. and Musser, S. M. (2003). Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. *J. Am. Soc. Mass Spectrom.* **14**, 342–351.

Wybo, I., De Bel, A., Soetens, O., Echahidi, F., Vandoorslaer, K., Van Cauwenbergh, M. and Piérard, D. (2011). Differentiation of *cfiA*-negative and *cfiA*-positive *Bacteroides fragilis* isolates by matrix assisted laser desorption ionization – time of flight mass spectrometry. *J. Clin. Microbiol.* **49**, 1961–1964.

Wybo, I., Soetens, O., De Bel, A., Echahidi, F., Vancutsem, E., Vandoorslaer, K. and Piérard, D. (2012). Species identification of clinical *Prevotella* isolates by matrix assisted laser desorption ionization time-of-flight mass spectrometry. *J. Clin. Microbiol.* **50**: 1415–1418.

6

Differentiation of Closely Related Organisms using MALDI-TOF MS

Mark A. Fisher

Assistant Professor of Pathology, University of Utah School of Medicine, Medical Director, Bacteriology and Antimicrobials, ARUP Infectious Diseases Laboratory, Salt Lake City, UT, USA

6.1 Introduction

The introduction of matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) into the clinical microbiology laboratory has resulted in a major change in the way microorganisms are identified. Since the time of Robert Koch, microbiologists have relied on detecting the metabolic capabilities of microbes for classification and identification. Methods to detect fermentation, oxidation, and other forms of metabolism dominated the diagnostic laboratory for over a century. These methods survived with only minor modifications because they were practical, relatively inexpensive, and could be partially automated, simplifying the process of routine identification in ever-shrinking clinical laboratories. Molecular biology techniques such as nucleic acid probes, PCR, and DNA sequencing have found niche roles in the microbiology laboratory, but have generally not replaced traditional culture and biochemical identification. Cost is a major factor, but breadth of coverage is also a major limitation. Of the routinely available DNA- or RNA-based methods, only DNA sequencing of universal targets, such as ribosomal RNA genes, can match the extremely broad capabilities of biochemical identification. Unfortunately, this approach is too costly for routine identification of typical bacteria and suffers from the nature of its breadth in that it is difficult to apply efficiently to samples containing mixed populations of microorganisms. MALDI-TOF MS is a molecular biology method that rivals biochemical identification of microbes in not only its breadth of coverage (nearly 2300 species in the Bruker Biotyper 5627 spectrum database), but also in its speed and relative cost. Most biochemical systems require multiple panels, the choice of which is determined by at least basic knowledge of the growth and Gram stain characteristics of the microorganism, to cover the full range of organisms within its capabilities. Because metabolic reactions are the basis of identification in these systems, an incubation period of several hours to overnight is generally required. The most rapid tests, designed to detect preformed enzymes, can reduce that delay to a few hours. In contrast, MALDI-TOF MS

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

can be applied regardless of the prior knowledge of the phenotypic characteristics of most bacteria routinely encountered in the clinical microbiology laboratory. Further, once organisms are isolated on culture media, identification can be achieved in minutes. The application of this technology to the analysis of positive blood cultures and even direct specimens is poised to transform the microbiology laboratory from a service that is perceived to provide important, but often supportive data (e.g., after clinical decisions have been made), to an active participant in the early diagnostic process [1,2].

The capability of MALDI-TOF MS to identify an extremely broad range of microorganisms has been shown in many publications in recent years [3–9]. However, several have pointed out limitations in the commercial systems' abilities to resolve certain closely related organisms, including numerous significant pathogens such as *Streptococcus mitis* and *S. pneumoniae*, *Yersinia pseudotuberculosis* and *Y. pestis*, the *Mycobacterium tuberculosis* complex, and *Escherichia coli* and *Shigella* spp. [3–7,10–12]. It is not surprising, nor unexpected, that a single method with such broad identification capabilities could suffer from lack of resolution at the fine level of detail needed to distinguish closely related organisms. In fact, some error must be tolerated in the identification algorithms in order to allow for natural variation within bacterial species and in data acquisition [13,14]. Neither biochemical identification methods, nor the arguable gold standard for microbial identification, rRNA gene sequencing, are immune to such difficulties. Interestingly, there is a significant degree of overlap between organisms that challenge the resolution of these methods [3–7,10–12,15]. Because MALDI-TOF MS is such a broadly applicable technology, and likely to become an integral part of routine clinical microbiology, it seems important to pursue its limits in addressing these difficult challenges. To that end, in this study, a solution to the apparent inability of MALDI-TOF to distinguish between *E. coli* and *Shigella* spp. was sought.

Numerous investigators using a variety of methods, such as DNA–DNA hybridization, multilocus enzyme electrophoresis, and both housekeeping gene and whole genome comparative analyses, have indicated that *Shigella* spp. and *E. coli* are likely the same species [16–19]. *E. coli* is thought to have diverged from other bacteria roughly 8–22 million years ago, yet *Shigella* spp. and a very similar *E. coli* pathovar, enteroinvasive *E. coli* (EIEC), appear to have diverged from the parent *E. coli* several times within the last 300,000 years [20]. Although these organisms should be considered the same species, differences in pathology necessitate their distinction. *Shigella* spp. are highly infectious, with an infectious dose of roughly 10 bacteria, and there are approximately 80–165 million cases and 600,000 deaths per year worldwide from *Shigella* infection [21]. Biochemical methods are usually capable of distinguishing typical *E. coli* isolates from *Shigella* spp., which is a major reason they have traditionally been considered different species. However, atypical *E. coli* isolates, such as metabolically "inactive" strains, and even the approximately 5% of isolates that do not ferment lactose, are much more challenging for clinical laboratories to correctly distinguish from *Shigella* spp. As current routine MALDI-TOF MS identification systems, and even the gold-standard method of bacterial identification, 16S rRNA gene sequencing, fail to reliably differentiate even the typical isolates of these organisms (Figure 6.1) [5,10,15], a rapid, inexpensive MALDI-TOF MS method that would fit into routine workflow would be very beneficial to the clinical microbiology laboratory.

Mo	Detected Species	Log(Score)
	Escherichia coli ATCC 25922 THL	2.481
	Escherichia coli DSM 1103 QC DSM	2.423
	Escherichia coli MB11464 1 CHB	2.355
	Escherichia coli DSM 682 DSM	2.354
	Escherichia coli ATCC 25922 CHB	2.340
	Escherichia coli DSM 1576 DSM	2.308
	Escherichia coli ESBL EA RSS 1528T CHB	2.270
	Escherichia coli DH5alpha BRL	2.230
	Escherichia coli RV412 A1 2010 06a LBK	2.206
	Escherichia coli Nissl VML	2.205

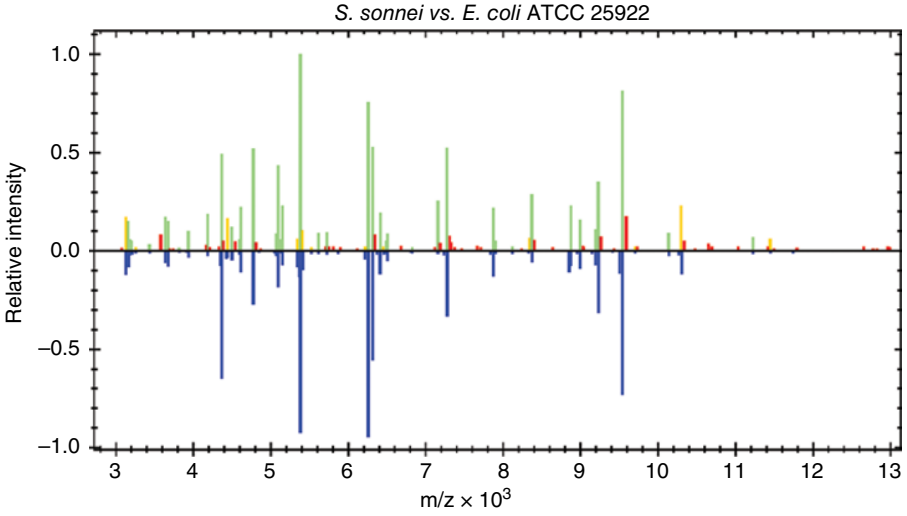


Figure 6.1 Representative MALDI Biotyper analysis of a *S. sonnei* isolate. The database lacks *Shigella* spp. reference spectra, and highly reliable species-level matches occur for *E. coli*.

6.2 Experimental Methods

6.2.1 Strains and Traditional Identification

A total of 138 isolates (72 *E. coli*, 66 *Shigella* spp.) were evaluated, including 127 clinical isolates from 17 US states submitted to ARUP Laboratories (Salt Lake City, UT, USA) between 2006 and 2012 for reference identification. Strains were isolated from a variety of clinical specimens and body sites such as stool, urine, genital, wound, tissue, respiratory, and blood. The remaining 11 isolates were composed of 5 ATCC reference strains: *E. coli* (25922), *S. sonnei* (25931), *S. flexneri* (12022), *S. boydii* (8700), and *S. boydii* (BAA-1247); 2 *S. dysenteriae* and 2 *S. boydii* isolates submitted to the Utah Department of Health (gifts from Chad Campbell); and 2 related *S. dysenteriae* isolates, CVD1254 and CVD1255 lacking the *stxAB* and *guaBA, stxAB,* and *sen* genes, respectively (gifts from Dr. Eileen Barry, University of Maryland) [22]. In all, there were 31 typical and 41 inactive *E. coli*, 35 *S. sonnei*, 23 *S. flexneri*, 4 *S. boydii,* and 4 *S. dysenteriae*. Isolates were

identified by traditional and automated biochemicals and/or serogrouping according to standard clinical microbiology procedures [23]. Isolates with discrepant identifications were further analyzed by PCR for the *ipaH* and *lacY* genes [24,25]. "Inactive" *E. coli* were defined as *E. coli* isolates that lacked two or more of gas production from glucose, motility, or lactose fermentation [23]. Unless otherwise indicated, isolates were grown for 18 to 24 h at 35 °C on Columbia sheep blood agar (SBA) for biochemical and phenotypic testing and on MacConkey agar for MALDI-TOF MS (Hardy Diagnostics, Santa Maria, CA, USA).

Inactive *E. coli* and *Shigella* spp. were serogrouped using antisera specific for groups A-D (BD Diagnostics) according to manufacturer's recommendations. Briefly, bacteria were harvested from SBA plates and used to make a dense suspension in 0.85% saline. Bacterial suspensions were mixed 1:1 with each of the anti-A, -B, -C, and -D reagents (corresponding to *S. dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*, respectively), rocked gently for 1 min, and observed for 3-4+ agglutination, indicating a positive reaction. Isolates that displayed biochemical or phenotypic characteristics consistent with *Shigella* but failed to agglutinate strongly in antisera were reanalyzed after boiling the suspension for 60 min to remove envelope antigens, which may block specific antibody binding. Boiled suspensions were centrifuged (1000 rpm, 15 min), the supernatant discarded, and the pellet resuspended again in saline prior to reanalysis.

Routine phenotypic and biochemical methods were used to identify isolates according to standard clinical microbiology procedures [23]. Automated (Phoenix NID panel, BD Diagnostics, Sparks, MD, USA) and/or traditional biochemical analyses using lysine-iron agar (LIA), motility-indole-ornithine (MIO), triple sugar iron (TSI), and MacConkey agars (Hardy Diagnostics) were interpreted according to manufacturer's recommendations.

6.2.2 PCR Identification

Real-time PCR assays targeting the *ipaH* (invasion plasmid antigen H) and *lacY* (β-galactoside permease) genes were developed to distinguish *E. coli* from *Shigella* species as previously described [24,25]. DNA was extracted from pure cultures (MagaZorb DNA Mini-Prep, Promega, Madison, WI) and quantified by spectrophotometry. Quantitative PCR (qPCR) was performed using appropriate positive and negative controls as described below on the SmartCycler real-time PCR instrument (Cepheid, Sunnyvale, CA) using the dsDNA binding dye LCGreen Plus + (BioFire Diagnostics, Salt Lake City, UT).

The *ipaH* gene qPCR reactions (25 µl) contained 1X Colorless GoTaq Flexi DNA Polymerase (Promega), 3 mM of $MgCl_2$, 0.3 mM dNTP Blend (Promega), 0.5 µM each of forward (5′-TCGATAATGATACCGGCGCTC-3′) and reverse (5′-CTGCGAGCATGGTCTGGAA-3′) primer, 0.5X LCGreen Plus+, and 100 ng genomic DNA. PCR cycling conditions consisted of an initial melt at 95 °C for 2 min, then 30 cycles of 95 °C for 20 s, 55 °C for 30 s, and 72 °C for 20 s, with a terminal extension of 72 °C for 5 min and a final melt curve to confirm the 147 bp *lacY* PCR product. Positive (*S. flexneri*) and negative (*E. coli*) control genomic DNA samples were analyzed on every run. An isolate was called positive for the *lacY* gene when the crossing threshold (Ct) value was within ±3.3 cycles (10-fold concentration) of the positive control and had a characteristic melt peak at 83.5 °C, and negative when the Ct value was less than 9.97 cycles (0.001-fold concentration) of the positive control.

The *lacY* gene qPCR reaction and data interpretation conditions were identical to the *ipaH* PCR except for the use of 20 ng genomic DNA and 58 °C annealing temperature to generate a 102 bp amplicon using the forward 5′-CTGCTTCTTTAAGCAACTGGCGA-3′ and reverse primer 5′-ACCAGACCCAGCACCAGATAAG-3′. Positive (*E. coli*) and negative (*S. sonnei*) control genomic DNA samples were analyzed on every run.

6.2.3 MALDI-TOF MS Identification

Standard extractions were performed on isolates from MacConkey agar as previously described [5,26]. Briefly, 5–10 mg of bacteria were washed and inactivated by resuspension in 300 µl dH_2O, then mixed by inversion with 900 µl absolute ethanol. Bacteria were harvested by centrifugation (16000 × g, 2 min), and any residual ethanol was removed by aspiration following a subsequent centrifugation (16000 × g, 2 min). Cells were resuspended in 70% formic acid (50 µl), vortexed for 1 min, and mixed with an equal volume of pure acetonitrile. The bacterial extract was transferred to a fresh tube following a final centrifugation (16000 × g, 2 min). Triplicate 1 µl aliquots were allowed to air-dry on polished steel targets and were overlaid with 1 µl matrix (saturated alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile–2.5% trifluoroacetic acid) that was allowed to air-dry prior to analysis.

Mass spectra were acquired on a microflex LRF instrument (Bruker Daltonics, Billerica, MA) between 2000 and 20000 *m/z* in linear positive ionization mode. Standard MALDI Biotyper (Bruker Daltonics) automated acquisition settings were used except that each spectrum was a sum of 500 shots collected at intervals of 100. Isolates with low-quality spectra, defined as failure to be identified as *E. coli* with a Biotyper score of ≤1.9, were reanalyzed using manual data acquisition. Isolates whose spectra failed recalibration in ClinProTools 2.2 (Bruker Daltonics) as described below were regrown, reextracted, and spectra re-collected as described above.

Models based on statistical analysis of peaks found in each group of organisms (e.g., classes) were developed using three distinct approaches with the goal of identifying reliable organism classifiers. In each approach, spectra from subsets of the organisms, or model generation cohorts, were used to identify peaks that could distinguish between classes of isolates (e.g., species). Following model generation, known test cohort spectra were identified, or "classified," by the model to evaluate its performance. ClinProTools was used in model development, and in two approaches, for final isolate classification. The classification in ClinProTools involved two steps: the first was used to differentiate between only two classes (*Shigella* spp. and *E. coli*, or "genus-level" distinction); the second step was used to differentiate all five classes of organisms (*S. sonnei, S. flexneri, S. boydii, S. dysenteriae,* and *E. coli*, or species-level distinction). If the classification results from the two-class and five-class models were concordant (e.g., if both models agreed at the genus level), then the species-level identification was accepted. Inconsistent results, or those which disagreed at the genus level, were flagged for additional analysis. In a typical laboratory, this would lead to supplemental testing by biochemical approaches and/or serotyping. MALDI-TOF MS accuracy was determined by agreement with the reference identification, which was derived from serotyping, biochemical methods, and/or PCR, as described above.

To generate the semiautomated models, potentially distinguishing peaks were identified from ClinProTools' Peak Statistic Tables, which were calculated after comparison

of each class of organisms. FlexAnalysis (Bruker Daltonics) was used to confirm, by visual inspection, that peaks were real and reliably present in at least one class. Each spectrum was processed in FlexAnalysis for baseline subtraction, smoothing, and calculation of mass lists using default parameters. The frequencies at which potential biomarker peaks were present within isolates of a given class were calculated using custom MATLAB code (Mathworks, Natick, MA; code available upon request). Biomarker peak frequencies were assembled into "reference peak profiles," which were compared to mass lists from each test isolate. Pearson's correlation coefficients were calculated for each reference peak profile, and the profile with the highest score was deemed the identification of the test isolate.

Classification models generated by the automated approach relied on ClinProTools for both data preparation and model generation, as well as the final spectrum classification step. Spectra were prepared for analysis and model generation in ClinProTools using the following parameters: baseline subtraction using top hat with minimal baseline width of 10%, normalization using total ion current, recalibration (maximal peak shift = 1,000 ppm, 30% match to calibrant peaks, spectra unable to be recalibrated were excluded), average spectrum calculation using a resolution of 800, average peak list calculation using a signal-to-noise threshold of 5, individual spectrum peak calculation, and peak list normalization. Multiple model generation algorithms (genetic algorithm, supervised neural network, quick classifier) were initially evaluated using default settings. The genetic algorithm [27] was chosen for routine analysis and was used with the following settings: up to 15 peaks acceptable, initial number of peak combinations automatically detected, up to 50 generations of evolution, 0.5 crossover rate, 0.2 mutation rate, 3 neighbors, and no varying random seed. As part of the model generation process, ClinProTools calculates two statistics, Cross Validation and Recognition Capability, for each model. Cross Validation is calculated by randomly dividing the input spectra into a model subset and a test subset. A model is generated and used to classify spectra in the test subset. This process is iterated, and the results are used to calculate a normalized Cross Validation score, or the average percentage of correct identifications across these iterations [28]. This statistic describes the model's reliability and may be used to predict its future performance. Recognition Capability describes the model's ability to correctly identify the individual spectra used to build the final model. It is defined as the number of spectra classified correctly by the model itself, divided by the total number of spectra used in model generation, or the percentage of spectra that were correctly classified by the model. The choice of which model generation algorithm to use was based on maximizing both of these scores. Unknown or test spectra were identified using the ClinProTools Classify function, and at least two of the three replicate spectra per isolate were required to agree in order to accept an identification.

A hybrid approach was developed by using the same data preparation and model generation settings as the automated approach, except that selected peaks from the semiautomated approach were included using the ClinProTools Force Peak option. Peaks were selected for the hybrid model if they increased the Cross Validation and Recognition Capability scores relative to the automated model.

To evaluate statistical significance, the Fisher's exact test was used to assess interaction of categorical variables using the statistical computing software R (v.2.15.0; http://www.R-project.org). P values < 0.05 were considered statistically significant.

6.3 Results

Preliminary manual analyses using FlexAnalysis software comparing extracts of *S. flexneri* and typical *E. coli* indicated that slightly different mass profiles could be observed under routine MALDI-TOF MS conditions. These differences were more pronounced when strains were grown on MacConkey agar compared to Columbia sheep blood or Hektoen Enteric agars. Because MacConkey agar is routinely used in the clinical laboratory, it was determined that an assay based on growth from this media would fit well in a normal laboratory workflow, and this approach was used for the remainder of the study.

6.3.1 Semiautomated Models

The semiautomated approach relied on an initial statistical analysis of the presence, absence, and intensity of peaks found in each organism category. Significant manual review was necessary to validate that peaks were of sufficient intensity and frequency of occurrence to be reliable in a classification model. An early semiautomated model based on four organism classes (typical *E. coli,* inactive *E. coli, S. flexneri,* and *S. sonnei*) revealed 13 peaks that were reasonably distinguishing (Table 6.1). Application of Pearson's correlation analysis to each test isolate resulted in scores that could be used for identification. This early model performed well in identifying *S. sonnei* and *S. flexneri* (95% accuracy), but was unable to reproducibly identify *E. coli* correctly (44% accuracy). Figure 6.2 shows that isolates generally showed the highest correlation coefficient with their cognate model; however, inactive *E. coli* showed significant overlap with all four models, and 14% of typical *E. coli* isolates were misidentified as *S. sonnei*. Further development of this model by consolidation of all *E. coli* isolates into a single class, addition of *S. dysenteriae* and *S. sonnei* (four each), and refinement of distinguishing peaks (added 7288 and 7302 *m/z* and removed peak 2874 *m/z* to give 14 final peaks) failed to significantly improve its performance: 94% (31 of 33) of *Shigella* but only 56% (20 of 36) of *E. coli* test cohort isolates were correctly identified by this five-class semiautomated model.

6.3.2 Automated Models

Because the semiautomated models were unable to reliably identify *E. coli*, ClinProTools was used to generate models in an automated fashion using the genetic algorithm [27]. This approach differed from the semiautomated approach in that peaks were evaluated for the ability to distinguish classes based on relative intensity in addition to presence or absence. As a result, a different, but partially overlapping, set of peaks was identified to distinguish among two (genus-level model: *E. coli* vs. *Shigella* spp.) or five classes of organisms (species-level model) (Table 6.2). Although, as with the semiautomated models, the chosen peaks were not absolutely conserved across classes (Figure 6.3), this approach clearly outperformed the semiautomated models overall, correctly identifying 64 of 69 isolates (94%, including 100% of *Shigella* spp. and 89% of *E. coli*) at the genus level, and 63 of 69 (91%, including 85% of *Shigella* spp. and 97% of *E. coli*) at the species level. Several isolates were correctly identified by one model (e.g., genus or species level), but not the other. To handle these discrepancies, a two-step algorithm was implemented that required agreement across both the two-way (genus) and five-way

Table 6.1 Semiautomated model peak frequency profiles across four organism classes.

Marker peaks (m/z)	*S. flexneri*	*S. sonnei*	Typical *E. coli*	Inactive *E. coli*
2400	0.0	0.0	75.0	16.7
2874	0.0	65.9	61.1	29.5
3791	0.0	0.0	83.3	17.9
4162	98.0	83.0	0.0	0.0
4855	0.0	0.0	86.1	91.0
4869	94.1	72.7	0.0	1.3
5096	17.6	100	100	100
5752	5.9	89.8	81.9	62.8
8323	98.0	93.2	5.6	3.8
8457	96.1	5.7	1.4	2.6
9711	0.0	0.0	91.7	88.5
9736	90.2	79.5	13.9	9.0
10458	70.6	3.4	1.4	32.1

Values indicate the percentage of spectra containing the indicated peak.

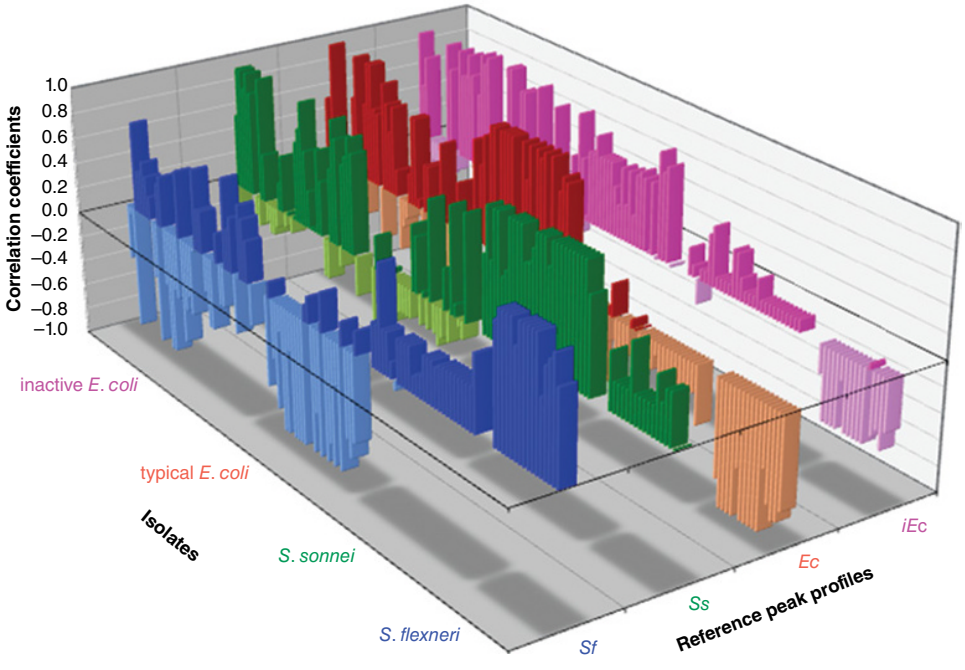


Figure 6.2 Pearson's correlation coefficients for isolates tested by the four-class semiautomated model (Table 6.1). Three-dimensional plot showing the degree of correlation between each isolate and the four-peak frequency profiles (Table 6.1): *S. flexneri* (Sf), *S. sonnei* (Ss), typical *E. coli* (Ec), and inactive *E. coli* (iEc). A score of +1 indicates a direct correlation and −1 indicates a perfect inverse correlation.

Table 6.2 Biomarker peaks (*m/z*) identified in the automated MALDI-TOF MS approaches.

Automated		Hybrid	
Genus-level model	**Species-level model**	**Genus-level model**	**Species-level model**
2848	2701	2400 [a]	2400 [a]
3577	3673	3577 [b]	3578 [b]
3673	5096	3673 [b]	3673 [b]
5120	5136	3792 [a]	5096 [a, b]
5326	8324	4162 [a]	5136 [b]
6507	8444	4856 [a]	6668 [b]
6668	9533	5326 [b]	8324 [b]
6825	10135	6507 [b]	8444 [b]
6857	12222	6668 [b]	8455 [a]
7157	13601	7157 [b]	9533 [b]
8349	14725	8349 [b]	10135 [b]
9223		9223 [b]	13601 [b]
9264		9448 [b]	
9448		9711 [a]	
11706		11731	

[a] Peaks from the semiautomated approach selected for inclusion in ClinProTools models.
[b] Peaks from the automated approach selected for inclusion in ClinProTools models.
Adapted with permission from Khot, P. D. *et al.*, *J. Clin. Microbiol.*, **51**, 3711–3716 (2013), doi:10.1128/JCM.01526-13, Copyright © American Society for Microbiology.

(species) models to secure an identification. Disagreement between models resulted in an inconclusive result that would flag the isolate for further testing, such as serogrouping or PCR. Using this two-step algorithm, the automated model was able to correctly identify 86% of the 69 test isolates, including 85% of *Shigella* spp. and 86% of *E. coli*. Nine isolates showed discrepant results between genus- and species-level models, including two *S. sonnei* and one *S. boydii* called *E. coli* by the species-level model, but only a single isolate was misidentified after applying the algorithm. The lone misidentification was a *S. flexneri* isolate that was reported as *S. boydii* by the species-level model.

6.3.3 Hybrid Models

Although the automated models together outperformed the semiautomated models, manual review of spectra suggested that some peaks identified in the semiautomated approach might improve the performance of the automated models. ClinProTools allows peaks to be manually selected for, or "forced into," a model, and by forcing several peaks from the semiautomated approach along with peaks chosen in the automated approach, new hybrid models were generated. Three semiautomated model peaks were included in the species-level and five in the genus-level model (Table 6.2). Recognition Capability scores were nearly perfect for both sets of models (99.8%–100%). However, Cross Validation scores indicated that the hybrid genus-level model was slightly better (99.8% vs. 99.4%), and the hybrid species-level model was markedly better (97.2% vs.

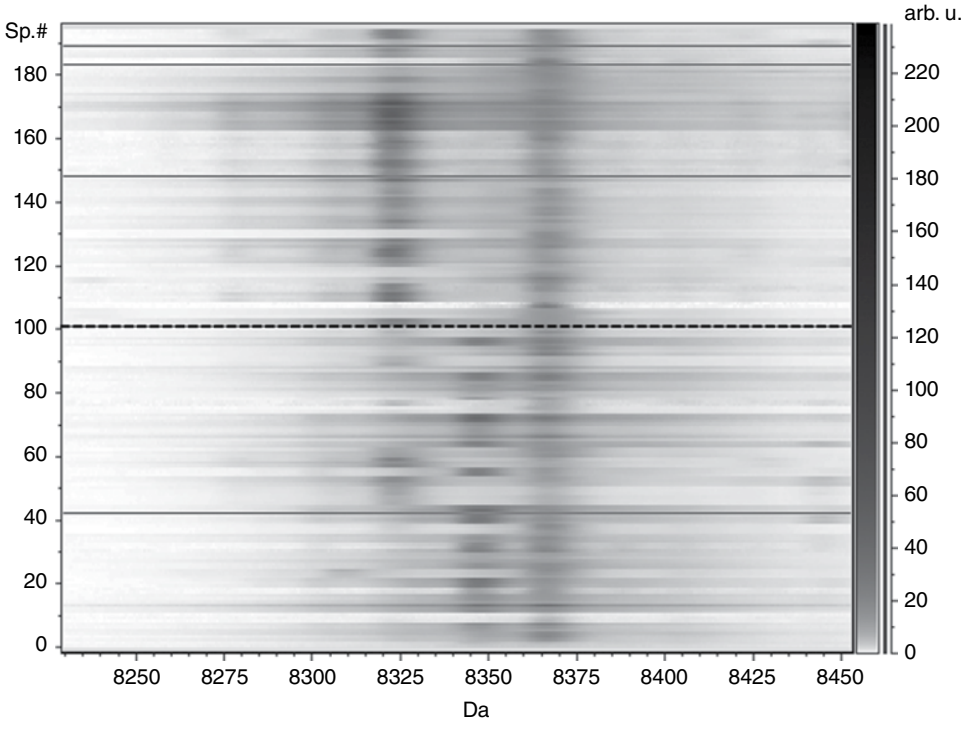


Figure 6.3 An example of peaks chosen by the ClinProTools automated model generation algorithm. Spectra are shown in a gel-like view with *E. coli* spectra below and *Shigella* spp. spectra above the dotted line. Peak 8349 was chosen by the algorithm for the genus-level model, 8324 for the species-level model, and 8366 was considered nondiscriminatory.

90.3%), than the corresponding automated models. The genus-level model accurately identified 66 (96%) of the test isolates, whereas the species-level model correctly identified 63 (91%). Using the two-step algorithm requiring agreement between both models, 90% of the 69 test isolates were correctly identified, including 91% of *Shigella* spp. and 89% of *E. coli* (Table 6.3). Five isolates were reported as inconclusive, and two were misidentified: a *S. flexneri* was reported as *S. boydii,* and a typical lactose-fermenting *E. coli* was reported as *S. sonnei* (Table 6.4). In the latter case, this typical *E. coli* would generally not have resulted in an incorrect final report because, due to its lactose-fermenting phenotype, it would not routinely be tested using this method. As with the automated models, the application of the two-step algorithm reduced the total number of correct identifications relative to the species-level model, albeit by a single isolate, but it also reduced the number of misidentifications from four to two (Table 6.4).

6.3.4 MALDI-TOF MS versus Traditional Identification Methods

The success of the hybrid model indicated that this approach could be used routinely in the clinical laboratory. However, replacing traditional methods may be difficult to justify because of their established position in recommendations and guidelines. To provide evidence to justify the use of this novel approach, the accuracy of identification

Table 6.3 Accuracy of the hybrid MALDI-TOF MS assay, serogrouping, and automated biochemicals (Phoenix) relative to the reference identification for test isolates.

	No. of isolates (%) identified correctly				
	MALDI-TOF MS Result				
Organism (no. tested)	**Genus-level model**	**Species-level model**	**Final result [a]**	**Serotyping**	**Phoenix**
S. sonnei (18)	18 (100)	17 (94)	17 (94)	13 (72)	16 (89)
S. flexneri (11)	11 (100)	10 (91)	10 (91)	11 (100)	8 (73)
S. boydii (2)	2 (100)	2 (100)	2 (100)	2 (100)	2 (100)
S. dysenteriae (2)	1 (50)	1 (50)	1 (50)	2 (100)	1 (50)
All *Shigella* spp. (33) [b]	32 (97)	30 (91)	30 (91)	28 (85)	27 (82)
E. coli, typical (16)	15 (94)	15 (94)	15 (94)	—	16 (100)
E. coli, inactive (20)	19 (95)	18 (90)	17 (85)	15 (75)	16 (80)
All *E. coli* (36) [c]	34 (94)	33 (92)	32 (89)	—	32 (89)
Total (69)	66 (96)	63 (91)	62 (90)	43 (81) [d]	59 (86)

[a] Final MALDI-TOF identification was accepted when results from genus-level and species-level models were in agreement.
[b] All *Shigella* species combined.
[c] Typical and inactive *E. coli* combined.
[d] 53 isolates tested by serogrouping. Typical *E. coli* isolates were not tested.
Adapted with permission from Khot, P. D. *et al.*, *J. Clin. Microbiol.*, **51**, 3711–3716 (2013), doi:10.1128/JCM.01526-13, Copyright © American Society for Microbiology.

Table 6.4 Discrepant results between reference identification and hybrid MALDI-TOF MS results.

		ID based on MALDI			PCR amplification	
Isolate	**Reference identification**	**Genus-level**	**Species-level**	**Final**	***lacY* gene**	***ipaH* gene**
22 [a]	*E. coli* (inactive)	*E. coli*	*S. sonnei*	Further workup	Neg	Neg
82	*S. sonnei*	*Shigella* spp.	*E. coli*	Further workup	Neg	Pos
102	*E. coli* (typical)	*Shigella* spp.	*S. sonnei*	*S. sonnei*	Pos	Neg
123	*E. coli* (inactive)	*Shigella* spp.	*E. coli*	Further workup	Pos	Neg
124	*E. coli* (inactive)	*E. coli*	inconclusive	Further workup	Pos	Neg
129	*S. flexneri*	*Shigella* spp.	*S. boydii*	*S. boydii*	Neg	Pos
136	*S. dysenteriae*	*E. coli*	inconclusive	Further workup	Neg	Pos

[a] Although the *lacY* gene for isolate #22 did not amplify, it was determined to be *E. coli* on the basis of *ipaH* gene PCR and biochemical tests.
Adapted with permission from Khot, P. D. *et al.*, *J. Clin. Microbiol.*, **51**, 3711–3716, (2013), doi:10.1128/JCM.01526-13, Copyright © American Society for Microbiology.

using the traditional approaches of automated biochemical and serogrouping were compared to that of the hybrid MALDI-TOF MS model. Surprisingly, the hybrid model, with 90% overall accuracy, outperformed both traditional methods when directly compared as stand-alone tests (Table 6.3). The automated biochemical method (Phoenix) identified 86% of isolates correctly, whereas serogrouping accurately identified only 81% of tested isolates. The latter comparison excluded typical *E. coli* because they are not routinely tested by serogrouping. In addition, because of the redundancy of the two-step algorithm in the MALDI-TOF MS approach, only two of the seven (28.6%) discrepant isolates were incorrectly reported, whereas both the Phoenix and serotyping incorrectly reported nine of ten (90%) discrepant isolates. The results of this direct comparison indicate that the advanced MALDI-TOF MS approach described here may be able to replace traditional identification methods while maintaining patient care.

6.4 Discussion and Implications

Since the initial discovery of *S. dysenteriae* in the late 1800s, it has been considered important to distinguish *Shigella* spp. from the mostly commensal *E. coli* isolated from diarrheal stools [29,30]. Although genetically considered the same species [16–19]), typical isolates can usually be distinguished using traditional biochemical methods, including lactose fermentation, indole production, motility, lysine decarboxylase activity, and gas production from glucose. Unfortunately, atypical and "inactive" isolates are much less reliably differentiated using these and other biochemical tests [23]. Interestingly, previously published mass spectrometry studies described some level of resolution of *E. coli* and *Shigella* spp. [31–34]; however, the currently available commercial MALDI-TOF MS systems are unable to reliably differentiate these organisms (Figure 6.1) [5,10]. Taken together, this suggested that distinguishing *E. coli* and *Shigella* spp. might be feasible using MALDI-TOF MS coupled with advanced data analysis methods.

ClinProTools is a software package for the visualization and statistical analysis of mass spectrometry data. This software allows users to view statistical parameters describing peaks found among "training sets," or isolates of known identity chosen for model generation. This data, known as the Peak Statistic Table, allows for a hands-on approach to biomarker discovery, as was used in our semiautomated model approach. ClinProTools also includes advanced algorithms for the generation of classification models. The goal of such "classifiers" is to allow an organism's identity (at the genus, species, strain, or other classification level) to be predicted based on spectral patterns observed among training set isolates. The software allows the user to rapidly assess models generated by three different algorithms: the genetic algorithm, supervised neural network, and quick classifier. It is able to generate models in an automated manner with little user intervention, yet is flexible enough to allow the user to modify the algorithm parameters, and even the peaks ultimately included in or excluded from the model. As part of the model generation process, Cross Validation and Recognition Capability statistics are calculated to help assess the likely performance of the models on unknown isolates. These statistics were used to determine that the genetic algorithm provided the most robust models, and that our hybrid model was superior to the automated model generated using default ClinProTools settings. Recent studies have also

used ClinProTools in the development of models to distinguish isolates at the strain level (*Staphylococcus aureus*) or closely related species (*Streptococcus pneumoniae* and *S. mitis*) [35,36]. In both of these studies, the genetic algorithm was either the optimal method, or performed as well as other methods in classifying these organisms. Neither of these studies required such complex models as described here. In fact, single models using only three peaks were successful in both cases, whereas our models required 12 to 15 peaks and a two-model interrogation algorithm to achieve optimal performance. This may very well be a testament to the high degree of spectral overlap between *E. coli* and *Shigella* spp., but may also be due to the fact that our models were designed to distinguish up to five classes of organisms (*E. coli* and four *Shigella* spp.).

Among the three sets of models described here, the semiautomated models were the least effective overall. However, they correctly identified most *Shigella* spp., indicating that MALDI-TOF does have the discriminatory power to resolve these closely related organisms. It is interesting to note that a pair of discriminatory peaks identified using this approach (9710 and 9737 *m/z*) were shown using TagIdent (http://web.expasy.org/tagident/) to match very well with the mature forms of the acid stress chaperone HdeA from *E. coli* and *Shigella*, respectively. These proteins differ by two amino acids (K92Q and S94N in *E. coli* vs. *Shigella*), which accounts for the observed mass difference of 27 Da. Unfortunately, some strains of *E. coli* carry the 9737 *m/z* isoform of this protein, leading to somewhat reduced specificity of these markers (Figure 6.4). In fact, the semiautomated models lacked specificity overall in that the majority of misidentifications

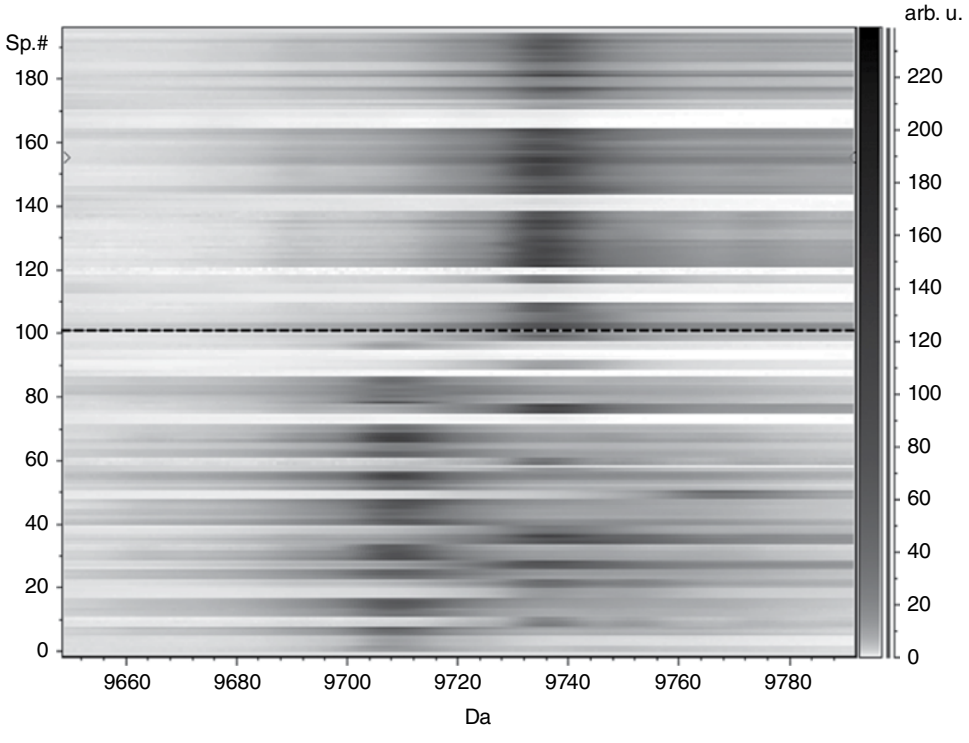


Figure 6.4 Peaks likely corresponding to amino acid changes in the acid resistance chaperone HdeA.

(88%) were *E. coli* isolates reported as *S. sonnei*. Although it would seem preferable to overcall rather than miss *Shigella* spp., these models were insufficiently specific for routine clinical use, forcing the evaluation of more sophisticated modeling approaches.

The automated models showed significantly better performance in correctly identifying *E. coli* compared to the semiautomated model. There was surprisingly little overlap between the models generated by these two approaches – only the 5096 and 8324 *m/z* peaks were shared. Interestingly, there were similarities among discriminatory peaks from other studies. Everley *et al.* found two peaks at 7287 ± 2 and 16886 ± 2 *m/z* that were present in *S. flexneri* but not *S. sonnei* or *E. coli* [33]. The first likely corresponds to the 7288 *m/z* peak in our semiautomated model, and the second may correspond to the 8444 *m/z* peak (as $[M+2H]^{2+}$ ions) in our automated species-level model. The TagIdent tool suggested that the 7288 peak corresponds to the 50S ribosomal protein L29 of *S. flexneri*, which differs from the same proteins in *E. coli* and *Shigella sonnei/boydii* by +15 Da (7273 Da) and in *S. dysenteriae* by −14 Da (7302 Da). The 7302 peak was also identified as discriminatory for *S. dysenteriae* in our semiautomated approach. In thesis work by van de Wiel *et al.*, features corresponding to the 9736, 10458, and 11706 *m/z* peaks in our models were described as being discriminatory [37]. Although these matching peaks substantiate our data, the lack of more corroborating peaks may indicate that standardizing MS-based differentiation of these organisms across instruments or platforms could be difficult. Consistent with this concern, subsequent attempts to analyze data acquired on different instruments have been met with limited success, likely due to multiple factors ranging from sample preparation to variations in instrument performance.

Because of the limited overlap between our initial models, it seemed counterintuitive that forcing peaks from the semiautomated model could improve upon the automated models. However, the improvements in the Cross Validation and Recognition Capability scores between the automated and hybrid models were notable, especially for the species-level model. As suggested in the ClinProTools documentation, these scores predicted the improved performance of the hybrid models and may generally serve as useful indicators of models' potential for success relative to each other. The fact that manually selected peaks improved the automated models stressed the idea that visual and statistical analyses may complement each other, and one should not simply trust a model without a detailed review of the selected peaks. Although the hybrid model outperformed the automated model to a modest degree, the automated model still performed well – at a level equivalent to the FDA-cleared Phoenix automated biochemical system and the traditional gold-standard method of serogrouping (Table 6.3).

Other analytical approaches have been used in an attempt to resolve *E. coli* and *Shigella* spp. In a study by van de Wiel *et al.*, a more complex experimental and statistical approach using the elastic net modeling method was applied in order to improve the distinction of these species [37]. The authors considered a hierarchical analysis approach to be important in order to capitalize on the stepwise nature of bacterial evolution. This work provides an interesting alternative to the approaches used in the present study and further demonstrates that the current inability to distinguish between closely related organisms might be due more to insufficient data analysis than limitations in the technology.

Serogrouping by the agglutination of cells using antisera directed against the O-antigen has long been used in the identification of *Shigella* spp., and is still recommended by most

authorities [23,38]. This method, although somewhat subjective in its interpretation – weak agglutination is considered negative and only strong agglutination should be used for positive identification – remains one of the few relatively reliable tests to distinguish atypical *E. coli* isolates from *Shigella* spp. It was somewhat unexpected that, although serogrouping performed reasonably well in identifying *Shigella* spp. overall (85% of 33 test isolates), it faltered with *S. sonnei*, failing to correctly identify 5 of the 18 isolates upon initial testing (Table 6.3). These were discovered only through PCR analyses and repeat serogrouping, which would not routinely be performed in most clinical laboratories, leaving some doubt as to the real-world performance of serogrouping in the identification of *Shigella* spp.

Automated microbiology instruments are widely used in clinical laboratories because of their generally reliable identification and antimicrobial susceptibility testing (AST) capabilities. The performance of the Phoenix system in this study was acceptable for the identification of *E. coli* (89% of 36 test isolates), but the system struggled with *Shigella* spp., with only 82% correctly identified (Table 6.3). Not surprisingly, the system also had difficulties with inactive *E. coli*, with only 16 of 20 (80%) isolates being correctly identified. These observations are not unprecedented, as the Phoenix and other automated systems have struggled to identify *Shigella* spp. and some *E. coli* isolates in other studies [39–42]. In one recent study, 4 of 23 *Shigella* spp. isolates were incorrectly identified as *E. coli* [39]. Among these, 75% were *S. sonnei* – by far the most commonly isolated *Shigella* species in the United States – which could have significant consequences not only in individual patients, but also in missed opportunities to limit the spread of these highly infectious organisms [21,43].

Among the primary limitations of this study was the small numbers of *S. dysenteriae* and *S. boydii* isolates, due to very low prevalence in the United States [43]. One approach to mitigating such a limitation would be to perform confirmatory testing on any isolate identified as either species by the models. Given the low frequency of isolation of these species, this approach would be required only rarely [43]. There may have also been limitations imposed by the modest numbers of other isolates. Because model generation was based on only half of the available isolates (to allow for an equally sized set of test isolates), the entire scope of diversity among our collection could not be encompassed. It is possible that a larger set of isolates could result in models with even better performance than those described here. One factor limiting the expansion of these models is the inability of ClinProTools to analyze spectra acquired at different digitizer sample rates. This could prevent the retrospective incorporation of data collected using different instrument settings over time into comprehensive, and potentially more accurate, models. A minor limitation is the fact that this method relied on protein extracts from isolates grown on MacConkey agar as opposed to whole-cell analysis of isolates harvested from SBA. Although the latter approach would certainly be more convenient, early investigations suggested this approach would produce less discriminatory models. Although the extraction method is simple and MacConkey agar is a routine laboratory medium, developing successful models from a more streamlined analytical process using a universal growth medium would be a substantial improvement over the current process. Other more complicated mass spectrometric methods utilizing liquid chromatography or nanoparticle-associated capture, digestion and subsequent MS/MS analysis have been shown to distinguish *E. coli* from some *Shigella* spp. [31,33], but the instruments and processing methods employed in these studies are not routinely available in clinical microbiology laboratories. Such methods would require further validation and optimization to be made compatible with the typical microbiology laboratory workflow.

The methods described here demonstrate that a simple, routine sample preparation method combined with a standard linear MALDI-TOF MS instrument generate data that, with advanced analytical tools, can address an important limitation in current commercial systems. Our models outperformed not only the Phoenix automated biochemical instrument, but also serogrouping, the gold-standard method. This approach could simplify the algorithm used by clinical laboratories for identification of *Shigella* spp., while reducing the response time for public health officials to intervene and control outbreaks.

Acknowledgments

I would like to acknowledge the contributions of Dr. Prasanna D. Khot, who performed the majority of the experiments and data analysis in this study. Ann Croft and the ARUP Bacteriology Laboratory provided substantial support in initial identification, isolate collection, and test interpretation. Dr. Eileen Barry, U. Maryland School of Medicine, generously provided two *S. dysenteriae* strains, and Chad Campbell, Utah Department of Health, provided an additional two isolates each of *S. dysenteriae* and *S. flexneri*.

References

1 J. H. Chen, P. L. Ho, G. S. Kwan, K. K. She, G. K. Siu, V. C. Cheng, K. Y. Yuen and W. C. Yam, Direct bacterial identification in positive blood cultures by use of two commercial matrix-assisted laser desorption ionization-time of flight mass spectrometry systems, *J. Clin. Microbiol.*, **51**, 1733–1739 (2013).

2 L. Ferreira, F. Sanchez-Juanes, J. L. Munoz-Bellido and J. M. Gonzalez-Buitrago, Rapid method for direct identification of bacteria in urine and blood culture samples by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: Intact cell vs. extraction method, *Clin. Microbiol. Infect.*, **17**, 1007–1012 (2011).

3 A. Bizzini, C. Durussel, J. Bille, G. Greub and G. Prod'hom, Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory, *J. Clin. Microbiol.*, **48**, 1549–1554 (2010).

4 A. Cherkaoui, J. Hibbs, S. Emonet, M. Tangomo, M. Girard, P. Francois and J. Schrenzel, Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level, *J. Clin. Microbiol.*, **48**, 1169–1175 (2010).

5 P. D. Khot, M. R. Couturier, A. Wilson, A. Croft and M. A. Fisher, Optimization of matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis for bacterial identification, *J. Clin. Microbiol.*, **50**, 3845–3852 (2012).

6 S. A. Neville, A. Lecordier, H. Ziochos, M. J. Chater, I. B. Gosbell, M. W. Maley and S. J. van Hal, Utility of matrix-assisted laser desorption ionization-time of flight mass spectrometry following introduction for routine laboratory bacterial identification, *J. Clin. Microbiol.*, **49**, 2980–2984 (2011).

7 P. Seng, M. Drancourt, F. Gouriet, B. La Scola, P. E. Fournier, J. M. Rolain and D. Raoult, Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted

laser desorption ionization time-of-flight mass spectrometry, *Clin Infect Dis*, **49**, 543–551 (2009).

8 S. Q. van Veen, E. C. Claas and E. J. Kuijper, High-throughput identification of bacteria and yeast by matrix-assisted laser desorption ionization-time of flight mass spectrometry in conventional medical microbiology laboratories, *J. Clin. Microbiol.*, **48**, 900–907 (2010).

9 D. Martiny, L. Busson, I. Wybo, R. A. El Haj, A. Dediste and O. Vandenberg, Comparison of the Microflex LT and Vitek MS systems for routine identification of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry, *J. Clin. Microbiol.*, **50**, 1313–1325 (2012).

10 Y. He, H. Li, X. Lu, C. W. Stratton and Y. W. Tang, Mass spectrometry biotyper system identifies enteric bacterial pathogens directly from colonies grown on selective stool culture media, *J. Clin. Microbiol.*, **48**, 3888–3892 (2010).

11 P. G. Saleeb, S. K. Drake, P. R. Murray and A. M. Zelazny, Identification of mycobacteria in solid-culture media by matrix-assisted laser desorption ionization-time of flight mass spectrometry, *J. Clin. Microbiol.*, **49**, 1790–1794 (2011).

12 P. Gerome, P. Le Fleche, Y. Blouin, H. C. Scholz, F. M. Thibault, F. Raynaud, G. Vergnaud and C. Pourcel, *Yersinia pseudotuberculosis* ST42 (O:1) Strain misidentified as *Yersinia pestis* by mass spectrometry analysis, *Genome Announc.*, **2**, (2014).

13 W. Kallow, M. Erhard, H. N. Shah, E. Raptakis and M. Welker, MALDI-TOF MS for microbial identification: Years of experimental development to an established protocol, in *Mass Spectrometry for Microbial Proteomics*, H. N. Shah and S. E. Gharbia (Eds.), John Wiley & Sons, Ltd, Chichester, 2010.

14 M.-F. Lartigue, G. Héry-Arnaud, E. Haguenoer, A.-S. Domelier, P.-O. Schmit, N. van der Mee-Marquet, P. Lanotte, L. Mereghetti, M. Kostrzewa and R. Quentin, Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by matrix-assisted laser desorption ionization-time of flight mass spectrometry, *J. Clin. Microbiol.*, **47**, 2284–2287 (2009).

15 C. A. Petti, P. P. Bosshard, M. E. Brandt, J. E. Clarridge, T. V. Feldblyum, P. Foxall, M. R. Furtado, N. Pace and G. Procop, *Interpretive Criteria for Identification of Bacteria and Fungi by DNA Target Sequencing, Approved Guideline (MM18-A)*, Clinical Laboratory and Standards Institute, Wayne, 2008.

16 D. J. Brenner, G. R. Fanning, G. V. Miklos and A. G. Steigerwalt, Polynucleotide sequence relatedness among *Shigella* species., *Int. J. Syst. Evol. Microbiol.*, **23**, 1–7 (1973).

17 O. Lukjancenko, T. M. Wassenaar and D. W. Ussery, Comparison of 61 sequenced *Escherichia coli* genomes, *Microb. Ecol.*, **60**, 708–720 (2010).

18 G. M. Pupo, D. K. Karaolis, R. Lan and P. R. Reeves, Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies, *Infect. Immun.*, **65**, 2685–2692 (1997).

19 G. M. Pupo, R. Lan and P. R. Reeves, Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics, *Proc. Natl. Acad. Sci. U S A*, **97**, 10567–10572 (2000).

20 M. J. van den Beld and F. A. Reubsaet, Differentiation between *Shigella*, enteroinvasive *Escherichia coli* (EIEC) and noninvasive *Escherichia coli*, *Eur. J. Clin. Microbiol. Infect. Dis.*, **31**, 899–904 (2012).

21 CDC, Shigellosis, in *CDC Health Information for International Travel 2014*, G. W. Brunette (Ed.), Oxford University Press, New York, 2014.

22 T. Wu, C. Grassel, M. M. Levine and E. M. Barry, Live attenuated *Shigella dysenteriae* type 1 vaccine strains overexpressing shiga toxin B subunit, *Infect. Immun.*, **79**, 4912–4922 (2011).

23 J. Versalovic, *Manual of Clinical Microbiology*, 10th edition, ASM Press, Washington DC, 2011.

24 M. Pavlovic, A. Luze, R. Konrad, A. Berger, A. Sing, U. Busch and I. Huber, Development of a duplex real-time PCR for differentiation between *E. coli* and *Shigella* spp, *J. Appl. Microbiol.*, **110**, 1245–1251 (2011).

25 D. T. Vu, O. Sethabutr, L. Von Seidlein, V. T. Tran, G. C. Do, T. C. Bui, H. T. Le, H. Lee, H. S. Houng, T. L. Hale, J. D. Clemens, C. Mason and D. T. Dang, Detection of *Shigella* by a PCR assay targeting the *ipaH* gene suggests increased prevalence of shigellosis in Nha Trang, Vietnam, *J. Clin. Microbiol.*, **42**, 2031–2035 (2004).

26 M. R. Couturier, E. Mehinovic, A. C. Croft and M. A. Fisher, Identification of HACEK clinical isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry, *J. Clin. Microbiol.*, **49**, 1104–1106 (2011).

27 J. H. Holland, *Adaptation in Natural and Artificial Systems : An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press, Cambridge, 1992.

28 M. Kearns, Y. Mansour, A. Y. Ng and D. Ron, An experimental and theoretical comparison of model selection methods, *Mach. Learn.*, **27**, 7–50 (1997).

29 J. R. Johnson, *Shigella* and *Escherichia coli* at the crossroads: Machiavellian masqueraders or taxonomic treachery?, *J. Med. Microbiol.*, **49**, 583–585 (2000).

30 S. K. Niyogi, Shigellosis, *J. Microbiol.*, **43**, 133–143 (2005).

31 W. J. Chen, P. J. Tsai and Y. C. Chen, Functional nanoparticle-based proteomic strategies for characterization of pathogenic bacteria, *Anal. Chem.*, **80**, 9612–9621 (2008).

32 G. C. Conway, S. C. Smole, D. A. Sarracino, R. D. Arbeit and P. E. Leopold, Phyloproteomics: Species identification of Enterobacteriaceae using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *J. Mol. Microbiol. Biotechnol.*, **3**, 103–112 (2001).

33 R. A. Everley, T. M. Mott, S. A. Wyatt, D. M. Toney and T. R. Croley, Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and *Shigella* species, *J. Am. Soc. Mass Spec.*, **19**, 1621–1628 (2008).

34 C. J. Keys, D. J. Dare, H. Sutton, G. Wells, M. Lunt, T. McKenna, M. McDowall and H. N. Shah, Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases, *Infect., Genet. Evol.*, **4**, 221–242 (2004).

35 S. R. Boggs, L. H. Cazares and R. Drake, Characterization of a *Staphylococcus aureus* USA300 protein signature using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *J. Med. Microbiol.*, **61**, 640–644 (2012).

36 L. N. Ikryannikova, A. V. Filimonova, M. V. Malakhova, T. Savinova, O. Filimonova, E. N. Ilina, V. A. Dubovickaya, S. V. Sidorenko and V. M. Govorun, Discrimination between *Streptococcus pneumoniae* and *Streptococcus mitis* based on sorting of their MALDI mass spectra, *Clin. Microbiol. Infect.*, **19**, 1066–1071 (2013).

37 L. van de Wiel, Differentiating *Shigella* from *E. coli* using hierarchical feature selection on MALDI-ToF MS data, Thesis for Computing Science, Radboud University Nijmegen, 2014.

38 L. S. Garcia, Clinical Microbiology Procedures Handbook, 3rd edition, ASM Press, 2010.

39 K. C. Carroll, B. D. Glanz, A. P. Borek, C. Burger, H. S. Bhally, S. Henciak and D. Flayhart, Evaluation of the BD Phoenix automated microbiology system for identification and antimicrobial susceptibility testing of Enterobacteriaceae, *J. Clin. Microbiol.*, **44**, 3506–3509 (2006).

40 P. J. Gavin, J. R. Warren, A. A. Obias, S. M. Collins and L. R. Peterson, Evaluation of the Vitek 2 system for rapid identification of clinical isolates of gram-negative bacilli and members of the family Streptococcaceae, *Eur. J. Clin. Microbiol. Infect. Dis.*, **21**, 869–874 (2002).

41 S. Gupta, C. Aruna and S. Muralidharan, Misidentification of a commensal inactive *Escherichia coli* as *Shigella sonnei* by an automated system in a critically ill patient, *Clin. Lab.*, **57**, 767–769 (2011).

42 C. M. O'Hara and J. M. Miller, Evaluation of the MicroScan rapid neg ID3 panel for identification of Enterobacteriaceae and some common gram-negative nonfermenters, *J. Clin. Microbiol.*, **38**, 3577–3580 (2000).

43 CDC, *National Shigella Surveillance Annual Report, 2011*, US Department of Health and Human Services, CDC, Atlanta, 2013.

7

Identification of Species in Mixed Microbial Populations using MALDI-TOF MS

Pierre Mahé,[1] Maud Arsac,[1] Nadine Perrot,[2] Marie-Hélène Charles,[1] Patrick Broyer,[1] Jay Hyman,[3] John Walsh,[3] Sonia Chatellier,[2] Victoria Girard,[2] Alex van Belkum,[2] and Jean-Baptiste Veyrieras[1]

[1] *bioMérieux SA, Marcy-l'Etoile, France*
[2] *bioMérieux SA, La Balme-les-Grottes, France*
[3] *bioMérieux Inc, Durham, NC, USA*

7.1 Introduction

Blood samples are among the most medically relevant patient specimens, and a positive bacterial culture derived from such a sample stimulates clinical interest. Antimicrobial treatment is imminent, and therapy is usually started by using empiric, broad-spectrum antibiotics. In most cases, blood cultures are monomicrobial: only between 5% and 10% may contain a mixture of different species. The nature of such species may differ depending on underlying disease, prior antibiotic usage, local prevalence of the organism and the ratio in which the bacterial species were present in the original clinical specimen. Adequate diagnostics of mixed infection depends on classical microbiology because purification of the individual species is needed, and this is easily managed by looking at individual colonies on a solid growth medium. This process usually involves overnight incubation and is considered slow, delaying the moment at which more tailored antibiotics can be prescribed to the patient.

In search of a faster diagnostic technology, matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) was explored as an alternative to selective cultivation. This involved developing sample preparation protocols that are able to quickly and efficiently extract bacterial cells from positive blood cultures, hence avoiding growing the microorganisms on solid agar plates (Fothergill *et al.*, 2013; Croxatto *et al.*, 2014; Mestas *et al.*, 2014). Although this approach has proved successful for the identification of monomicrobial cultures, commercially available systems failed to adequately detect mixtures of bacteria as such, and then also failed to identify two or more causative species (Chen *et al.*, 2013). Indeed, although studies were limited in specimen numbers, only a single species was identified in most cases (Kok *et al.*, 2011; Buchan *et al.*, 2012; Lagacé-Wiens *et al.*, 2012; Fothergill *et al.*, 2013). One of the main issues with this type of mixed specimens is in the software components of MALDI-TOF MS systems; these are unable to detect spectra that are built from a combination of

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

more than a single species fingerprint, and even if they do so correctly, then the subsequent step, the splicing of the combined spectrum into the individual fingerprints, becomes a problem. Although mixed cultures can sometimes be identified, this usually involves lowering the confidence score required to make an identification. Besides, this requires the intervention of a human expert to interpret the results and decide between an actual mixture of spectra, or simply a limited degree of discrimination between the species proposed. Although this approach can, and does, work in some cases where the mixed species are present in equal cell densities, one wonders if it will work for unbalanced mixtures. Indeed, in this latter case, as will be shown later in Figure 7.1, the spectral fingerprint of the minority species may only be partly revealed in the overall combined spectrum. Its score may therefore be significantly lower than that of the majority species, and probably also lower than that of its close relatives (e.g., other species of the same genus). This calls for alternative, or at least complementary, algorithms explicitly interpreting the MALDI-TOF MS spectrum as a potential combination of microbial fingerprints, in order to deal with polymicrobial samples.

Up to now, no method has been described that automatically and adequately infers the composition of a polymicrobial sample after MALDI-TOF MS, and in those few cases where software has been developed, their performance has not been studied in detail (Wahl *et al.*, 2002; Schleif *et al.*, 2011). In this chapter, we introduce a new method developed to infer the composition of polymicrobial samples on the basis of a single mass spectrum. Starting from a MALDI-TOF mass spectrum preprocessed by the VITEK®MS system (bioMérieux, France), as is done for all culture-based MS identifications, the method automatically predicts, within certain limitations, how many and which species are present in the sample, and provides an estimation of their relative concentrations. Our approach relies on a penalized nonnegative linear regression framework making use of species-specific prototypes that can be derived directly from the routine reference database of single-species spectra. Building upon our previous work in this area (Mahé *et al.*, 2014), we present here further proof of concept toward the clinical application of this method to identify microbial cells extracted from positive blood cultures, and lay the groundwork for future developments.

7.2 A New Algorithm to Identify Mixed Species in a MALDI-TOF Mass Spectrum

7.2.1 Mixed Spectrum Model

In the following sections, we consider a peak-list representation of mass spectra in which a mass spectrum is represented by a vector $x \in \mathbb{R}^p$, where p is the number of channels or "bins" involved in the peak-list representation, and each entry x_b is derived from the intensity of the peak(s) falling in the bth bin. Several schemes have been proposed to define such a representation (Coombes *et al.*, 2007). In this work, we rely on the approach embedded in the VITEK®MS system, which provides a representation defined on $p = 1300$ bins, covering a mass window ranging from 3 to 17 kDa. We also assume that we have a reference database of mass spectra covering a panel of K microbial species, which in our case corresponds to, or, depending on the application, may be extracted from, the reference database embedded in the VITEK®MS system. Given this

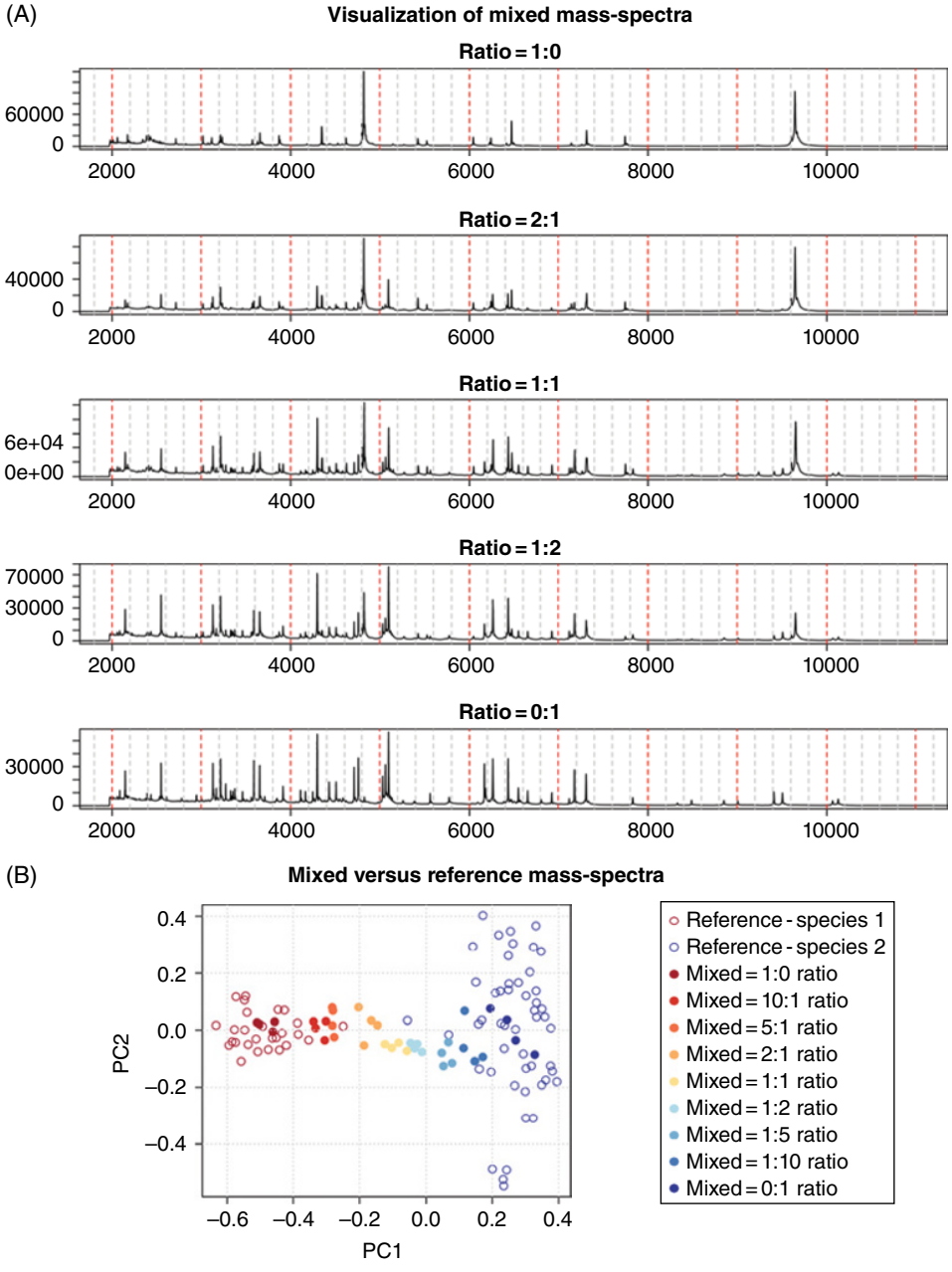


Figure 7.1 **Artificial mixtures prepared in vitro.** Artificial mixtures were prepared in vitro by diluting and mixing two calibrated monobacterial suspensions to obtain bimicrobial suspensions at the following relative concentrations: 1:0, 10:1, 5:1, 2:1, 1:1, 1:2, 1:5, 1:10, and 0:1. (A) *Visualization of mixed mass spectra.* Five smoothed mass spectra obtained at concentrations 1:0, 2:1, 1:1, 1:2, and 0:1 are shown on a 2–12 kDa grid. The top (resp. bottom) panel represents the pure spectra of the first (resp. second) species of the mixture, and the panels in between represent the spectra obtained when the proportion of the second species increases. We note that peaks specific for the second species gradually appear and increase with its relative proportion, whereas the peaks specific of the first species gradually decrease and disappear. (B) *Mixed versus reference mass spectra.* A principal component analysis (PCA) was carried out from the spectra of the reference database corresponding to the two species involved in these mixtures, shown as red and blue empty circles. The spectra obtained from the in vitro mixtures were then projected in the PCA space and shown as filled circles, with the color turning from red to blue as the relative proportion of the second species increases. We note a remarkably smooth transition from the first to the second species as their relative concentration varies. This PCA analysis was carried out from peak-list representation of the spectra. These artificial mixed spectra correspond to the mixture E involved in Mahé *et al.*, 2014, and are available online at the UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/MicroMass).

dataset, we address the problem of trying to predict which of the K species are actually present in a spectrum to analyze.

For that purpose, we model a spectrum x as a positive linear combination of species-specific prototypes built from the reference database. In its simplest form described below, our algorithm considers a single prototype per species, but it can easily be extended to accommodate a variable number of prototypes per species to optimally capture their degree of spectral variability. Our mixed spectrum model can therefore be formally written as follows:

$$x = \sum_{i=1}^{K} \beta_i P_i + \varepsilon,$$

Where $P_i \in \mathbb{R}^p$ is the prototype spectrum representing species i, the coefficient $\beta_i \geq 0$ accounts for its contribution in explaining the input spectrum x, and $\varepsilon \in \mathbb{R}^p$ is a vector of independent and identically distributed random residuals, assumed to be normally distributed.

Although this model may seem naïve and overly simplistic, it has worked in practice (see Mahé *et al.* and the results described therein) and seems to properly reflect the combined mass spectra. Indeed, as shown in Figure 7.1, we observed on artificial bimicrobial mixtures prepared in vitro that mixed spectra exhibited a smooth transition between their two constituent species as their relative concentrations varied.

7.2.2 Algorithm Description

Given a spectrum to identify, the problem therefore translates into that of obtaining an estimation $\widehat{\beta}$ of the vector $\beta = [\beta_1, \ldots, \beta_K]$ on the basis of the species prototypes $P_1, \ldots P_K$. A species is then predicted to be present in the sample whenever its corresponding entry in the vector $\widehat{\beta}$, estimated from the acquired mass spectrum, is positive. We note that the spectrum can be predicted to be polymicrobial as soon as more than one of these coefficients is positive. In addition, an estimation of the relative concentration c_i of the microbial species $i = 1, \ldots, K$ can be empirically derived from the model parameters according to, for instance, $\widehat{c_i} = \widehat{\beta_i} \Big/ \sum_{j=1}^{k} \widehat{\beta}_j$.

In its simplest form, our algorithm is a three-step process. The first step amounts to defining the species prototypes from the reference database of mass spectra. For that purpose, we carry out a simple procedure in which we consider a minimum frequency threshold to introduce a peak at a given position in a prototype: a peak is introduced in the prototype provided that it appears sufficiently often in the individual spectra of the corresponding species. Its intensity is then defined as the median intensity of the reference spectra that exhibit this peak. This procedure is done off-line, once and for all, and does not need to be repeated, because it forms part of the basic knowledge base from that moment onward. Given a spectrum to identify, we then start by defining a list of candidate (poly)microbial compositions. For typical clinical specimens, we expect the number of distinct species found in the sample to be relatively small. We therefore rely on a penalized nonnegative linear regression framework making use of the L1 penalty to favor sparsity in the vector β of model coefficients. More precisely, we consider the following optimization problem to estimate:

$$\widehat{\beta} = \underset{\beta \in \mathrm{R}^{K+}}{\arg\min} \left\| x - P\beta \right\|^2 + \lambda \left| \beta \right|_1$$

where $P = [P_1,\ldots,P_K] \in \mathbb{R}^{P\times K}$ is a matrix containing the K prototypes, and $|\beta|_1 = \sum_{i=1}^{K}\beta_i$ is the L1 norm of β. This problem is the standard Lasso problem (Tibshirani, 1996) with the addition of a nonnegativity constraint. Solving this problem for increasing values of the λ parameter provides solutions with a decreasing number of non-null coefficients in β, and hence a collection of models achieving various trade-offs between reconstruction error (small λ values leading to many positive coefficients) and sparsity of the solution (high λ values resulting in few positive coefficients). In practice, efficient algorithms make it possible to access the whole collection of solutions that can be reached when the parameter λ is varied, collectively referred to as the *regularization path* of the problem. We therefore propose to use the LARS-EN algorithm (Zou and Hastie, 2005) for that purpose, which can be easily modified to accommodate an additional nonnegativity constraint (Efron *et al.*, 2004). This procedure therefore provides us with a list of (usually nested) spectrum compositions involving an increasing number of species. Finally, to select the most plausible solution, we need to choose where to stop on the regularization path and find the number of species prototypes achieving the appropriate trade-off between the quality of the spectrum approximation and the sparsity of the solution. To do so, we base our model selection strategy on the Bayesian information criterion (BIC). The log-likelihood component of the BIC is derived from a standard (unpenalized) linear regression model and is directly related to a least-squares residual, which can naturally be interpreted in this case as the spectrum reconstruction error. The BIC penalizes this error by the complexity of the reconstruction model, which corresponds here to the number of components involved in the reconstruction, and we select the model that minimizes the BIC along the regularization path.

In practice, however, we observed a correlation structure between prototypes that reflects the taxonomic proximity of the corresponding species. Therefore, even if a spectrum to identify involves a single microbial fingerprint, its optimal decomposition in terms of residual sum of squares may involve the contribution of several prototypes, depending on their level of correlation with the prototype of the species actually present in the sample. This phenomenon is highly species dependent, and can lead to the prediction of additional erroneous components in the decomposition obtained by the above procedure. To overcome this issue, we proposed, in the spirit of Lindner and Renard (2013), to introduce $\gamma = [\gamma_1,\ldots,\gamma_K]$, the vector of unknown positive contributions of each of the K species to the spectrum, and to redefine $\beta_i = \sum_j a_{ij}\gamma_j$, where a_{ij} is a predefined measure of similarity between species i and j. We postulated that the estimations obtained by the procedure described above were noisy because a species actually present in the spectrum may "turn on" other prototypes of similar species. To estimate the vector of actual species contributions γ, we note that the original mixed spectrum model $x = \sum_{i=1}^{K}\beta_i p_i + \epsilon$ can be equivalently expressed as

$$x = \sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij}\gamma_j P_i + \varepsilon = \sum_{j=1}^{K}\gamma_j P_j^{(a)} + \epsilon, \qquad \textit{where} \quad P_j^{(a)} = \sum_{i=1}^{K} a_{ij} P_i.$$

We can therefore obtain an equivalent mixed spectrum model, but expressed in terms of the actual species contributions, after an appropriate redefinition of the prototypes, which we shall refer to as *adjustment*. Using the adjusted prototypes instead of the original ones in the above procedure therefore allows one to estimate directly the vector of coefficients γ that is used ultimately to predict the composition of the spectrum. We have empirically observed that the prototype adjustment is of critical importance for the success of the method. Although several definitions can be considered for the similarity measure a_{ij}, we have found the Jaccard coefficient to be a safe choice. Figure 7.2 illustrates the whole procedure, and we refer the interested reader to Mahé *et al.* (2014) for further details about this algorithm and its application.

7.2.3 A Simulation Framework to Optimize the Model Parameters

In its current form, the above algorithm involves three main tuning parameters: the frequency threshold involved in the prototype construction method, the spectrum similarity measure to consider for the adjustment, and the choice of the likelihood component involved in the BIC (which may involve an intercept or not). These parameters are difficult to choose a priori. Instead, we propose to rely on a simulation framework, which is described in detail in Mahé *et al.* (2014) and that we briefly recall here.

The idea underlying this simulation framework is to generate artificial mixtures by combining spectra of the reference database, that is, monomicrobial spectra obtained from colonies, using the same additive linear model as the one used for the decomposition algorithm. Large databases of mixed spectra can therefore be generated and used to study in detail the behavior of the algorithm, which is useful in two respects. First, it provides an objective way to set the algorithm parameters. It may indeed be necessary, or at least beneficial, to optimize these parameters with respect to the reference dataset considered, which can be application dependent. Moreover, it allows one to estimate the level of performance that may be reached for the reference dataset considered and to study the mixture identification performance on a (pair of) species per (pair of) species basis. Indeed, our previous work has shown that the performances obtained by simulation and on real in vitro samples were relatively consistent when the simulated dataset matched the experimental one (Mahé *et al.*, 2014). This was especially the case for balanced mixtures, the performance obtained by simulation being optimistic for unbalanced mixtures, and quite consistently observed across various pairs of mixed species. This simulation framework could therefore be useful, in particular, to identify pairs or groups of species that are too close to be efficiently separated from a mixed fingerprint and that thus should be grouped as a composite class.

7.3 Toward Direct-Sample Polymicrobial Identification from Positive Blood Cultures

In this section, we present a proof of concept toward the application of the algorithm presented in the previous section to correctly identify microbial cells extracted from positive bloodcultures containing either single or mixed species. Experiments involved mono- and bimicrobial spectra obtained in silico, generated by the simulation framework mentioned above, and in vitro, by spiking microorganisms into blood culture

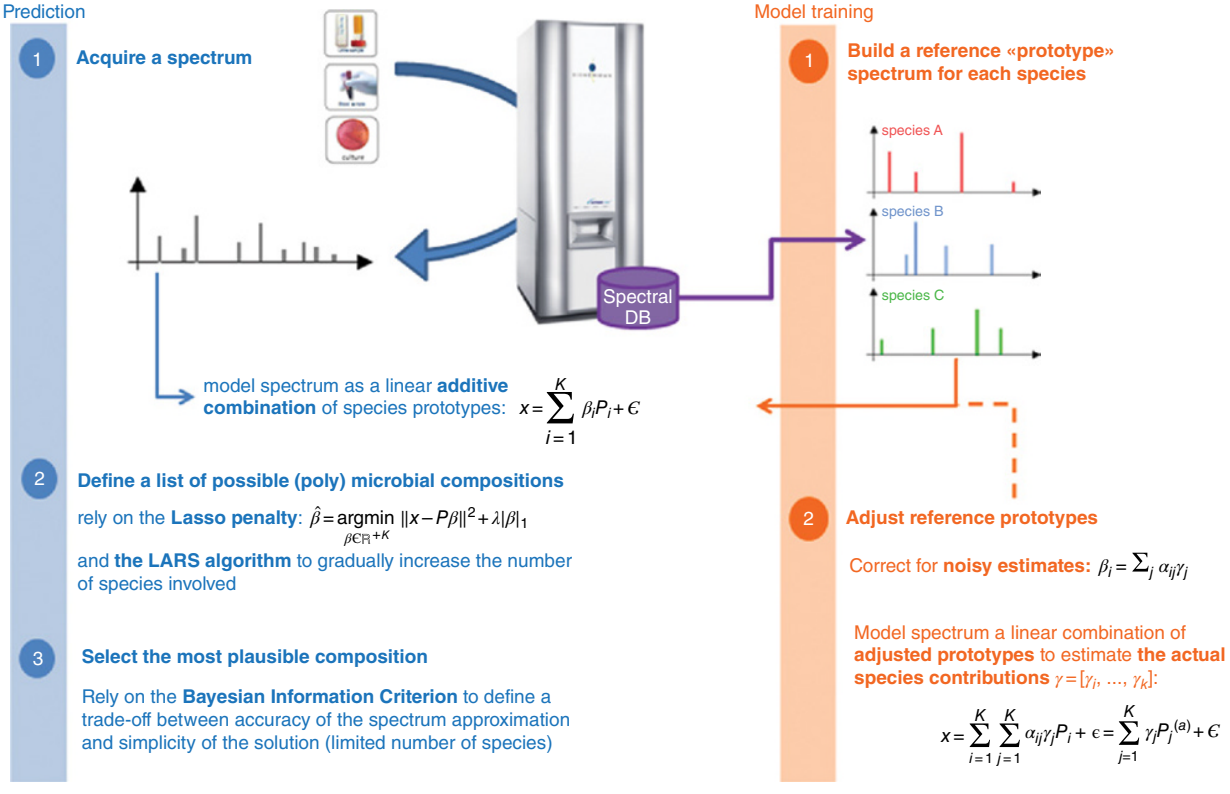


Figure 7.2 **The mixed identification algorithm.** Shown on the right-hand side (orange) are the operations done off-line to prepare the species-specific prototypes. First, a prototype must be built for each reference species considered (shown here in red, blue, and green for species A, B, and C). These prototypes are derived from the reference spectra database embedded in the commercial VITEK®MS system. Then, although not strictly mandatory, the prototypes should be adjusted in order to limit the risk of obtaining erroneous decompositions in subsequent analyses. Shown on the left-hand side (blue) are the operations carried out to predict the composition of a sample on the basis of these reference prototypes. First, a spectrum is acquired and preprocessed by the standard algorithm used for routine (culture-based) identification, in order to extract its prominent peaks. A list of candidate (poly)microbial compositions is then obtained by means of a nonnegative linear regression framework involving the Lasso penalty and the LARS algorithm. Finally, the most plausible decomposition, achieving a trade-off between the accuracy of the approximation of the spectrum as a combination of prototypes and the number of species involved, is selected by the Bayesian information criterion (BIC).

bottles. Before discussing the results obtained, we first present the reference microbial panel considered and define the indicators used to measure the performance of the algorithm.

7.3.1 Microbial Panel Considered

We relied on a small-sized microbial panel involving 26 bacterial and fungal species chosen for their high prevalence in bloodstream infections. This list, given in Table 7.1, covers a large fraction of species frequently documented in recent clinical studies related to direct microbial species identification from positive blood cultures (e.g., Lagacé-Wiens *et al.*, 2012; Chen *et al.*, 2013; Fothergill *et al.*, 2013). For each of these species, between 28 and 171 spectra could be extracted from the spectral database embedded in the VITEK-MS database, leading altogether to a reference database of 1617 spectra. Although this setting might look somehow restrictive, it allowed one to conveniently study the behavior of the method, and, as will discussed later on, constitutes a natural first step toward more challenging settings involving a larger number of species.

7.3.2 Qualifying the Success of the Identification

The main objective of our work was to evaluate the ability of the method to *detect* a microbial mixture and to *identify* its components. We defined the *sensitivity* and *specificity* of mixture detection, respectively, as the proportion of mixed and pure spectra detected as such by the method. Detection of a mixture was considered to be successful if two or more species were predicted. A mixture was said to be correctly identified whenever all of its species components – and only those components – were detected, and partially identified when only some of its species components were detected. A misidentification occurred whenever a species that was not part of the spectrum was predicted.

Table 7.1 List of species considered for the proof of concept related to direct identification from microbial cells extracted from positive blood culture bottles.

Acinetobacter baumannii	*Staphylococcus aureus*
Citrobacter freundii	*Staphylococcus capitis*
Enterococcus faecium	*Staphylococcus epidermidis*
Enterococcus faecalis	*Staphylococcus hominis*
Enterobacter aerogenes	*Staphylococcus saprophyticus*
Enterobacter cloacae	*Staphylococcus haemolyticus*
Escherichia coli	*Streptococcus agalactiae*
Haemophilus influenzae	*Streptococcus oralis*
Klebsiella oxytoca	*Streptococcus pneumoniae*
Klebsiella pneumoniae	*Streptococcus pyogenes*
Proteus mirabilis	*Candida albicans*
Pseudomonas aeruginosa	*Candida glabrata*
Serratia marcescens	*Candida parapsilosis*

7.3.3 In Silico Experiments

We begin by presenting results obtained in silico by using exclusively the spectra of the reference database, which were hence obtained according to the standard workflow used in clinical routine in which the corresponding microorganisms were grown on an agar plate.

The first experiment we considered relied on the simulation framework described above and aimed (1) to set the algorithms parameters and (2) to evaluate the performance that could be achieved on a (pairs of) species per (pairs of) species basis. For that purpose, we generated a simulated dataset by repeating 20 times the following procedure in which, for every pair of species of the reference panel, we first randomly picked two spectra from the reference database that were then used to simulate mixtures with relative proportions in 100%–0%, 90%–10%, 80%–20%, …, 20%–80%, 10%–90%, 0%–100% by means of linear combinations. The weights used to implement these linear combinations were the same as those used in our previous work, and we refer the interested reader to the original publication for a more detailed description. This procedure led to a dataset of 71.500 spectra,[1] which was then used to evaluate the performance obtained with several parameterizations of the algorithm, differing in the following:

- The peak frequency threshold involved in the prototype construction, taken as the fractions 0.2, 0.3, 0.4, 0.5, and 0.6.
- The similarity criterion used to adjust the prototypes, which was defined as a binary function (hence corresponding to not adjusting the prototypes), as their Jaccard or as their cosine coefficient.
- The definition of the linear model involved in the likelihood component of the BIC, and hence in the final model selection, that could integrate an intercept or not.

The resulting 30 configurations were evaluated by decomposing the spectra of the simulated dataset: prototypes were built from the reference dataset, adjusted if necessary, and used to decompose each spectrum according to the LARS/BIC procedure.[2] The results obtained with the selected configuration are shown in Figure 7.3. We first noted (Figure 7.3A) that 94.1% of the 13.000 spectra that were actually monomicrobial were correctly identified. Bimicrobial spectra (58.500 spectra, 82% of the simulated dataset) were correctly identified in 80.8% of the cases. The vast majority of the remaining mixed spectra (16.3%) were partially identified, and globally, only 3.4% of the spectra were misidentified. We noted, however (Figure 7.3B), that the probability of correctly identifying a mixture depended on the species involved. In particular, mixtures involving *Pseudomonas aeruginosa* or any of the three *Candida* species could be correctly identified in more than 87% of the cases. Conversely, *Enterobacter cloacae* was the hardest to identify, with a correct identification rate of 71.2%. Among the 26 species, 21 had a misidentification rate below 3%. The 5 remaining ones had error rates between 3% and 6%, except *Streptococcus oralis*, which showed a misidentification rate of 16.7%. Taking a closer look at the misidentifications, we realized that more than half (2% out of 3.4%

1 $(26 \times 25/2)$ possible combinations of 2 species, times 11 relative concentration, times 20 repetitions.

2 We note, however, that this procedure was actually repeated 6500 times; that is, for each of the 325 $(26 \times 25/2)$ pairs of species and each of the 20 repetitions, in order to exclude the spectra used to generate the simulated spectra from the reference dataset before building the prototypes.

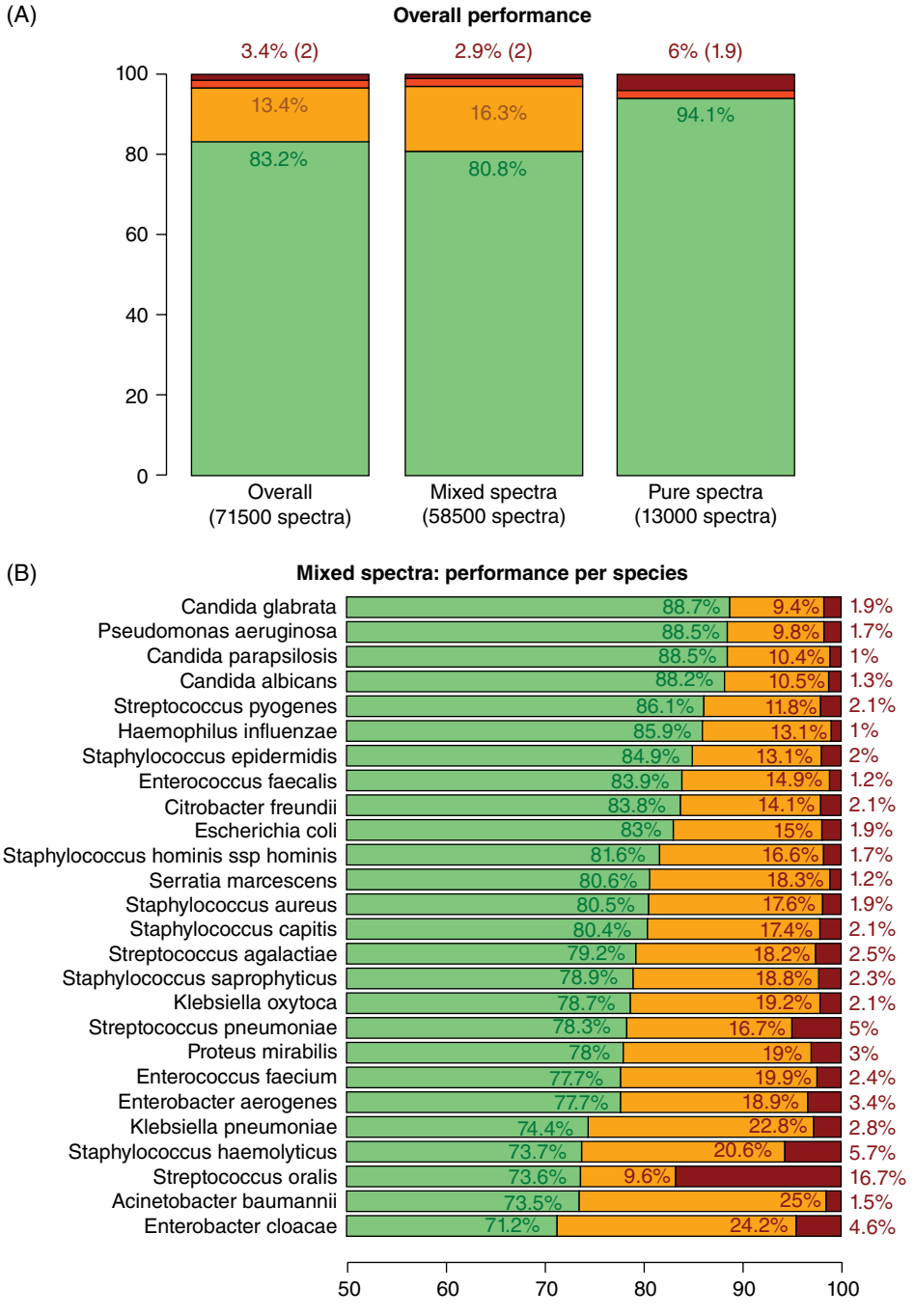


Figure 7.3 **In silico results – simulated mixtures.** (A) *Overall performance* in terms of correct (green), partial (orange), and erroneous (red) identification computed globally (left), and on mixed and pure spectra only (middle and right, respectively). The proportions of erroneous identification are split into same-genus misidentifications that involve prediction of species that belong to the same genera as the ones actually present in the samples (light red, second figure shown on top of the bar) and those involving species of other genera (dark red). (B) *Mixed spectra:* Species-level performance. Performance obtained among mixed spectra involving each of the species of the reference dataset. (C) *Same-genus misidentifications*: Proportion of genera involved. The great majority of same-genus errors involve confusing streptococci and to a lesser extent staphylococci. (D) Streptococci *confusion*. The great majority of same-genus errors involving *Streptococci* amount to confusing *Streptococcus oralis* and *S. pneumoniae*.

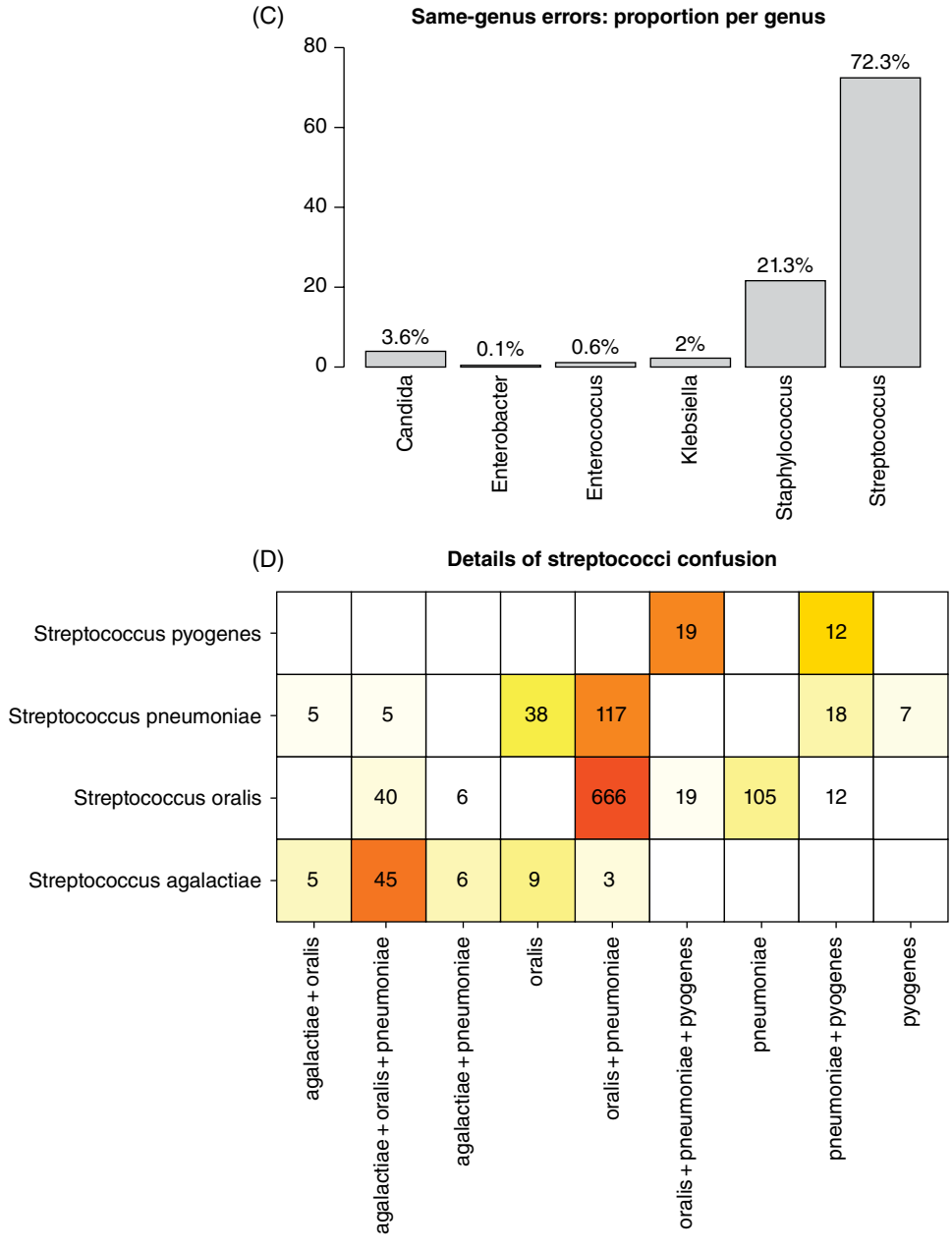


Figure 7.3 (Continued)

of misidentified spectra) involved mistaking a species for another one of the same genus, which was predicted instead of it, or in addition to it. It turned out that the probability of such same-genus misidentifications depended on the genus itself. Indeed, as shown in Figure 7.3C, more than 70% and 20% of these same-genus errors, respectively, involved *Streptococcus* and *Staphylococcus*. Although issues with *Streptococcus* have been observed independently (Buchan *et al.*, 2012), this observation may be explained in part

by the fact that these two genera are the most represented in the reference dataset considered (four and six species, respectively). Nevertheless, a detailed analysis of the streptococcal misidentifications, shown in Figure 7.3D, revealed that more than 80% involved predicting *Streptococcus pneumoniae* instead of, or together with *S. oralis*, and vice versa. This indicated that such same-genus misidentifications may not be uniformly distributed but involve specific pairs of species. We also noted that the probability of mistaking *S. oralis* for *S. pneumoniae* was much higher than the opposite. This may be an artifact of our algorithm that may be fixed with more elaborate ways of defining the prototypes. As a final remark, we noted that the configuration of the algorithm selected by this simulation study was slightly different than the one retained in our previous work. Here, the peak frequency threshold was set to 0.3, whereas it was set to 0.4 with the reference dataset considered earlier. This illustrated the relevance of this simulation framework to objectively set the algorithm parameters with respect to the (application-dependent) reference dataset.

The second in silico experiment aimed to further evaluate the specificity of the algorithm, that is, its ability to correctly detect and identify a monomicrobial spectrum as such. For that purpose, we relied on a cross-validation procedure using the spectra of the reference dataset. The dataset was evenly split in ten subsets, subsequently called folds. We then repeated ten times the following procedure in which species prototypes were built and adjusted from nine of the ten folds, using the parameters selected by the previous simulation study, and used to identify the spectra of the remaining fold. We noted that the method was highly specific: 96.9% of these monomicrobial spectra were detected as such, and 93.6% were correctly identified at the species level. We noted, however, that on the same cross-validation experiment, a support vector machine algorithm could reach up to 98.8% of correct identification. This therefore indicated, as could be expected, that the propensity of the algorithm to identify polymicrobial samples comes at a price in its ability to correctly identify pure cultures. The proposed algorithm nevertheless remained competitive with this purely discriminative approach, which often reached state-of-the-art performance on various classification tasks. Moreover, a natural synergy between both approaches could be envisioned, in which the identification of a sample predicted to be monomicrobial by the mixture algorithm could be ultimately provided by such a discriminative algorithm. This procedure would most likely achieve an intermediate level of performance for pure cultures.

7.4 In Vitro Experiments

In a second step, we relied on in vitro experiments: experiments involving spiking and growing microorganisms into blood culture bottles, to evaluate the performance of the algorithm.

The first experiment we considered aimed to evaluate the specificity of the algorithm, its ability to identify monomicrobial spectra, and therefore we spiked a single microorganism into blood culture bottles. The motivation followed directly that of the previous cross-validation experiment, with the striking difference that identification was made directly and in "semi-clinical" practice from microbial cells extracted from these blood cultures. More precisely, mass spectra were acquired according to the following protocol. Microbial suspensions were prepared by growing microorganisms on blood agar

(COS) plates, from which some colonies were then harvested and diluted in a suspension buffer to reach a concentration of 1000 CFU/ml. An inoculum of 0.4 ml was then injected into blood culture bottles of type BacT/ALERT® SA (bioMérieux, reference 259789) along with 10 ml of SPS anticoagulated human blood from healthy donors. The bottles were then placed into an incubator at 35 °C until positivity, determined by checking the colorimetric sensor change due to CO_2 production by the multiplying bacteria. The microbial concentration after positivity approximately ranged from 10^7 to 10^9 CFU/ml. Once the bottles were positive, two sample preparation protocols were carried out in parallel:

- A reference protocol, which consisted in streaking an inoculum of 0.05 ml on a COS agar plate. The plate was then placed in an incubator for 24 h, after which a (portion of a) colony was picked with a 1 µl loop.
- A direct-sample protocol involving a lysis-filtration method described earlier (Fothergill *et al.*, 2013).

In both cases, the microorganisms collected were spotted four times on a MALDI slide, and mass spectra were acquired with the VITEK®MS system.

Nineteen strains of 19 bacterial and fungal species were included in this experiment. The above protocols were carried out only once for most of them, and twice for five of them, which therefore provided us with 96 spectra[3] for each protocol. Results were shown in Figure 7.4, where we first confirmed (Figure 7.4A) that the algorithm was highly specific: 94.7% and 95.7% of the spectra obtained with the reference and the direct-sample protocols, respectively, were predicted to be monomicrobial. A correct identification rate of 85.1% was obtained with the direct-sample protocol, and was lower than that of 92.6% obtained with the reference protocol. As shown in Figure 7.4B, the majority of misidentifications were consistent across replicate mass spectra. In particular, the four spectra obtained from an *S. oralis* experiment were systematically predicted as representing *S. aureus* (either alone, or together with another *Staphylococcus* species) with both protocols. Interestingly, these spectra were also identified as *S. aureus* specific with the routine VITEK®MS system, which indicated that these misidentifications did not specifically arise from using our algorithm. A second experiment involving the same strain carried out another day led to a correct identification of all spectra, with both protocols. Regarding the reference protocol, only three other misidentifications were obtained, originating from three distinct strains. In all cases, the spectrum was predicted to be derived from a mixture of two species, the actual one representing the majority of the mixture, and the three remaining replicates were correctly identified. The lower rate of correct identification obtained with the direct-sample protocol could mainly be attributed to two strains that were almost systematically misidentified. In the first case, four spectra obtained from a *Staphylococcus capitis* extraction were predicted to be a mixture of *S. capitis* and *S. aureus*. In the second case, three out of four spectra obtained from a *S. pneumoniae* extraction were predicted as *S. oralis*. In both cases, these observations were consistent with the simulation study reported previously (Figure 7.3C).

The second experiment involved bimicrobial cultures prepared in vitro. However, instead of growing each pair of microorganisms in a common blood culture bottle, each

3 Fourteen strains with 4 replicate spectra + 5 strains with 2 × [4 replicate spectra].

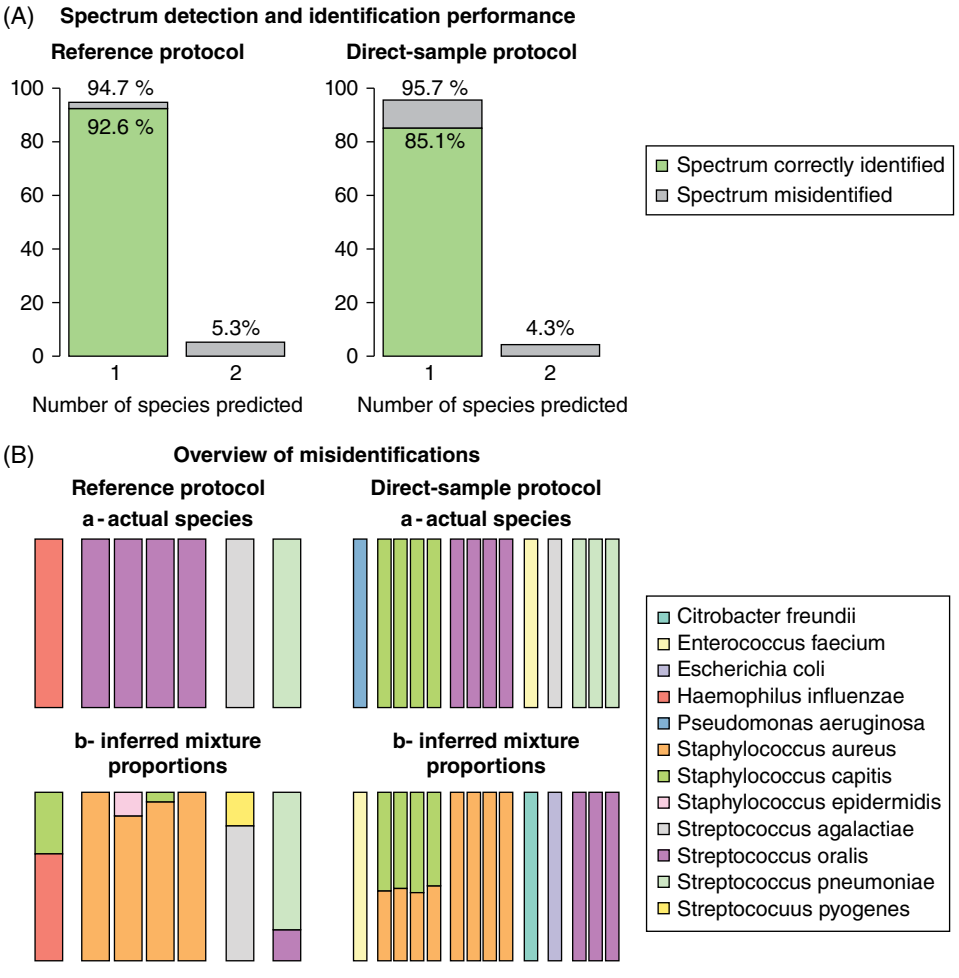


Figure 7.4 In vitro results – spiked monomicrobial blood cultures. (A) *Spectrum detection and identification performance.* Overall performance of the algorithm in detecting and identifying spiked monomicrobial blood cultured processed according to the reference (left) and direct-sample (right) protocols. Each bar represents the proportion of the spectral dataset for which a given number of species (here, 1 or 2) was predicted. The green and gray fractions of the bars represent the proportions of spectra correctly and misidentified, respectively. (B) *Overview of misidentifications.* Details about the misidentifications obtained using the reference protocol (left) and the direct-sample protocol (right). In both cases, the top row gives the actual relative concentration of the species in the mixture, and the bottom row gives the relative concentrations inferred by the algorithm, obtained by normalizing the vector γ of coefficients obtained by the algorithm, as described in the text.

microorganism was grown in its "own" bottle, and the mixtures were prepared afterward by mixing the resulting positive cultures. This procedure allowed one to precisely control the relative concentrations of the microorganisms within the blood cultures prior to the subsequent steps of sample preparation and spectral acquisition. This would have been harder to control if pairs of microorganisms were spiked in the same bottle, where they could grow at different rates or influence each other's growth

patterns. More precisely, mass spectra were acquired according to the following protocol. Microbial suspensions were prepared by growing microorganisms on Tryptic Soy Agar with 5% sheep blood for bacteria, or Sabouraud Dextrose Agar for yeast, overnight at 36 °C. The microorganisms were then harvested from isolated colonies and diluted in Tryptic Soy Broth to reach a concentration of 40–400 CFU/ml. An inoculum of 0.4 ml of these suspensions of individual microorganism strains were then injected into blood culture bottles of type BacT/ALERT® SA (bioMérieux, reference 259789) along with 10 ml of SPS anticoagulated human blood from healthy donors. The inoculated bottles were grown until flagged positive by the BacT/ALERT 3D system and then immediately placed at 2 °C–8 °C until processing to prevent further growth. Mixtures of the positive blood culture broths were prepared from the pure cultures at predetermined volumetric ratios in order to reach relative concentrations of 100%–0%, 90%–10%, 50%–50%, 10%–90%, and 0%–100%. For this purpose, the concentration of cells in a blood culture was quantified in terms of CFU/ml, which was measured by making a dilution series of broth from each positive bottle, growing them on the appropriate agar plates, and counting the resulting colonies. We note, however, that dry cell mass is a more important consideration than CFU/ml for MS analysis, and that the latter was chosen because of its simplicity. Microorganisms were then extracted by means of a lysis-centrifugation protocol. Blood cells were first lysed with a similar approach to that of Fothergill *et al.* (2013). The resulting lysate was then layered on top of a density cushion and centrifuged, which had the effect of letting the lysed blood and media above the cushion. The pelleted microorganisms were then resuspended with purified water. One microliter of these suspensions was applied to a MALDI-TOF target plate together with 1 µl of 50% formic acid, and, after drying at ambient temperature, 1 µl of matrix solution. Mass spectra were finally acquired with the VITEK®MS instrument.

Four pairs of microorganisms were considered in this experiment:

- Mixture A involved *Staphylococcus epidermidis* and *S. aureus*.
- Mixture B involved *Escherichia coli* and *S. aureus.*
- Mixture C involved *E. coli* and *P. aeruginosa.*
- Mixture D involved *Candida albicans* and *S. aureus.*

Results were shown in Figure 7.5, where we noted that the method correctly identified the majority of the spectra (17 out of 20), and in particular every 50%–50% mixture. Two partial identifications were obtained for unbalanced (10%–90%) mixtures, where the majority species was solely identified, and a misidentification involved predicting *Citrobacter freundii* instead of *E. coli*, in a case where it accounted for 10% of the mixture.

7.5 Discussion and Perspectives

We have developed a fully automatic in silico procedure to characterize a polymicrobial sample on the basis of a single mass spectrum. Our method builds upon the same reference database as the one used to identify pure cultures in the commercial VITEK®MS system, and hence could be easily translated for and used in routine clinical microbiology practice. Although this method was previously evaluated in a comprehensive way on a large and challenging spectral dataset obtained from in vitro mono- and bimicrobial

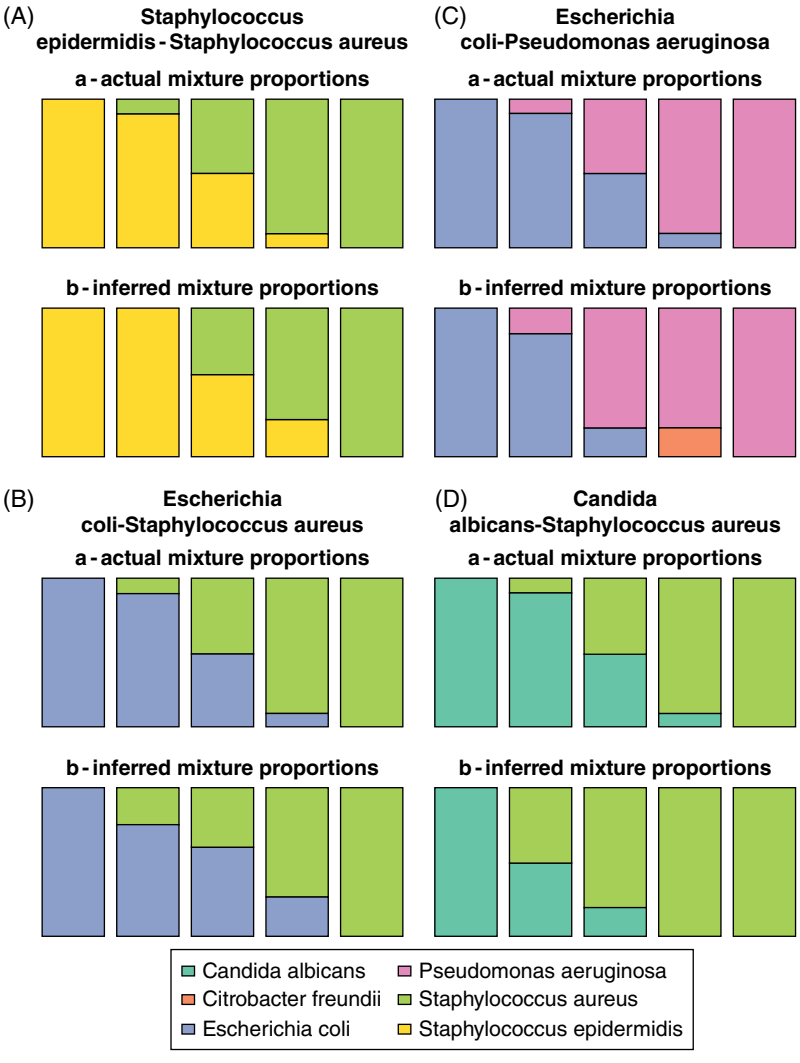


Figure 7.5 In vitro results – mixed positive blood cultures. From top left to bottom right: mixtures A, B, C, and D. In each case, the top row gives the actual relative concentration of the species in the mixture (100%–0%, 90%–10%, 50%–50%, 10%–90%, or 0%–100%), and the bottom row gives the relative concentrations inferred by the algorithm, obtained by normalizing the vector γ of coefficients obtained by the algorithm, as described in the text.

samples, we have further described here a proof of concept of its applicability for direct identification from positive blood cultures. Our experiments involved spectra generated in silico and in vitro, and although preliminary, we think these results are encouraging. First, we could confirm and validate the specificity of the algorithm: we have confidence in its ability to correctly detect and even identify a pure culture. Moreover, although the number of in vitro bimicrobial samples available was limited, most of them could be correctly identified. This was indeed the case for every balanced mixture and for the majority of the highly unbalanced ones (10%–90% proportions), among which we noted that

the major species was always correctly identified. Finally, the overall mixture performance estimated in silico was encouraging. Interestingly, we revealed a species-dependent level of performance that should be further understood, and hopefully improved when possible, in future work. As a concluding remark, we note that this algorithm can be useful to simply *detect* a mixed infection. Indeed, although not ultimately actionable in terms of diagnostics, being able to quickly do so, even if the sample is not properly identified to the bacterial species level yet, may already be useful to a clinical microbiologist to initiate additional cultivation of the microorganisms on a solid medium. In the setting considered in this study, we observed that the algorithm was almost as good in detecting a mixture and in identifying its composition. Indeed, the algorithm reached a rate of correct mixture identification of 80.8% (Figure 7.3A), while 83.5% of mixed spectra were recognized as such. Figure 7.5 shows such a case of correct detection but misidentification obtained from an in vitro sample, where a mixed *E. coli*/*P. aeruginosa* spectrum was identified as *C. freundii*/*P. aeruginosa*.

This study does not provide answers to at least three important questions. First, considering that a limited reference dataset was used in this study, proof of concept would most likely constitute a serious limitation from a routine clinical microbiology perspective. Whether the method would be compatible with a larger and more diverse microbial panel and a larger reference database remains an open question. Preliminary experiments suggested, however, that the method would scale up reasonably well up to a certain point. Three additional larger (nested) microbial panels were considered in a preliminary assessment. The first one consisted of 49 species and involved additional *Streptococcus* and *Staphylococcus*, as well as species of new genera such as *Salmonella enterica*, *Listeria monocytogenes*, and several *Aspergillus* species. The second one involved 138 species, and included every possible species of additional genera such as *Bacillus*, *Lactobacillus*, *Bacteroides*, *Propionibacterium*, *Campylobacter*, or *Prevotella*, to name a few. The largest one involved 269 species, and basically involved every species available for any genus represented in the previous lists. The experiments presented before were reproduced considering a reference dataset defined for each of these panels. In terms of specificity, a drop of up to 10% in terms of mixture detection and identification could be observed, based on the cross-validation procedure carried out from the reference datasets. In terms of mixture identification, performances estimated by simulation remained relatively steady when 26 to 138 species were considered. When the largest list of 269 species was considered, however, a serious drop in the rate of correct identification was observed (66.1% instead of 82.5% with the 26-species panel), that lead to a corresponding increase of the partial identification rate (22.6% instead of 13.6%). This analysis indicated that, in its current form, our method is not able to scale to very large reference panels with the same level of performance. Scaling the method to 100 species or so should remain feasible, but our approach would probably be hard to deploy on the whole VITEK®MS database, for instance. We emphasize, however, that these experiments did not take any prior information about the species involved, and hence some pairs of species considered to generate a mixture may be unrealistic with respect to a bloodstream infection application. This therefore means that the actual performance that may be reached in this setting may even more significantly differ from the above figures. We note moreover that the current VITEK®MS medical database cleared by the FDA comprises 193 species, somewhere in between the two most challenging settings (138 and 269 species) considered in this preliminary analysis. Regarding

the direct positive blood culture application, we would therefore recommend to design a reference panel involving up to around 100 species, which should allow one to at least detect, and hopefully identify, mixed cultures with a sufficient level of performance. The discriminative algorithm embedded in the routine identification system could be used to confirm the identification of a sample predicted to be monomicrobial by our algorithm, which should constitute the majority of positive blood cultures encountered in clinical routine.

Second, the limit of detection of the minority species of a mixture has yet to be demonstrated. Indeed, although we could empirically observe in our previous work (Mahé *et al.*, 2014) that the probability of correctly identifying a mixture decreased when the mixture was unbalanced (between 78% for balanced mixtures, down to 65% for 5:1 and 40% for 10:1 mixtures), we also know that our simulation framework is optimistic for unbalanced mixtures, and hence does not allow one to properly estimate this limit of detection. In the direct positive blood culture setting, we also note that the relative concentrations present in the positive bottle may be different from the ones originally present in the sample, because different species will grow at different rates. This limit-of-detection issue can therefore be approached at two levels: at the sample level and at the positive bottle level. Our future work will involve spiking several bacterial and/or fungal strains at various relative concentrations in several types of blood culture bottles to investigate this question.

A third question raised by this study is that of mixtures involving more than two species. Although we could not evaluate our method on current real in vitro spectra in this setting, a simulation study described in our previous work suggested that the approach could work reasonably well for balanced mixtures involving two to five different species. Although the proportion of mixtures correctly identified steadily decreased for an increasing number of species, mixed spectra were in general detected as such, and remained partially identified with only a single component missed in most cases. Regarding the direct positive blood culture application, however, it seems that such an event would be rare, according to recent studies involving clinical samples (Kok *et al.*, 2011; Lagacé-Wiens *et al.*, 2012; Chen *et al.*, 2013; Fothergill *et al.*, 2013).

In terms of perspectives, our future work will address these issues, which will require defining the most appropriate reference microbial panel to consider for a real-life scenario, spiking several bacterial and fungal species at various relative concentrations in several types of blood cultures, and subsequently validating this method using actual clinical samples.

References

Buchan, B. W. *et al.* (2012). Comparison of the MALDI Biotyper system using Sepsityper specimen processing to routine microbiological methods for identification of bacteria from positive blood culture bottles. *J. Clin. Microbiol.*, **50**(2), 346–352.

Chen J. H. K. *et al.* (2013). Direct bacterial identification in positive blood cultures by use of two commercial matrix-assisted laser desorption ionization-time of flight mass spectrometry systems. *J. Clin. Microbiol.*, **51**(6), 1733–1739.

Coombes, K. R. *et al.* (2007).Pre-processing mass-spectrometry data. In W. Dubitzky *et al.* (Eds.), *Fundamentals of Data Mining in Genomics and Proteomics*. Springer, US.

Croxatto, A. *et al.* (2014). Preparation of a blood culture pellet for rapid bacterial identification and antibiotic susceptibility testing. *J. Vis. Exp.*, **92**, e51985.

Efron, B. *et al.* (2004). Least angle regression, *Ann. Stat.*, **32**, 407–499.

Fothergill, A. *et al.* (2013). Rapid identification of bacteria and yeasts from positive-blood-culture bottles using a lysis-filtration method and matrix-assisted laser desorption ionization time of flight mass spectrum analysis with the SARAMIS database, *J. Clin. Microbiol.*, **51**(3), 805–809.

Kok, J. *et al.* (2011). Identification of bacteria in blood culture broths using matrix-assisted laser desorption ionization Sepsityper and time of flight mass spectrometry. *PLoS ONE*, **6**(8), e23285.

Lagacé-Wiens, P. R. S. *et al.* (2012). Identification of blood culture isolates directly from positive blood cultures by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry and a commercial extraction system: analysis of performance, cost, and turnaround time, *J. Clin. Microbiol.*, **50**(10), 3324–3328.

Lindner, M. S. and Renard, B. Y. (2013). Metagenomic abundance estimation and diagnostic testing on species level. *Nucleic Acids Res.*, **41**(1), e10.

Mahé, P. *et al.* (2014). Automatic identification of mixed bacterial species fingerprints in a MALDI-TOF mass-spectrum. *Bioinformatics*, **30**(9), 1280–1286.

Mestas, J. *et al.* (2014). Direct identification of bacteria from positive BacT/ALERT blood culture bottles using matrix-assisted laser desorption ionization-time-of-flight mass spectrometry, *Diagn. Microbiol. Infect. Dis.*, **80**(3), 193–196.

Schleif, F.-M. *et al.* (2011). Hierarchical deconvolution of linear mixtures of high-dimensional mass spectra in microbiology. In *Proceedings of the IASTED International Conference of Artificial Intelligence and Applications*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B*, **58**(1), 267–288.

Wahl, K. *et al.* (2002). Analysis of microbial mixtures by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.*, **74**, 6191–6199.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B*, **67**, 301–320.

8

Microbial DNA Analysis by MALDI-TOF Mass Spectrometry

Part 8A DNA Analysis of Viral Genomes using MALDI-TOF Mass Spectrometry

Christiane Honisch

Director Microbiology Markets, Illumina, Inc., San Diego, CA, USA

8A.1 Introduction

Matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS) has been applied to sequencing for over 25 years (Köster *et al.*, 1996; Tang *et al.*, 1995). Today the technology is being used in an increasing number of clinical laboratories around the world. A breakthrough has been the successful implementation of whole cell analysis of cultured microorganisms about a decade ago, which is now broadly accepted as an innovative tool for genus and species identification (van Belkum *et al.*, 2012). Two whole cell analysis systems, the bioMeriéux Vitek MS and the Bruker MALDI Biotyper as well as a nucleic acid analysis system, the Agena IMPACT Dx mass spectrometer (former Sequenom MassARRAY), have received FDA approval.

In contrast to clinical microbiology, where culture-based methods are still widely used to confirm the viability of the organism, especially in the context of antibiotic treatment regimes, clinical virology has shifted to molecular nucleic-acid-based tests owing to increased sensitivity, specificity, and faster turnaround times when compared to viral cultures (Buchan and Ledeboer, 2014). In general, the molecular detection and identification of viruses is based on the specific recovery and detection of certain genomic fragments inclusive for the viral species of interest and exclusive for additional species in the sample flora and nearest phylogenetic neighbors. Low viral loads can be detected quantitatively in complex samples after DNA or RNA purification owing to the efficient and specific amplification of a target region by polymerase chain reaction (PCR). Nucleic-acid-based MALDI-TOF MS has successfully been used for the molecular detection and identification of viral species, for the characterization of their genetic heterogeneity, and for tracking of transmissions (Ganova-Raeva and Khudyakov, 2013).

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

8A.2 The Molecular Detection and Identification of Viruses

Nucleic-acid-based MALDI-TOF MS applications for the detection and identification of viruses use specific or broad-range PCR primers in combination with different post-PCR biochemistries. The most widely used biochemistry and variations thereof, a PCR/primer extension assay (Figure 8A.1) (Storm *et al.*, 2002; Jurinke *et al.*, 2004) has been applied to the detection and genotyping of RNA and DNA viruses with a multiplexing level and throughput that support large-scale epidemiological research studies and might improve diagnosis of infections and co-infections in clinical applications.

The simultaneous detection of several viruses in a single sample has been demonstrated for the differentiation of human herpesviruses. Herpes simplex virus (HSV) types 1 and 2, varicella-zoster virus (VZV), Epstein-Barr virus (EBV) types A and B, cytomegalovirus (CMV), human herpesvirus 6 (HHV6) types A and B, and HHV7 and HHV8 (Kaposi's sarcoma-associated herpesvirus) infect humans and are often asymptomatic in healthy subjects. They can cause oral and genital lesions (HSV), chicken pox or shingles (VZV), infectious mononucleosis (EBV and CMV), or roseola (HHV6) and are associated with even more severe diseases in immunologically compromised hosts (Murray *et al.*, 1998), for example, transplant patients. Herpesvirus infections during pregnancy have been associated with birth defects or premature delivery

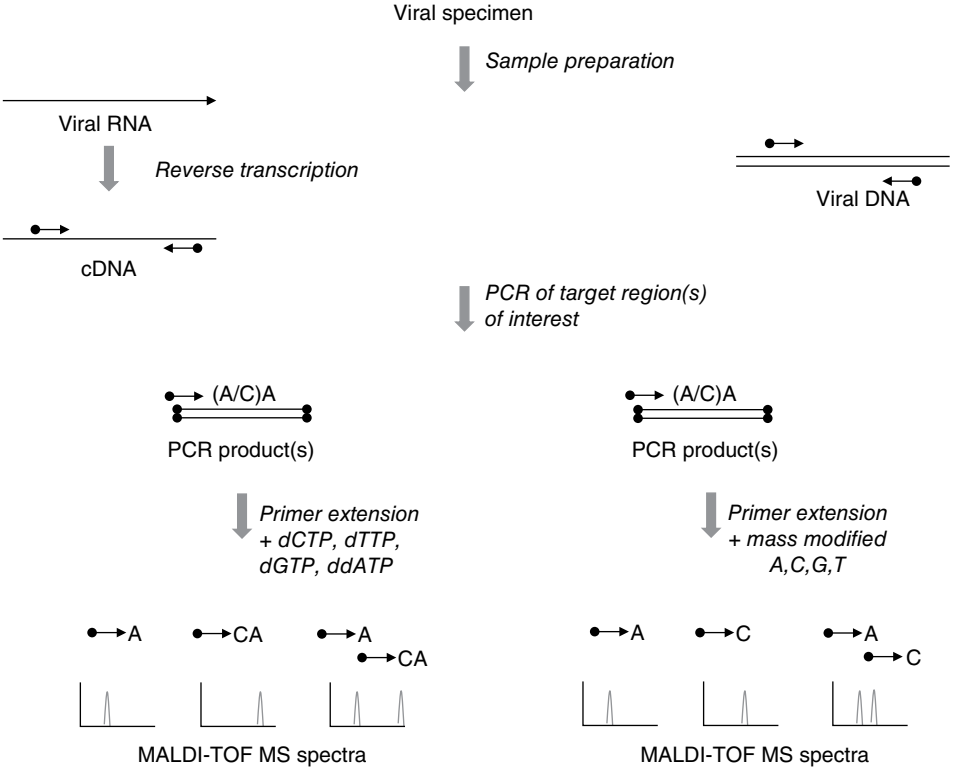


Figure 8A.1 Workflow of PCR/primer extension MALDI-TOF MS assays. PCR and primer extension with dNTP/ddNTP stop mixes (left); PCR and primer extension with mass-modified deoxyribonucleotides (right).

(Arvin, 1996). The detection limit of the developed MALDI-TOF MS assay at 100% sensitivity was found to be at five copies for EBV, HHV6A, and HHV6B and 10 copies for HHV7 as well as 100 copies for HSV-1, HSV-2, CMV, and HHV8. This qualifies the assay for multiplex HHV detection in large-scale research studies, for example, on archival samples and shows potential for validation of the test in patient samples (Sjöholm *et al.*, 2008).

The early detection and accurate identification of human enteric viruses, which are the most common cause of illnesses during early childhood, ranging from gastroenteritis to life-threatening diseases such as hand, foot, and mouth disease or neurological complications (Chan *et al.*, 2003; Tu *et al.*, 2007), is crucial to patient management and control of infection. A multiplex MALDI-TOF MS assay for the simultaneous detection of eight distinct enteric viruses, including poliovirus (PLV), coxsachievirus A16 (CoxA16), enterovirus 71 (EV71), hepatitis E virus (HEV), echovirus (ECHO), norovirus (NVG), astrovirus (ASTRV), and reovirus (REV), showed sensitivity levels ranging from 100 to 1000 copies/reaction and better agreement with results from direct sequencing than real-time RT-PCR (Piao *et al.*, 2012).

The examples demonstrate that the concordance rate between nucleic-acid-based MALDI-TOF MS and reference methods such as real-time PCR, dideoxy sequencing, and oligonucleotide microarrays for viral detection and identification is generally high, and the detection limits are comparable. MALDI-TOF MS detects an intrinsic physical property of the intact analyte – the molecular mass-to-charge ratio (m/z) of the extension product. This means no fluorescent or radio labeling is required; thus, multiplexing levels are not limited to the available number of chemical labels on a probe, for example, the limitation of four or five colors of real-time PCR.

When compared to a technology such as dideoxy sequencing, data acquisition times are much faster, requiring only a few seconds per sample and multiplex on a MALDI-TOF MS instrument. On commercially available systems, 24, 96, or 384 samples can be analyzed in one run in a maximum data acquisition time of 1 h. The accuracies and resolution of today's benchtop mass spectrometers allow for the detection of hundreds of nucleic-acid-specific signals per spectrum, which translates into multiplexing levels of 40 to 60 biallelic loci or single-nucleotide polymorphisms as well as insertions and deletions in one reaction and spectrum. A distinct advantage of the multiplexing capabilities of nucleic-acid-based MALDI-TOF MS is an easy extension of existing assays by adding new type-specific primers (Cobo, 2013).

The latest development of the PCR/primer extension-based assay utilizes a single mutation-specific chain terminator labeled with a moiety for solid phase capture. Captured, washed, and eluted products are interrogated for mass and mutational genotypes using MALDI-TOF MS. An ultrasensitive detection of mutations down to 0.125% of a mutant in the background of wild type has been acchieved as demonstrated for the detection of somatic cancer mutations (Mosko *et al.*, 2015) and would be applicable to viral detection.

8A.3 Viral Quantification

Viral vaccines contain live attenuated virus. Successful viral vaccines prevent, for example, yellow fever, measles, rubella, and mumps. The quality control to determine vaccine safety requires monitoring of small quantities of mutants or revertants that may indicate incomplete or unstable attenuation.

A PCR/primer extension assay and MALDI-TOF MS can be used to analyze frequencies of mutants by calculating the areas of the extension peaks of wild type and mutant. An application of this quantitative approach was demonstrated in a study supporting mumps vaccine control. Viral quasispecies of the mumps virus were determined between Jeryl Lynn substrains in live, attenuated mumps/measles vaccine on the basis of five distinct nucleotide positions in the viral genome. Feasibility was shown in reference with the existing QC methodology used by the Federal Drug Administration (FDA). Quantitative analysis of mutants by MALDI-TOF MS can thus be used to monitor the genetic stability of viruses during clinical trials of vaccines, for epidemiological surveillance of new virus isolates, and for screening of emerging drug-resistant viral strains in the course of antiviral therapies (Amexis *et al.*, 2001).

8A.4 The Characterization of Viral Genetic Heterogeneity

In addition to the detection and quantification of viruses, PCR/primer extension assays on MALDI-TOF MS enable the characterization of viruses by genotyping.

High-risk human papillomavirus infections (HPV), especially infections with HPV types 16 and 18, are associated with cervical cancer and in some cases with head/neck and schistosomiasis-associated bladder cancers. HPV detection and the differentiation of high and low risk types are an established and useful tag to screen for cervical cancer and monitor the efficacy of newly introduced HPV vaccines. A type-specific, MALDI-TOF-MS-based multiplex competitive PCR/primer extension assay of the viral E6 region was designed that detects and differentiates the 15 most important high-risk HPV subtypes (16, 18, 31, 33 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, and 73). The differentiation of all 15 types and the resultant sensitivity was shown to be superior to the standard four-color real-time fluorescent PCR-based assays and performed at least as well as the commercially available Hybrid Capture 2 High-Risk HPV DNA test (HC2; Digene Corp., Gaithersburg, MD) (Patel *et al.*, 2009). The sensitivity of the MALDI-TOF MS method extended down to individual molecules while specificity was maintained (Stenmark *et al.*, 2013; Yang *et al.*, 2004).

The technology was further successfully applied to the identification of hepatitis C (HCV) by genotype-specific point mutations in the 5′ untranslated region (5′ UTR) of the virus. Even with treatment regimes available, HCV is the leading cause of liver transplantation (Hugo and Rosen, 2011). Genotypes and subtypes of the virus vary by geographical region and in treatment outcome (Ilina *et al.*, 2005).

Hepatitis B virus (HBV) is another causative agent of liver disease and cancer. The genetic diversity of HBV resulted in the classification of the virus into 8 genotypes and 24 subtypes (Gish and Locarnini, 2007; Roque-Afonso *et al.*, 2003). The genotype of the virus has been associated with the severity and progression of chronic hepatitis B, the response to therapy, and the risk of liver cancer (Mahtab *et al.*, 2008; Palumbo, 2007).

PCR/primer extension assays have successfully been applied to genotype HBV and to detect up to 60 single-nucleotide drug resistance mutations in the HBV polymerase gene (Luan *et al.*, 2009; Malakhova *et al.*, 2009).

Comparative sequencing by MALDI-TOF MS uses PCR and four nucleotide-specific cleavage reactions of RNA molecules transcribed in vitro from the PCR products to a generated MS pattern that can be automatically compared to a simulated pattern based

on sequences in a reference database (Figure 8A.2) (Stanssens *et al.*, 2004). The results are returned with a scoring system and probability call (Honisch *et al.*, 2010). Target regions are identified, and sequence polymorphisms can be discovered (Böcker, 2003). The principle is in analogy with tryptic digestion of proteins and the subsequent peptide analysis by MS and database comparison. Sensitivity of two levels, exponential amplification by PCR combined with linear amplification in in vitro transcription, is achieved.

The application of comparative sequencing by MALDI-TOF MS on the *S* gene of HBV allowed for the automated and accurate identification of all eight virus-specific genotypes. As opposed to the targeted PCR/primer extension biochemistry, comparative sequencing by base-specific cleavage is amenable to the detection of novel HBV variants down to a single-nucleotide polymorphism (Honisch *et al.*, 2007). Mass peak patterns were automatically compared to sequences from known and database-deposited HBV genotypes. The assay is highly reproducible because of the built-in redundancy of the four cleavage reactions. Parameters such as viral titer, genotype, heterogeneity, quality of PCR, and MS patterns were carefully evaluated. The quality of the PCR product, and thus the quality of the PCR template, was the only parameter found to have a significant impact on the accuracy of the assay. Assay performance showed complete concordance with the gold-standard dideoxy sequencing results and phylogenetic analysis (Ganova-Raeva *et al.*, 2010). The application of the protocol achieved sensitivities of 64 genome copies of HBV/ml or 11 IU/ml. Mixtures of wild-type and mutant alleles can be identified if the minority type is present at >10% (Ehrich *et al.*, 2005). Due to its high throughput and semi-automation, the assay was commended

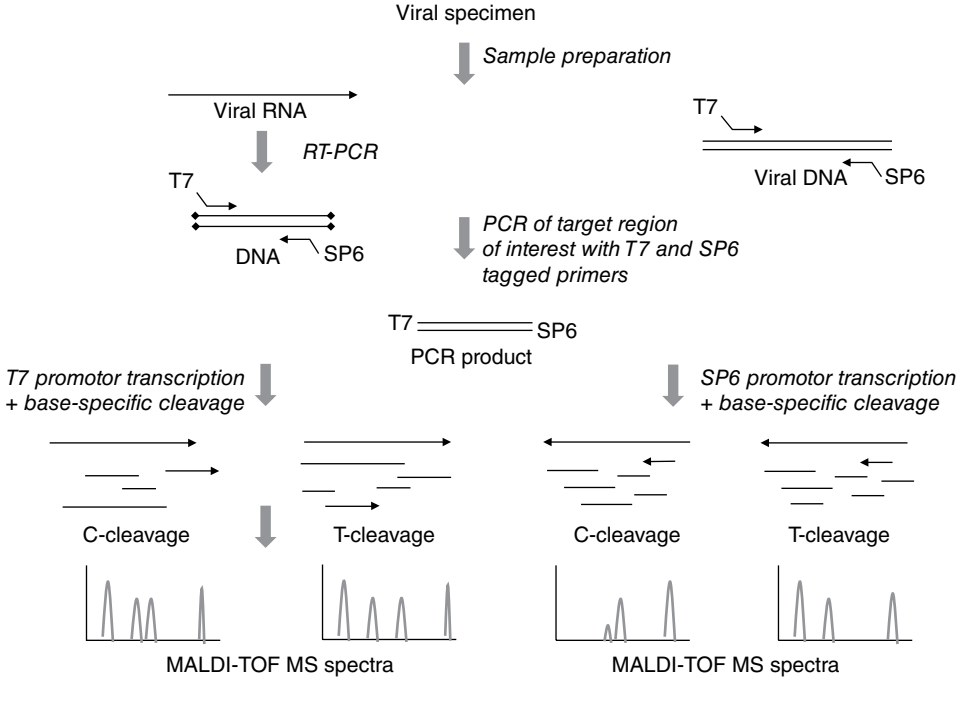


Figure 8A.2 Schematic workflow of PCR, T7-, and SP6- directed in vitro transcription and base-specific cleavage for comparative sequencing.

for molecular surveillance of HBV infections at the Center of Control and Prevention (CDC, Atlanta, GA) as a low-cost alternative to dideoxy sequencing (Ganova-Raeva *et al.*, 2012).

8A.5 Viral Transmission Monitoring

The global outbreak of severe acute respiratory syndrome (SARS) in February 2003 appears to have started in Guangdong Province, China, in November 2002 and spread explosively to Singapore, Hanoi and Toronto with exported cases on every continent. Caused by a coronavirus (SARS-CoV), SARS was the first disease to show the damage possible in a globalized world. Near the end of June 2003, the total number of cases was 8456 in 30 countries, of which 809 resulted in death (Fleck, 2003). The Genome Institute in Singapore developed a MALDI-TOF MS PCR/primer extension assay based on 21 single-nucleotide variations (SNVs) in the viral genome to track the transmission route of the virus in Singapore. MALDI-TOF MS analysis showed high sensitivity, providing a successful detection of the virus more than 95% of the time at viral concentrations of 75 copies per reaction, which was found to be close to the detection limit of 5–58 copies of viral RNA per reaction in RT-PCR-based diagnostic tests and within the concentration range reported in respiratory and plasma samples (Liu *et al.*, 2005). The study demonstrates the importance of genetic analysis and informative molecular markers for rapid and high-throughput epidemiological studies of pathogen transmission.

Hepatitis C (HCV) is an RNA virus with extremely variable genomic RNA. Its genetic heterogeneity is a hallmark of the virus, which exists in each infected individual as quasispecies, a heterogeneous and dynamic mixture of mutants (Argentini *et al.*, 2009). HCV is classified into 6 major genotypes and over 50 subtypes (Tellinghuisen *et al.*, 2007) with several distinct types differing by as much as 33% over the entire genome (Nolte *et al.*, 2003). Genotypes vary in disease outcome and response to therapy. Consensus sequencing of the hypervariable region HVR1 and regions in the NS5a and b gene are commonly used to identify HCV transmission. However, a consensus sequence cannot represent the entire HCV population present in the host, particularly in chronically infected patients.

The specific mass peak pattern of all four concatenated base-specific cleavage reactions of the comparative sequencing application on MALDI-TOF MS reflect the sequence context, heterogeneity, and diversity of HCV sample populations and can be mathematically presented as a numeric vector of the detected masses and their corresponding normalized peak intensities. Distances between these vectors correspond to the genetic relatedness of HCV strains among infected patients and allow for the accurate molecular detection of HCV transmission. The mass peak patterns are applicable to evaluate phylogenetic relationships between HCV populations. The approach has been found to match the accuracy of dideoxy sequencing, which had been implemented for this analysis after time-consuming limited dilution PCR or subcloning. Hundreds of isolated viral species and sequencing reactions had to be performed to obtain a snapshot of the population structure of a single sample in contrast to just one PCR and four base-specific cleavage reactions on MALDI-TOF MS.

An alternative applicable technology is next-generation sequencing (NGS), which is challenged by NGS sequencing errors per single DNA read, in a complex mixture of

high intra-host viral heterogeneity as well as extensive data processing required for NGS data analysis (Ganova-Raeva *et al.*, 2013).

8A.6 Additional Nucleic Acid Applications of MALDI-TOF MS

The separation of polymerase-chain-reaction-amplified and restriction-digested DNA fragments (PCR-RFLP) by MALDI-TOF MS as a replacement for gel electrophoresis was demonstrated for avian influenza viruses (Deyde *et al.*, 2011; Harder *et al.*, 2009; Hong *et al.*, 2008), HCV (Ilina *et al.*, 2005), drug-resistant HBV variants (Han *et al.*, 2011), and HPV genotyping (Hong *et al.*, 2008). The biochemistry introduces TypeIIS restriction endonuclease recognition sites (e.g., *Fok*I and *Bts*CI) by PCR surrounding a genotype-specific motif of interest. TypeIIS restriction enzymes cleave DNA subsequently at a fixed distance from their recognition site and make the assay independent of restriction sites within the target region of interest. Enzymatic digestion releases a pair of double-stranded fragments representative of the genotypic information. Both strands are analyzed by MALDI-TOF MS in parallel, providing a level of internal confirmation.

8A.7 Conclusion

A fully automated and integrated nucleic-acid-based MALDI-TOF MS platform has not yet been commercialized, even through the workflow from sample preparation through PCR amplification and cleanup is highly amenable to liquid handling automation.

Although the application of MS to nucleic-acid-based pathogen detection and identification is described in detail in the literature, the mere use of MS as a detection system in place of, for example, real-time PCR, gel electrophoresis, or sequencing does not resonate with clinical laboratories, mainly because of the seemingly high complexity of the technology and the cost of the instrumentation.

Multiplexing of up to 40–60 targets without a fluorescent-dye-based channel limitation and the possibility of quantifying the analyte can be beneficial when testing specimens from patients presenting with nonspecific symptoms attributable to a number of different pathogens. MS is the superior technology when it comes to the detection of various biomarker molecules including proteins, lipids, small molecules, carbohydrates, and nucleic acids on a single platform. MS as well as next-generation sequencing (NGS) have the potential to analyze specimens in a massively parallel fashion and measure molecular details that will enable precise diagnosis of infectious agents.

References

Amexis, G., Oeth, P., Abel, K., Ivshina, A., Pelloquin, F., Cantor, C. R., Braun, A., and Chumakov, K. (2001). Quantitative mutant analysis of viral quasispecies by chip-base matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci. USA* 98(21), 12097–12102.

Argentini, C., Genovese, D., Dettori, S., and Rapicetta, M. (2009). HCV genetic variability: From quasispecies evolution to genotype classification. *Future Microbiol.* 4, 359–373.

Arvin, A. M. (1996). Varicella-zoster virus. *Clin. Microbiol. Rev.* 9, 361–381.

Böcker, S. (2003). SNP and mutation discovery using base-specific cleavage and MALDI-TOF mass spectrometry. *Bioinformatics* 19, 44–53.

Buchan B. W. and Ledeboer, N. A. (2014). Emerging technologies for the clinical microbiology laboratory. *Clin. Microbiol. Rev.* 27(4), 783–822.

Chan, K. P., Goh, K. T., Chong, C. Y., Teo, E. S., Lau, G. and Ling, A. E. (2003). Epidemic hand, foot and mouth disease caused by human enterovirus 71, Singapore, *Emerg. Infec. Dis.* 9(1), 78–85.

Cobo, F. (2013). Application of MALDI-TOF mass spectrometry in clinical virology: A review. *Open Virol. J.* 7, 84–90.

Deyde, V. M., Sampath, R., and Gubareva, L. V. (2011). RT-PCR/electrospray ionization mass spectrometry approach in detection and characterization of influenza viruses. *Expert Rev. Mol. Diagn.* 11(1), 41–52.

Ehrich, M., Böcker, S., and van den Boom, D. (2005). Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS. *Nucleic Acids Res.* 33, e38.

Fleck, F. (2003). How SARS changed the world in less than 6 months. *Bull. World Health Organization* 81(8), 625–626.

Ganova-Raeva, L. M., Dimitrova, Z. E., Campo, D. S., and Khudyakov, Y. E. (2012). Application of mass spectrometry to molecular surveillance of hepatitis B and C viral infections. *Antivir. Ther.* 17(7 Pt.B), 1477–1482.

Ganova-Raeva, L. M., Dimitrova, Z. E., Campo, D. S., Lin, Y., Ramachandran, S., Xia, G., Honisch, C., Cantor, C. R., and Khudyakov, Y. E. (2013). Detection of hepatitis C virus transmission by use of DNA mass spectrometry. *JID* 207, 999–1006.

Ganova-Raeva, L. M., and Khudyakov, Y. E. (2013). Applications of mass spectrometry to molecular diagnostics of viral infections. *Expert Rev. Mol. Diagn.* 13(4), 377–388.

Ganova-Raeva, L. M., Ramachandran, S., Honisch, C., Forbi, J. C., Zhai, X., and Khudyakov, Y. (2010). Robust hepatitis B virus genotyping by mass spectrometry. *J. Clin. Microbiol.* 48(11), 4161–4168.

Gish, R., and Locarnini, S. (2007). Genotyping and genomic sequencing in clinical practice. *Clin. Liver. Dis.* 11, 761–795.

Han, K. H., Hong, S. P., Choi, S. H., Shin, S. K., Cho, S. W., Ahn, S. H., Hahn, J. S., and Kim S. O. (2011). Comparison of multiplex restriction fragment mass polymorphism and sequencing analysis for detecting entecavir resistance in chronic hepatitis B. *Antivir. Ther.* 16(1), 77–87.

Harder, M. K., Mettenleiter, T. C., and Karger, A. (2009). Diagnosis and strain differentiation of avian influenza viruses by restriction fragment mass analysis. *J. Virol. Methods* 158(1–2), 63–69.

Hong, S. P., Shin, S. K., Lee, E. H., Kim, E. O., Ji, S. I., Chung, H. J., Park, S. N., Yoo, W., Folk, W. R., and Kim, S. O. (2008). High-resolution human papillomavirus genotyping by MALDI-TOF mass spectrometry. *Nat. Protocols* 3(9), 1476–1484.

Honisch, C., Chen, Y. and Hillenkamp, F. (2010). Comparative DNA sequence analysis and typing using mass spectrometry. In Shah, H. N. and Gharbia, S. E. (Eds.), *Mass Spectrometry for Microbial Proteomics*. John Wiley & Sons, Chichester, pp. 441–462.

Honisch, C., Chen, Y., Mortimer, C., Arnold, C., Schmidt, O., van den Boom, D., Cantor, C. R., Sha, H. N., and Gharbia, S. E. (2007). Automated comparative sequence analysis by base-specific cleavage and mass spectrometry for nucleic-acid–based microbial typing. *Proc. Natl. Acad. Sci. USA* 104, 10649–10654.

Hugo, R. and Rosen, M. D. (2011). Chronic Hepatitis C infection. *N. Engl. J. Med.* 364, 2429–2438.

Ilina, E. N., Malakhova, M. V., Generozov, E. V., Nikolaev, E. N., and Govorun, V. M. (2005). Matrix-assisted laser desorption ionization-time of flight (mass spectrometry) for hepatitis C virus genotyping. *J. Clin. Microbiol.* 43(6), 2810–2815.

Köster, H., Tang, K., Fu, D. J., Braun, A., van den Boom, D., Smith, C. L., Cotter, R. J., and Cantor, C. R. (1996). A strategy for rapid and efficient DNA sequencing by mass spectrometry. *Nat. Biotechnol.* 14(9), 1123–1128.

Jurinke, C., Oeth, P., and van den Boom, D. (2004). MALDI-TOF mass spectrometry: A versatile tool for high-performance DNA analysis. *Mol. Biotechnol.* 26(2), 147–164.

Liu, J., Lim, S. L., Ruan, Y., Ling, A. E., Ng, L. F. P., Drosten, C., Liu, E. T., Stanton, L. W., and Hibbert, M. L. (2005). SARS transmission pattern in Singapore reassessed by viral sequence variation analysis. *PLoS Medicine* 2, e43.

Luan, J., Yuan, J., Li, X., Jin, S., Yu, L., Liao, M., Zhang, H., Xu, C., He, Q., Wen, B., Zhong, X., Chen, X., Chan, H. L., Sung, J. J., Zhou, B., and Ding, C. (2009). Multiplex detection of 60 hepatitis B virus variants by MALDI-TOF mas spectrometry. *Clin. Chem.* 55, 1503–1509.

Mahtab, M. A., Rahman, S., Khan, M., and Karim, F. (2008). Hepatitis B virus genotypes: An overview. *Hepatobiliary Pancreat. Dis. Int.* 7, 457–464.

Malakhova, M. V., Rahman, S., Kan, M., and Karim, F. (2009). Hepatitis B virus genetic typing using mass spectrometry. *Bull. Exp. Biol. Med.* 147, 220–225.

Mosko, M. J., Nakorchevsky, A. A., Flores, E., Metzler, H., Ehrich, M., van den Boom, D. J., Sherwood, J., and Nygren, A. O. H. (2015). Ultrasensitive detection of multiplexed somatic mutations using MALDI-TOF mass spectrometry. *J. Mol. Diagn.* doi: http://dx.doi.org/ 10.1016/j.jmoldx.2015.08.001.

Murray, P. R. Rosenthal, K. S., Kobayashi, G. S., and Pfaller, M. A. (1998). Human herpesvirus. In Brown, M. (Eds.), *Medical Microbiology*, 3rd ed. Mosby, St. Louis, MO.

Nolte, F. S. A., Green, M. K., Fiebelkorn, R., Caliendo, A. M., Sturchio, C., Grunwald, A., and Healy, M. (2003). Clinical evaluation of two methods for genotyping hepatitis C virus based on analysis of the 5' noncoding region. *J. Clin. Microbiol.* 41, 1558–1564.

Palumbo, E. (2007). Hepatitis B genotypes and response to antiviral therapy: A review. *Am. J. Ther.* 14, 306–309.

Patel, D. A., Shih, Y. J., Newton, D. W., Michael, C. W., Oeth, P. A., Kane, M. D., Opipari, A. W., Ruffin, M. T., Kalikin, L. M., and Kurnit, D. M. (2009). Development and evaluation of a PCR and mass spectrometry (PCR-MS) method for quantitative, type-specific detection of human papillomavirus. *J. Virol. Methods* 160(1), 78–84.

Piao, J., Jiang, J., Xu, B., Wang, X., Guan, Y., Wu, W., Liu, L., Zhang, Y., Huang, X., Wang, P., Zhao, J., Kang, X., Jiang, H., Cao, Y., Zheng, Y., Jiang, Y., Li, Y., Yang, Y., and Chen, W. (2012). Simultaneous detection and identification of eneric viruses by PCR Mass-Assay. *PLoS ONE* 7, 42251e.

Roque-Afonso, A. M., Ferey, M. P., Mackiewicz, V., Fki, L., and Dussaix, E. (2003). Monitoring the emergence of hepatitis B virus polymerase gene variants during lamivudine therapy in human immunodeficiency virus coinfected patients: Performance of CLIP sequencing and line probe assay. *Antivir. Ther.* 8, 627–634.

Sjöholm, M. I. L., Dillner, J., and Carlson, J. (2008). Multiplex detection of human herpesviruses from archival specimens by using matrix-assisted laser desorption ionization-time off light mass spectrometry. *J. Clin. Microbiol.* 46(2), 540–545.

Stanssens, P., Zabeau, M., Meerseman, G., Remes, B., Gansemans, Y., Storm, N., Hartmer, R., Honisch, C., Rodi, C. P., Böcker, S., and van den Boom, D. (2004). High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res.* 14(1), 126–133.

Stenmark, M. H., McHugh, J. B., Schipper, M., Walline, H. M., Komarck, C., Feng, E. Y., Worden, F. P., Wolf, G. T., Chepeha, D. B., Prince, M. E., Bradford, C. R., Mukherji, S. K., Eisbruch, A., and Carey, T. E. (2013). Nonendemic HPV-positive nasopharyngeal carcinoma: Association with poor prognosis. *Int. J. Radiation Oncol. Biol. Phys.* 88(3), 580–588.

Storm, N. B., Darnhofer-Patel, B., van den Boom, D. and Rodi, C. P. (2002). MALDI-TOF mass spectrometry-based SNP genotyping. *Methods Mol. Biol.* 212, 241–262.

Tang, K., Fu, D. J., Kotter, S., Cotter, R. J., Cantor, C. R., and Köster, H. (1995). Matrix-assisted laser desorption/ionization mass spectrometry of immobilized duplex DNA probes. *Nucleic Acid Res.* 23, 3126–3131.

Tellinghuisen, T. L., Evans, M. J., von Hahn, T., You, S., and Rice, C. M. (2007). Studying hepatitis C virus: Making the best of a bad virus. *J. Virol.* 81(17), 8853–8867.

Tu, P. V., Thao, N. T. T., Perea, D., Huu, T. K., Tien, N. T. K., Thuong T. C., How, O. M., Cardosa, M. J., and McMinn, P. C. (2007). Epidemiologic and virologic investigation of hand, foot, and mouth disease, southern Vietnam, 2005. *Emerg. Infec. Dis.* 13(11), 1733–1741.

van Belkum, A., Welker, M., Erhard, M., and Chatellier, S. (2012). Biomedical mass spectrometry in today's and tomorrow's clinical microbiology laboratories. *J. Clin. Microbiol.* 50(5), 1513–1517.

Yang, H., Yang, K., Khafagi, A., Tang, Y., Carey, T. E., Opipari, A. W., Lieberman, R., Oeth, P. A., Lancaster, W., Klinger, H. P., Kaseb, A. O., Metwally, A., Khaled, H., and Kurnit, D. M. (2004). Sensitive detection of human papillomavirus in cervical, head/neck, and schistosomiasis-associated bladder malignancies. *Proc. Natl. Acad. Sci. USA* 102(21), 7683–7688.

Part 8B Mass Spectral Analysis of Proteins of Nonculture and Cultured Viruses

Vlad Serafim,[1] Nicola Hennessy,[2] David J. Allen,[3] Christopher Ring,[1] Leonardo P. Munoz,[1] Saheer E. Gharbia,[2] Ajit J. Shah[1] and Haroun N. Shah[1]

[1] *Department of Natural Sciences, Middlesex University, London, UK*
[2] *Genomics Research Unit, Public Health England, London, UK*
[3] *Virology, Public Health England, London, UK*

8B.1 Introduction

The progress that has been made in mass spectrometry has revolutionized proteomics research. New strategies have been developed not only for analyzing specific proteins but also for global protein expression. Microbial proteomics has attracted considerable attention worldwide as a clinical tool for new methods of identification that are fast, robust, and relatively inexpensive. An obvious trend would be to adapt the existing technology for viral detection and identification. Furthermore, through in-depth proteomics research, molecular interactions between virus and host and also between virus and antiviral drugs can be revealed, providing useful data that can be utilized for prevention and treatment of viral diseases (Trauger *et al.*, 2013). The present part of this chapter will address the advances that have been made in this field and the challenges that have been encountered. Using a combination of mass spectrometry proteomic-based approaches, a significant number of viruses (from human and animal origin) have been investigated. From these, the origins of over 600 proteins were identified; of these, more than 100 were novel proteins (Table 8B.1).

The analysis of proteins in bacterial cell lysate or intact organism using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) provides a rapid means of bacterial identification which, in many cases, provides a species-level identification (Chapters 1–6, 11; Shah and Gharbia, 2010). The approach used for viruses needs to be different because of the restricted range of target proteins and the inability to culture viruses in pure cultures in an accessible manner. Viruses are found in complex matrices where other proteins (sometimes with the same molecular weight) are much more abundant than viral proteins. Even in the case of cultivable viruses, the viral proteins are "masked" by other proteins from the cell or tissue culture that is used. Hence, it is necessary to extract the protein(s) in order to obtain a relatively clean preparation.

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

Table 8B.1 A selection of viruses that have been investigated using proteomic-based approaches.

Virus/host	Number of proteins detected/identified	Ref.
Adenovirus type 5/human	11/2	Chelius *et al.* (2002)
Channel catfish virus/fish	16/ND	Davison *et al.* (1995)
Vaccinia virus/human	80/13	Resch *et al.* (2007)
	75/10	Chung *et al.* (2006)
	63/2	Yoder *et al.* (2006)
Cytomegalovirus/human	21/6	Baldick and Shenk (1996)
	71/12	Varnum *et al.* (2004)
Murine cytomegalovirus/mouse	58/20	Kattenhorn *et al.* (2004)
Epstein Barr virus/human	41/ND	Johannsen *et al.* (2004)
Kaposi's sarcoma associated herpes virus/human	24 (5)	Zhu *et al.* (2005)
Rhesus radiovirus/monkey	33 (3)	O'Connor *et al.* (2006)
White spot syndrome virus/shrimp	18/14	Huang *et al.* (2002)
	45/13	Li *et al.* (2007)
SARS coronavirus/human	4/4	Zeng *et al.* (2004)
HIV-1/human	14/ND	Saphire and Bark (2006)
	10 (ND)	Chertova *et al.* (2006)
Murine gammaherpesvirus 68 Mouse	17 (5)	Bortz *et al.* (2003)
Poliovirus type 1, 2 and 3/human	15/9	Calderaro *et al.* (2014)

For extraction of viral proteins, sample preparation is usually a complex process that involves several centrifugation steps beginning with a low-speed centrifugation in order to remove cell debris and other large fragments and moving to high-speed centrifugation. The extract can be purified further using gradient centrifugation (Calderaro *et al.*, 2014). Despite its efficiency, gradient ultracentrifugation is an expensive, lengthy, and a potential hazardous process that cannot be applied in a routine diagnostic workflow. Therefore, attention has focused on other purification methods intended to capture and purify viruses or viral proteins. These exploit antigen–antibody binding (Chou *et al.*, 2011) or size exclusion filtration (Colquhoun *et al.*, 2006).

Methods based on antigen–antibody binding involve the use of different surfaces coated with monoclonal antibodies. Magnetic nanoparticles are commonly used as a substrate for monoclonal antibodies because they have the advantage of a large surface-to-volume ratio and thus capture a large quantity of viral particles (Chou *et al.*, 2011). Due to their magnetic properties, the nanoparticles are easy to handle and can be used with a minimum of laboratory hardware (Tian *et al.*, 2008).

The use of antibodies to capture a virus is highly specific and has been shown to be very efficient (Chou *et al.*, 2011). However, most viruses are highly variable entities, acquiring mutations in a relatively short period of time (Goering *et al.*, 2013). Thus, this high diversity would necessitate production of new antibodies every time a new strain emerges.

Following the initial sample purification and enrichment, the process continues with disassembling of the viral particles and solubilizing the proteins. The proteins can then be analyzed in intact form (Trauger *et al.*, 2013). However, this only provides the molecular weight of the protein. To unequivocally identify the protein, it has to be digested and the resulting peptides analyzed using tandem mass spectrometry. This can be achieved either by in-solution digestion followed by mass spectrometric analysis or separation of proteins using SDS-PAGE and in-gel digestion of proteins(s) of interest followed by MS and MS/MS analysis. The most frequently used protein separation method is polyacrylamide gel electrophoresis, which can be used to separate proteins in one dimension (separation according to size) or in two dimensions (based on the isoelectric point in the first dimension and size in the second) (Shevchenko *et al.*, 1996). In-gel digestion of protein is often achieved using trypsin because it generates a significant number of peptides since it cleaves the protein chain at the carboxyl side of lysine – arginine, two amino acids that are found in great abundance (Trauger *et al.*, 2013).

8B.2 Norovirus Identification using MS

Norovirus infection represents the principal cause of gastroenteritis in both hospitals and the community. Although it is rarely life threatening, severe economic losses occur as a result of the large burden of the disease. It is estimated in the United Kingdom that nosocomial gastroenteritis costs the NHS up to 12.5% of the total £930 million that is spent every year on tackling all nosocomial infections (Lopman *et al.*, 2004). In the community, norovirus has an estimated rate of 47 cases per 1000 person/year. Thus, norovirus is responsible for three million disease cases every year and 130,000 GP consultations in the United Kingdom (Tam *et al.*, 2012).

Norovirus is a naked virus with icosahedral symmetry, and the diameter of the virion is 27–32 nm (Hardy, 2005). The genome consists of a 7.7 kb positive-sense, single-stranded RNA. It encodes three open reading frames (ORFs). ORF 2 encodes the major capsid protein VP1, whereas ORF 3 encodes the minor structural protein VP2, which does not play any role in the particle assembly but is essential in the production of infectious virus. ORF 1 encodes a polyprotein that plays a role in viral replication and does not have a structural role. VP1 is made up of 530–545 amino acids and has a molecular weight of 58–60 kDa (Zheng *et al.*, 2006). Each viral particle consists of 180 copies of VP1 that are arranged in an icosahedral virion (Hardy, 2005).

The VP1 protein includes two domains: P (protruding, with two subdomains P1 and P2) and S (shell). P2 is responsible for the majority of cellular and immune interactions, and it has the most sequence divergence. The capsid protein plays both a structural role and an attachment function. It determines the antigenicity of the viral strains (Zheng *et al.*, 2006). Norovirus cannot be cultivated in laboratory conditions. However, VP1 protein can be expressed in a recombinant system. The VP1 that is expressed self-assembles in virus-like particles (VLPs), which are structurally and antigenically indistinguishable from native virions but do not contain any RNA (Ausar *et al.*, 2006).

8B.3 Sample Preparation Considerations

The fact that the norovirus capsid consist of a single protein in a large number of copies should, at least in theory, make it easily detectable using MS. The main challenge is the complexity of feces in which the norovirus is found. Feces contain carbohydrates, protein, fat, and fiber, which originate from food, gut microbiota, and human cells. All this highlights the need for a very efficient purification method. The easiest way to achieve this would be to cultivate the virus in a cell culture. However, norovirus cannot be cultivated outside the human body. One desired approach would be to isolate the virus directly from feces. The isolate should have a purity that would allow the detection and identification of the VP1, equivalent to the identification of norovirus.

Because the use of monoclonal antibodies to capture norovirus is limited by the high antigenic variability between strains, a generic purification method that could be used to capture more than one norovirus strain was tested. The approach was based on exploiting the binding capacity of norovirus to human histo-blood group antigens (HBGA) (Huang *et al.*, 2005) as it has been shown that these antigens are present on the surface of the epithelial cells of the gastrointestinal tract (Tian *et al.*, 2008). Capture of norovirus using synthetic HBGA will allow its concentration and purification from complex matrix and enable rapid detection and identification using MS. Human HBGA is difficult to obtain; therefore, a porcine gastric mucin (PGM) was used as an alternative. PGM contains HBGA similar to those found in humans, and it has been used successfully for capturing and concentrating norovirus prior to PCR analysis by Tian and coworkers (2008), the result being a two-log increase in sensitivity.

8B.4 Experimental Workflow

In a study carried out in our laboratory, VLPs were used for optimizing the method and for spiking norovirus-negative samples. The VLPs were produced in a Baculovirus recombinant system using a GII-4 norovirus strain circulating after the 2002 epidemic, as described by Allen *et al.* (2009). A number of norovirus-positive feces samples were used.

The approach used was based on the PGM capture method using magnetic beads as substrate. The extracts were then analyzed in a high mass range using MALDI-TOF MS and its variant, surface-enhanced laser desorption/ionization (SELDI) for detecting the intact VP1 protein. The extracts were further processed using SDS-PAGE. The bands corresponding to the VP1 were excised and subjected to in-gel digestion with trypsin. The resulting peptides were analyzed using MALDI-TOF MS and high-performance liquid chromatography coupled to an ion-trap time-of-flight mass spectrometer (LC-IT-TOF MS) (Figure 8B.1).

8B.5 Detection of Intact VP1 using MALDI-TOF and SELDI-TOF MS

Using MALDI-TOF and SELDI-TOF MS, the intact VP1 was detectable at concentrations as low as 6.25 μg/ml in VLP suspensions in phosphate-buffered saline PBS. Using MALDI-TOF, the VP1 protein was detected as a monomer and in a single-charge state

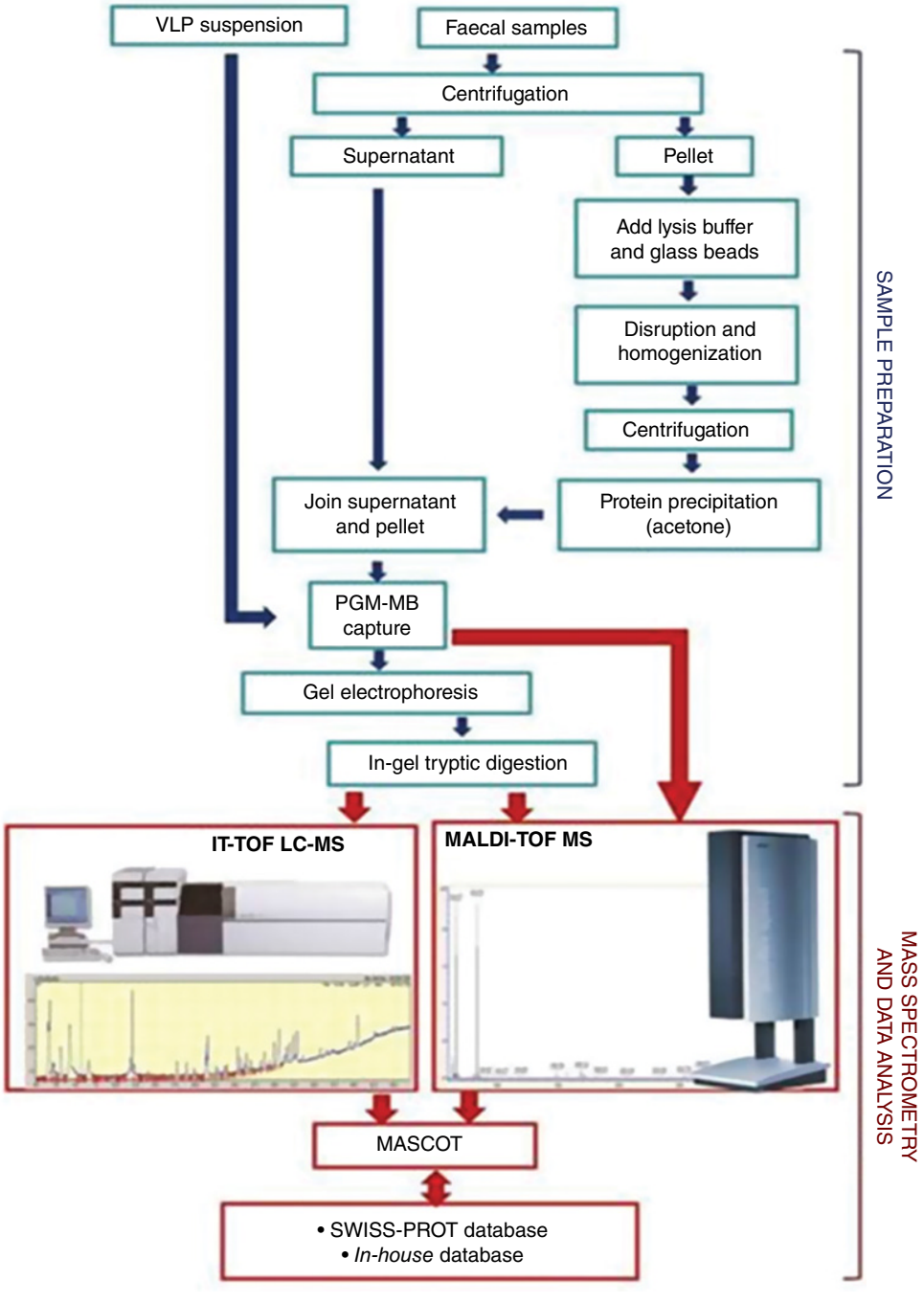


Figure 8B.1 Workflow for sample preparation and analysis of VP1 protein from norovirus.

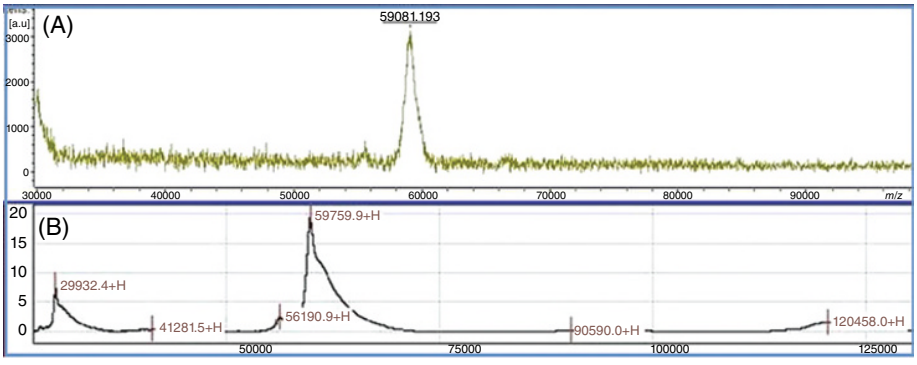


Figure 8B.2 Mass spectra showing VP1 detected using (A) MALDI-TOF MS and (B) SELDI-TOF MS. In the MALDI-TOF spectrum, only the single-charged monomer was detected. In the SELDI-TOF spectrum, besides the peak corresponding to the single-charged monomer (59,759 *m/z* Da), peaks corresponding to the double-charged monomer (29,932 *m/z* Da) and single-charged dimer (120,458 *m/z* Da) of VP1 were also detected.

at *m/z* of 59 kDa (Figure 8B.2). This corresponds to the calculated mass for the sequence that was used for producing the VLPs. The mass spectra of VP1 acquired using SELDI-TOF MS showed that in addition to the *m/z* of 59 kDa corresponding to the single-charged VP1 monomer, additional peaks at *m/z* of 30 kDa and 120 kDa were detected. These represented the double-charged monomer and dimer, respectively (Figure 8B.2).

However, intact VP1 could not be detected in fecal samples spiked with VLPs or norovirus-positive samples using these techniques. This may be attributed to contaminants in the extracts, which may affect the ionization and desorption processes.

8B.6 Peptide Mass Fingerprinting

As VP1 protein could not be detected in spiked fecal samples or norovirus-positive samples, the PGM-captured protein was further processed using polyacrylamide gel electrophoresis. The bands corresponding to VP1 protein (Figure 8B.3) were excised, and following reduction, alkylation incubated with trypsin. The resulting tryptic digests were analyzed using MALDI-TOF MS and LC-IT-TOF. The mass spectral data were analyzed against an in-house database that contains the sequence used for producing the VLPs and the all entries in the Mascot database.

Using MALDI-TOF MS for PMF, the VP1 protein was identified in clean VLP suspension. A coverage of up to 54% of the VP1 theoretic protein sequence was obtained. From PMF, the most intense ions were selected for MS/MS analysis, and the identity of five target peptides was confirmed (Table 8B.2). PMFs of VLP-spiked fecal specimens (even at concentrations of 25 µg VLPs/ml 10% fecal emulsion) and norovirus-positive samples did not produce any significant protein identifications.

In contrast, VP1 protein in fecal samples spiked with VLPs was detected using LC-IT-TOF. These samples contained approximately 100 mg of fecal material spiked with only 5 µg of VLPs, which represents 83 pmol of VP1 protein or 2.8×10^{11} VLPs. The initial PMF revealed a number of unmatched peptides, which suggests that other proteins are present

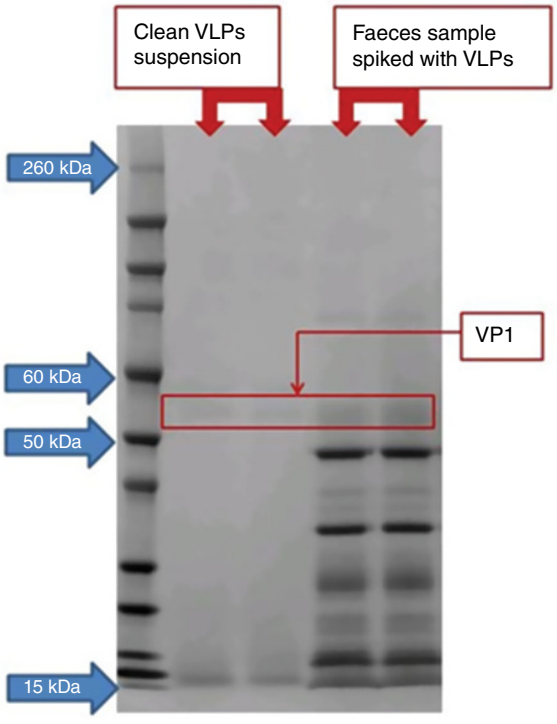


Figure 8B.3 Gel electrophoresis showing (left) separation of VP1 protein and proteins captured from the clean VLP suspension and (right) those in fecal samples spiked with VLPs.

in the sample in the 55–60 kDa range. To reduce the likelihood of false identification, a LC-MS/MS method was developed. From the initial PMF, three peptides were selected for MS/MS. These were selected on the basis of their intensity and the unicity of the corresponding peptide (for GII norovirus). The unicity of the peptides was evaluated by blasting them against an in-house, UniProt, and Swiss-Prot databases. The fragmentation of the three peptides provided information on their sequence and thus more confident identification of the VP1 protein. The tandem MS spectra of peptides with *m/z* of 905.5711, 488.865, and 1719.828 confirmed the identity of the peptides as FVNPDTGR, QLEPVLIPLPDVR, and NNFVQAPGGEFTVSPR, respectively (Table 8B.3).

8B.7 Conclusions

Developing an MS-based method for detecting norovirus in stool is a challenging task. The fact that the VP1 protein, which originates in native noroviruses, could not be identified may be due to the low concentration in feces, the complexity of the matrix, and the lack of a specific capturing method. These may be overcome in the near future by using more specific capture methods and higher-resolution mass spectrometers.

It was not feasible to detect intact VP1 in VLPs spiked into fecal samples and positive samples using MALDI-TOF MS. However, the presence of protein was confirmed

Table 8B.2 Observed and theoretical *m/z* of peptides detected using MALDI-TOF MS analysis of VLP tryptic digest together with error in ppm. The sequence of peptides highlighted in red was confirmed using tandem MS.

Observed m/z	Expected m/z	Mass accuracy (ppm)	Peptide sequence
843.4501	842.4902	−56.3	R.FPIPLEK.L
905.4137	904.4403	−37.4	**R.FVNPDTGR.V**
1052.521	1051.5411	−26.2	R.GDVTHIPGTR.T
1120.555	1119.5958	−43.4	R.IQGMLTQTTK.R
1206.656	1205.6842	−29.6	K.LIAMLYTPLR.A + Oxidation (M)
1281.584	1280.6249	−37.6	K.ATVSTGSADFTPK.L
1488.834	1487.8712	−30.1	R.QLEPVLIPLPDVR.N
1525.618	1524.6627	−34.3	R.ANNAGDDVFTVSCR.V
1572.732	1571.7693	−28.4	R.FTPVGVTQDGSTAHR.N
1688.743	1687.7954	−35.3	**R.NEPQQWVLPDYSGR.N**
1719.789	1718.8376	−32.3	**R.NNFVQAPGGEFTVSPR.N**
1804.866	1803.9268	−37.7	K.LFTGPSSAFVVQPQNGR.C
1844.744	1843.8125	−40.9	R.NNFYHYNQSNDSTIK.L
2062.992	2062.0769	−44.9	**R.TKPFTVPILTVEEMTNSR.F**
2078.992	2078.0718	−42.2	R.TKPFTVPILTVEEMTNSR.F + Oxidation (M)
2329.983	2329.0459	−30.1	**K.LGSIQFATDTDNDFETGQNTR.F**
2453.136	2452.2091	−32.7	R.CTTDGVLLGTTQLSPVNICTFR.G
2532.27	2531.3424	−31.4	R.VLTRPSPDFDFIFLVPPTVESR.T
2543.198	2542.2288	−14.9	R.FDSWVNQFYTLAPMGNGAGRRR.A
3246.403	3245.5088	−34.8	R.MNLASQNWNNYDPTEEIPAPLGTPDFVGR.I

Table 8B.3 Observed and theoretical *m/z* of peptides detected using LC-MS/MS analysis of VP1 originating from VLP-spiked feces together with error in ppm. The sequence of peptides highlighted in red was confirmed using tandem MS.

Observed m/z	Expected m/z	Mass accuracy (ppm)	Peptide
905.571	905.4476	136	**(R)FVNPDTGR(V)**
1488.865*	1487.858	−9.03	**R.QLEPVLIPLPDVR.N**
1525.66	1524.653	−6.56	R.ANNAGDDVFTVSCR.V
1572.756	1571.749	−13.1	R.FTPVGVTQDGSTAHR.N
1688.791	1687.784	−6.93	R.NEPQQWVLPDYSGR.N
1719.828*	1718.821	−9.85	**R.NNFVQAPGGEFTVSPR.N**
2330.036	2329.029	−7.36	K.LGSIQFATDTDNDFETGQNTR.F

using gel electrophoresis of captured protein and in-gel tryptic digestion of 55–60 kDa bands followed by LC-MS analysis, which lead to an increase in sample preparation time and require specialized equipment. However, without employing all these techniques together with LC-IT-TOF, the presence of VP1 could not be confirmed in fecal samples spiked with VLPs. MALDI-TOF MS was also used in this study because it offers a significant advantage in speed of analysis. However, its sensitivity and selectivity were not high enough to allow VP1 to be detected in VLP-spiked fecal samples because of the low levels of VP1 and impurities from the gel bands. The use of LC-MS, on the other hand, provides a means of separating the peptides, meaning that they will enter the mass spectrometer at different times, improving both the sensitivity and selectivity. In addition, with MALDI-TOF MS, due to lack of chromatographic separation, the more intense ions suppress the less intense ones. Also, using IT-TOF, pockets of ions can be accumulated in the ion trap and then separated in the TOF analyzer.

Mass spectral identification of VP1 following tryptic digestion requires a complete or partial sequence of the targeted virus. In the present study, the sequence used for producing the VLPs was also used for data analysis. This approach is somehow limited by the high genomic variability of norovirus. A proteomic-based detection method for norovirus (and any other highly viable virus) would require a permanently updated database that contains the sequence of every new strain. However, the three peptides chosen for MS/MS are found in the vast majority of the GII noroviruses and, following fragmentation, they are a good set of biomarkers to use for reliable identification. Mass spectral identification of other genotypes could be employed in a similar manner using other unique peptide sequences.

The present study represents an important step in developing assays for the detection of viral proteins from complex samples using various genotypes of norovirus to devise model for non-cultivable viruses. New forms of MS-based proteomics are developing together with new high-resolution instruments that may provide the tools for detection of viruses in complex matrices. The challenge for the future is to be able to find common proteins that are both genotype specific while retaining a common core of species-specific proteins.

8B.8 Bacteriophage Identification using MS

Bacteriophages are viruses that infect and replicates within bacteria. They are composed of proteins that encapsulate a DNA or RNA genome that may encode a few genes to as many as hundreds of genes. Bacteriophages have structural and functional similarities. A large proportion of bacteriophage genera possess a head and tail-like structure and infect bacteria by injecting their genome into its cytoplasm (Ackermann, 2009; McGrath and van Sinderen, 2007). The capsid is built by multiple copies of one or a few proteins, most often having an icosahedral structure. In some cases, they may have lipid envelope. The tail-like structure is responsible for attaching the bacterial wall and delivering the phage genome (Goering *et al.*, 2013).

The emergence of antibiotic-resistant bacteria represents a critical problem in today's medicine. Administration of bacteriophages for treating bacterial infections was considered even before the discovery of antibiotics. Although this approach was discarded in the United States and Western Europe, once antibiotics became widespread, phages continued to be used in Eastern Europe, where several studies were conducted (Bogovazova *et al.*, 1992; Rajagopala *et al.*, 2011; Sulakvelidze *et al.*, 2001). Considering the imperative need for finding an alternative to classic antibiotics, the results of early and more recent studies suggest that phages may be effective therapeutic agents and warrant further studies in the field (Sulakvelidze *et al.*, 2001).

Proteomic-based studies could address some of the problems of early therapeutic phage research. One of these is the insufficient purity of phage preparations. The cell lysates containing the phages also contained other contaminants such as endotoxins (Bogovazova *et al.*, 1992). Also, bacterial sterility has to be ensured (Carlton, 1999). MS combined with high-performance liquid chromatography can be used to evaluate the purity of these preparations. Also, the titer of phages could be determined indirectly by quantification of their structural proteins using quantitative MS. The use of MALDI-TOF MS and LC-MS are offering a rapid and efficient way to achieve this.

8B.9 Bacteriophages

A study was undertaken to evaluate the use of MS for rapid and accurate identification of bacteriophages. The phages were cultivated in *E. coli* cultures on agar plates. The plates exhibiting plaques (clear zones in the bacterial lawn, induced by phage) were selected. SM phage buffer was used for diluting the phages. An aliquot of the suspension was separated using SDS-PAGE. The protein bands were excised, and following reduction and alkylation treated with trypsin. Also, the suspension was precipitated with acetone, reconstituted in ammonium bicarbonate buffer, and subjected to reduction, alkylation, and tryptic digestion. The peptides obtained from both in-gel and solution digestion were analyzed using MALDI-TOF MS and LC-IT-TOF.

8B.10 Protein Identification

MALDI-TOF MS was used for peptide mass fingerprinting. LC-IT-TOF was used in MS/MS mode, and all ions detected were selected for tandem mass spectrometric analysis. The peptide mass fingerprint data obtained were compared to all entries from

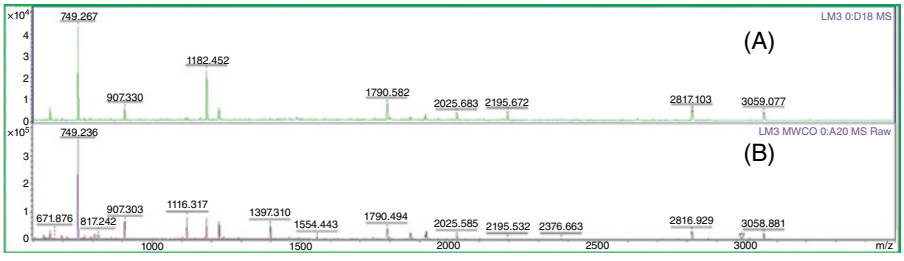


Figure 8B.4 PMF spectra of bacteriophage lambda major capsid protein digests showing the differences between (A) in-solution and (B) in-gel digestion. The spectra were recorded using MALDI-TOF MS in positive ion reflectron mode.

Table 8B.4 Proteins identified using LC-IT-TOF in MS/MS mode following in-solution tryptic digestion of bacteriophages protein lysates.

Protein	Species
Major capsid protein	Enterobacteria phage T7
Tail fiber protein	Enterobacteria phage T7
Capsid assembly scaffolding protein	Enterobacteria phage T7
Single-stranded DNA-binding protein	Enterobacteria phage T7
Portal protein	Enterobacteria phage T7
Tail tubular protein gp12	Enterobacteria phage T7
Protein D 11	Escherichia phage T5
L-alanyl-D-glutamate peptidase	Escherichia phage T5
Tail tube protein	Escherichia phage T5
Deoxyuridine 5'-triphosphate nucleotidohydrolase	Escherichia phage T5
Protein sciB	Escherichia phage T5
Probable portal protein	Escherichia phage T5
Major capsid protein	Escherichia phage T5
Probable tape measure protein	Escherichia phage T5
Probable baseplate hub protein	Escherichia phage T5
DNA polymerase	Escherichia phage T5
Major capsid protein	Enterobacteria phage lambda
DNA-packaging protein FI	Enterobacteria phage lambda
Capsid decoration protein	Enterobacteria phage lambda
Tail assembly protein GT	Enterobacteria phage lambda
Tail tube protein	Enterobacteria phage lambda
Serine/threonine-protein phosphatase	Enterobacteria phage lambda
Tail tube terminator protein	Enterobacteria phage lambda
Capsid assembly protease C	Enterobacteria phage lambda

the Swiss-Prot database. The MALDI-TOF MS PMF data identified one protein (major capsid protein) originating in enterobacteria phage lambda both from in-gel and in-solution digested samples (Figure 8B.4).

From IT-TOF data, various viral proteins were identified from both in-gel and in-solution digestion. Alongside Lambda, phages related to enterobacteria phages T5 and T7 were also identified (Table 8B.4).

8B.11 Conclusions

The number of matched peptides was significantly greater after the gel electrophoresis and in-gel digestion. However, the in-solution digestion followed by LC-IT-TOF analysis provided a confident identification of all three phage species. Furthermore, in-solution digestion is much easier to perform and takes less time to complete. In-solution digestion takes around 5 h to complete. One approach that can be used to reduce this is to use microwave-assisted digestion, which has been shown to reduce the incubation time to around 10 min (Lill, 2009). Although not assessed in this study, the process could also benefit from the use of magnetic beads (reducing the incubation time to as low as 30 s), which act by concentrating the protein on their surface and by absorbing radiation (Lill, 2009).

References

Ackermann, H-W. (2009). Phage classification and characterisation. In Clokie, M. R. J and Kropinski, A. M. (Eds.), *Bacteriophages: Methods and Protocols Volume 1. Isolation, Characterisation and Interactions*. Humana Press, USA, pp. 127–140.

Allen. J. A., Noad, R., Samuel, D., Gray, J. J., Roy, P., and Iturriza-Gomara, M. (2009). Characterization of GII-4 norovirus variant-specific surface-exposed site involved in antibody binding. *Virol J.* 6, 150.

Ausar, S. F., Foubert, T. R., Hudson, M. H., Vedvick, T. S., and Middaugh, C. R. (2006). Conformational stability and disassembly of Norwalk virus-like particles – effect of pH and temperature. *J. Biol. Chem.* 281, 19478–19488.

Bogovazova, G. G., Voroshilova, N. N., Bondarenko, V. M., Gorbatkova, G. A., Afanas'eva, E. V., Kazakova, T. B., Smirnov, V. D., Mamleeva, A. G., Glukharev, I. A., Erastova, E. I., Krylov, I. A., Ovcherenko, T. M., Baturo, A. P., Yalsyk, G. V., and Arefyeva, N. A. (1992). Immunobiological properties and therapeutic effectiveness of preparations from Klebsiella bacteriophages. *Zh Mikrobiol Epidemiol Immunobiol.* 3, 30–33.

Calderaro, A., Arcangeletti, M. C., Rodighiero, I., Buttrini, M., Gorrini, C., Motta, F., Germini, D., Medici, M. C., Chezzi, C., and De Conto, F. (2014). Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry applied to virus identification. *Scientific Reports*. [On-line] Available at: http://www.nature.com/articles/srep06803

Carlton, R. M. (1999). Phage therapy: past history and future prospects. *Arch Immunol Ther Exp.* 5, 267–274.

Chou, T. C., Hsu, W., Wang, C. H., Chen, Y. J., and Fang, J. M. (2011). Rapid and specific influenza virus detection by functionalized magnetic nanoparticles and mass spectrometry. *J. Nanobiotechnol.* Nov 16; 9, 52. doi: 10.1186/1477-3155-9-52.

Colquhoun, D. R., Schwab, K. J., Cole, R. N., and Halden, R. U. (2006). Detection of norovirus capsid protein in authentic standards and in stool extracts by matrix-assisted laser desorption ionization and nanospray mass spectrometry. *Appl. Erviron. Mirobiol.* 72, 2749–2755.

Goering, R., Dockrell, H., Zuckerman, M., Roitt, I., and Chiodini, P. L. (2013). *Mims' Medical Microbiology*, 5th edition. Elsevier.

Hardy, M. E. (2005). Norovirus protein structure and function. *FEMS Microbiol. Lett.* 253(1), 1–8.

Huang, P., Farkas, T., Zhong, W., Tan, M., Thornton, S., Morrow, A. L., and Jiang, X. (2005). Norovirus and Histo-blood group antigens: Demonstration of a wide spectrum of strain specificities and classification of two major binding groups among multiple binding patterns. *J. Virol.* 79(11), 6714–6722.

Lill, J. R. (2009). *Microwave Assisted Proteomics*. Royal Society of Chemistry, Cambridge.

Lopman, B. A., Reacher, M. H., Vipond, I. B., Hill, D., Perry, C., Halladay, T., Brown, D. W., Edmunds, W. J., and Sarangi, J. (2004). Epidemiology and cost of Nosocomial Gastroenteritis, Avon, England, 2002–2003. *Emerg. Infect. Dis.* 10, 1827–1834.

Mc Grath, S., and van Sinderen, D. (2007). *Bacteriophage: Genetics and Molecular Biology*. Caister Academic Press, Norfolk, UK.

Rajagopala, S. V., Casjens, S., and Uetz, P. (2011). The protein interaction map of bacteriophage lambda. *BMC Microbiol*.

Shah, H. N. and Gharbia, S. E. (2010). *Mass Spectrometry for Microbial Proteomics*. John Wiley & Sons, Chichester, UK.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins from silver-stained Polyacrylamide gels. *Anal. Chem.* 68(5), 850–858.

Sulakvelidze, A., Alavidze, Z., and Morris J. G. (2001). Bacteriophage therapy. *Antimicrob. Agents Chemother*.

Tam, C. C., Rodrigues, L. C., Viviani, L., Dodds, J. P., Evans, M. R., Hunter, P. R., Gray, J. J., Letley, L. H., Rait, G., Tompkins, D. S., and O'Brien, S. J. (2012). Longitudinal study of infectious intestinal disease in the UK (IID2 study): Incidence in the community and presenting to general practice. *Gut.* 61, 69–77.

Tian, P., Engelbrektson, A., and Mandrell, R. (2008). Two-log increase in sensitivity for detection of norovirus in complex samples by concentration with porcine gastric mucin conjugated to magnetic beads. *Appl. Erviron. Mirobiol.* 74, 4271–4276.

Trauger, S. A., Junker, T., and Siuzdak, G. (2013). Investigating viral proteins and intact viruses with mass spectrometry. *Open Virol. J.* 7, 84–90.

Zheng, D., Ando, T., Fankhauser, R. L., Beard, R. S., Glass, R. I., and Monroe, S. S. (2006). Norovirus classification and proposed strain nomenclature. *Virology.* 346: 312–323.

9

Impact of MALDI-TOF MS in Clinical Mycology; Progress and Barriers in Diagnostics

Cledir R. Santos,[1] Elaine Francisco,[2] Mariana Mazza,[3] Ana Carolina B. Padovan,[2] Arnaldo Colombo[2] and Nelson Lima[4]

[1] *Department of Chemical Sciences and Natural Resources, CIBAMA, BIOREN, University of La Frontera, Temuco, Chile*
[2] *Department of Medicine, Infectious Diseases Section, Federal University of São Paulo (UNIFESP), São Paulo, SP, Brazil*
[3] *Department of Mycology, INEI/Instituto Nacional de Enfermedades Infecciosas "Dr. Carlos G. Malbrán"- ANLIS. Ciudad Autónoma de Buenos Aires, Argentina*
[4] *CEB-Centre of Biological Engineering, Micoteca da Universidade do Minho, University of Minho, Braga, Portugal*

9.1 Introduction

The science of microbial identification can be considered to have arisen from the seminal work of the Dutch microbiologist Antonie van Leeuwenhoek. He provided the pioneering microscopic work in 1683 using rudimentary equipment to support the studies of the equally important French microbiologist Louis Pasteur, who described the biological functions of microbes. In Germany, Robert Koch demonstrated the aetiology of the infectious disease process in humans. The search for the accurate identification of microbial pathogens became crucial for underpinning medicine in treating diseases.

Florentine Pier Antonio Micheli published the *Nova Plantarum Genera* in 1729, which is an illustrated work detailing approximately 1900 'plant' species. Of these, about 900 were fungi and lichenized fungi. In fact, Micheli specialized in microfungi and is notable for having defined several important genera including *Aspergillus*. Even now, classical mycological identifications are performed by some of those classical methods based on micro- and macromorphology (Simões *et al.*, 2013).

In any discussion of the history of fungal identifications, it is fundamental to divide them into two groups: the single-celled yeasts and the filamentous fungi, as each taxonomy has developed differently. Yeasts have a limited morphology upon which it is impossible to circumscribe taxa sufficiently. However, the biochemical and physiological diversity is sufficiently large to allow tests such as growth on various carbon sources to be used, in a manner similar to those for bacteria. Filamentous fungi have a plethora of morphological characteristics such as shapes of hyphae, spores and, particularly, conidiophores, which allowed a greater range of taxonomic characters [e.g. Rapper and Fennell (1965) for *Aspergillus*]. In addition, many fungi show distinctive macromorphologies that vary with different agars, temperatures and water content which can be

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

used in classifications [e.g. Pitt (1973) for *Penicillium*]. Consequently, a sophisticated taxonomy was constructed for these organisms based on morphology. However, it was becoming apparent that morphologies for filamentous fungi had limitations in that not all taxa could be reliably identified, and so biochemical and physiological methods were increasingly introduced. The biochemical procedures were often based on chromatographic techniques, culminating in a wide range of methods potentially being employed (Frisvad *et al.*, 1989; Paterson and Bridge, 1994; Paterson *et al.*, 2006; Simões *et al.*, 2013). Indeed, the use of secondary metabolites in some groups has been particularly emphasized by Frisvad (e.g. Frisvad and Filtenborg, 1983; Frisvad and Thrane, 1987) and are still used in metabolomics (Alexander *et al.*, 2015).

Increasingly, it was realized that a single method was unavailable that could be used for each taxon and that a multidisciplinary, or polyphasic approach, using many different characters from various disciplines was optimal, which was first reported by Bridge *et al.* (1989a) for the terverticillate penicillia: it is now generally accepted as the optimal approach to cover a wide range of taxa. Careful consideration is required regarding which individual methods are to be used and on what to base the core taxonomy. To a large extent, these biochemical and physiological methods have themselves been complemented or superseded by molecular biology methods for filamentous fungi and yeasts in the clinical field (Liu, 2011; Paterson and Lima, 2015), with DNA barcoding finding prominence as a single-method procedure for fungal identifications (Robert *et al.*, 2015).

Polyphasic identification aims at the integration of different taxonomic characters (Simões *et al.*, 2013). Interestingly, Bridge *et al.* (1989b) were also the first to employ numerical taxonomy as an integration method for filamentous fungi, often used previously for other microorganisms. However, there remains the problem of variation, which can originate from the fungal specimen, and/or the analytical technique employed (Paterson, 1998). By using numerous techniques, it is assumed the level of variation in the technique can be reduced, although variation in the fungal specimen remains. The time and resources required to undertake a particular identification also have to be considered. What can be done in a fully equipped mycological centre is different from what can be done in a clinical diagnostic laboratory with limited resources or in an industrial food laboratory, where the practical approach based on antifungal resistance or mycotoxin characters, respectively, can provide answers to specific problems. These factors intensify the need for the scientific community to find new and complementary fungal identification tools.

In the ground-breaking paper by Cain *et al.* (1994), a new methodology for the identification of bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was presented, where sample preparation involved minimal purification of cells. Holland *et al.* (1996) described for the first time an improved method for the rapid identification of whole bacterial cells by MALDI-TOF MS, establishing the basis of the current methodology. This inspired the use of MALDI-TOF MS in fungal identifications (Kallow *et al.*, 2006). MALDI-TOF MS has now been applied routinely to analyze the chemical cellular composition of microorganisms, providing rapid and discriminatory proteomic profiles for identification and subtyping (Amiri-Eliasi and Fenselau, 2001; Dias *et al.*, 2011; Erhard *et al.*, 2008; Kallow *et al.*, 2006; Maier *et al.*, 2006; Nicolau *et al.*, 2014; Rodrigues *et al.*, 2011; Rodrigues *et al.*, 2014; Oliveira *et al.*, 2015; Passarini *et al.*, 2013; Pereira *et al.*, 2014; Santos *et al.*, 2010, 2011; Seyfarth *et al.*, 2008; Silva *et al.*, 2015; Stackebrandt *et al.*, 2005).

The application of this technique for the identification of clinical fungal samples is currently well established based on the remarkable reproducibility for the measurement of constantly expressed and highly abundant proteins, such as ribosomal proteins, that are used as biomarkers to generate a fingerprint profile that ranges between 2 and 20 kDa (Dias *et al.*, 2011; Lima-Neto *et al.*, 2014; Pereira *et al.*, 2014; Santos *et al.*, 2010; Santos *et al.*, 2011). For example, proteomic profiles by MALDI-TOF MS have been used in the identification of filamentous fungi isolated in complex matrices and from different clinical substrates (Becker *et al.*, 2014; Silva *et al.*, 2015).

MALDI-TOF MS has been used as (1) an important component in the polyphasic identification of clinical fungi and (2) a rapid diagnostic tool for routine clinical fungal identifications (Croxatto *et al.*, 2012; Schulthess *et al.*, 2014). However, there are some important limitations to MALDI-TOF MS, particularly in relation to the identification of closely related fungal taxa, such as the dimorphic fungi with mycelial-to-yeast phase transitions or highly encapsulated yeasts. Commercial databases of MALDI-TOF MS spectra have also limitations in their coverage of taxa, although this has expanded recently (Lohmann *et al.*, 2013; Mancini *et al.*, 2013).

Commercial databases are based on the different protocols used by the main MALDI-TOF MS manufacturers and are widely available. In some cases, sample preparation protocols have been changed over the time even for the same database, as MALDI-TOF MS evolved and sample preparation to produce databases may have been developed in a generalized manner without using clinical samples. For example, sample optimization may have occurred when comparing small differences in spectra from the same taxon. This evolution of sample preparation was not always accomplished with a data cleaning process or data storage conversion. Commercial databases for clinical fungi are less well established and comprehensive than for bacteria.

For fungal, and other, identifications by MALDI-TOF MS, commercial databases are built with software that uses a point system based on the peak list with mass signals weighted according to their specificity. Similarity between individual spectra is expressed as the relative or absolute number of mass signals matching after subjecting the data to a single link agglomerative clustering algorithm analysis. In general, the protocols used in different laboratories are not standardized. A large number of publications contain different (1) protocols for protein extraction, (2) matrices for MALDI-TOF MS analysis and (3) growth conditions for fungi are available in the scientific literature. These can affect negatively the spectra quality and consequently fungal identification with potentially serious shortfalls in clinical diagnoses. The sources of variation are protein extraction, matrices, growth conditions and databases without even considering the natural variation of fungal strains.

9.2 Evolution in Commercial Methodologies of Sample Preparation

9.2.1 Fungal Identification

The identification of clinically relevant fungi in the mycological laboratory has been attempted by MALDI-TOF MS. Rapid identification of some filamentous fungi and yeasts are possible without time-consuming subculturing and phenotypic identification

(Bader, 2013; Schulthess *et al.*, 2014). However, some fungal taxa remain problematic due to their wall cell rigidity, heterogeneity of biological materials and poor protein extraction generating low numbers of mass peaks in the spectra (Oliveira *et al.*, 2015). However, MALDI-TOF MS in routine diagnostic clinical mycology is rare, especially for filamentous fungi, mainly due to (1) extended sample preparation for filamentous fungi for good-quality spectra, (2) low numbers of representative taxa in databases and (3) the quality of databases (Schulthess *et al.*, 2014). Compared to bacterial sample preparation, fungal sample preparation is laborious and time consuming.

Different studies have reported the establishment of in-house database sets within commercially available databases covering mainly a single fungal genus, such as *Aspergillus*, *Candida*, *Fusarium* and *Lichtheimia*. In other cases, the (1) *Pseudallescheria*/*Scedosporium* complex, (2) Mucorales, (3) the dermatophytes and (4) clinically relevant yeasts are covered (De Carolis *et al.*, 2012; De Carolis *et al.*, 2014a; Lima-Neto *et al.*, 2014; Santos *et al.*, 2011; Schulthess *et al.*, 2014). Despite the proliferation of in-house and commercial databases, two main commercial platforms are currently available for use in clinical laboratories: MALDI Biotyper® (Bruker Daltonics®, Germany) and VITEK® MS (bioMérieux, France). Furthermore, the smaller platform of MS LT2-ANDROMAS (http://andromas.com/a-propos/, France) is also available and is discussed in the following.

9.2.2 MALDI Biotyper

Bruker Daltonics published an application note for the Biotyper MALDI-TOF MS as a fast and reliable method for the classification and identification of bacteria, yeasts and filamentous fungi using spectral fingerprinting (Maier *et al.*, 2006). This has applications in clinical diagnostics, environmental and taxonomical research and food-processing quality control using a straightforward and generic sample preparation protocol. According to the application note, the system allowed identifications of yeasts and fungi in few minutes. Sample preparation was described as very simple through the direct transfer of a thin layer of the microbial cells (without any treatment) directly from the culture plate to the MALDI-TOF sample plate, followed by the addition of a α-cyano-4-hydroxycinnamic acid (CHCA) matrix solution and spectral acquisition (Maier *et al.*, 2006). This method was considered optimal and fit for purpose for all microorganisms.

The diagnostic procedure employed by Biotyper is implemented by the MALDI Biotyper database, which generates cellular biocompound profiles which leads to microbial identification using well-developed protocols of sample preparation. Fresh fungal material should be used for analysis, whereas overnight growth is used for yeasts. In contrast, slow-growing fungi may need to be cultivated for several days before spectral analysis, which makes obtaining fresh material problematic. Fungi stored for more than 3 d, even at 4°C or lower, should be avoided because it has a negative impact on spectral quality and reproducibility. In addition, according to this manufacturer, culture plates with different recommended media and growth temperatures could also have a negative effect on final identification results.

Filamentous fungi are grown in liquid, using a standardized method. Briefly, tubes containing media are inoculated with the filamentous fungi, placed in Bruker's cultivation tubes, transferred to a shaker and incubated overnight or until enough biological material is observed. The samples are left for 10 min until the filamentous fungus sediments on the bottom of the tube. A sample (1.5 ml) is taken from the sediment and centrifuged for 2 min at full speed (e.g. 10,000g). The supernatant is carefully removed,

and 1 ml distilled water is added to the pellet. The sample is vortexed for 1 min with washing, and re-vortexed. An extraction with ethanol is performed prior to sample transfer to the MALDI-TOF sample plate, to which CHCA matrix solution in 50% acetonitrile and 1.5% trifluoroacetic acid is added. The sample is air-dried at room temperature and analyzed.

Analysis for yeast identification is mainly performed based on a standard microtube extraction protocol. A single colony is thoroughly suspended in 300 µl of distilled water, and then 900 µl of absolute ethanol is added. The cell suspension is vortexed and centrifuged at 10,000g for 2 min. The supernatant is removed, 50 µl of formic acid is added to the mixture after which it is re-vortexed. Acetonitrile (50 µl) is added and vortexed vigorously, and the tube is centrifuged at 10,000g for 2 min. One microlitre of the supernatant is spotted onto the MALDI-TOF sample plate and allowed to dry before overlying with CHCA matrix as described above for filamentous fungi. MALDI-TOF sample plates are applied to the MS, and the results are analyzed in automatic and linear mode as per instructions of the manufacturer.

Recently, a direct blood culture test using culture-independent methods for yeasts identification was developed (Morgenthaler and Kostrzewa, 2015). In this procedure, 1 ml of liquid blood culture is harvested and transferred to a microtube. After addition of 200 µl of lysis buffer (Bruker Daltonics), the sample is stirred for 10 s and then centrifuged for 1 min at 10,000g at room temperature. The supernatant is discarded, and pellet suspended in 300 µl of distilled water. The sample is washed twice with 900 µl ethanol and subjected to centrifugation cycles of 2 min at 10,000g. The pellet is suspended in 30 µl of 70% formic acid. Acetonitrile (30 µl) is added, and the sample is vortex and centrifuged. One microlitre of the extract is spotted onto a MALDI-TOF sample plate and analyzed according to the MALDI Biotyper standard procedure, as described above for the analysis of yeasts on pure culture.

The identification generated is associated with a numerical score. There are thresholds where values greater or equal to 2.0 represent proven species or particular groups of strains. Values from 1.99 to 1.7 indicate genus identification, whereas samples with lower than 1.7 need to be reanalyzed.

These databases are controlled by the company, and customers cannot introduce spectra obtained by different methodologies. However, it is possible to modify protocols and add information to in-house libraries in a private field of the MALDI Biotyper system. In this case, all generated spectra will only be available to each end user. The user is entirely responsible for the validation and use of results derived from library entries they generate in their in-house libraries (MALDI Biotyper, 2014).

In a study by performed by the authors, MALDI-TOF spectra of *Aspergillus udagawae* (Figure 9.1A), an emergent human pathogen associated with invasive aspergillosis, and *Trichosporon asahii* (Figure 9.1C), an emerging fungal pathogen seen particularly in immunologically compromised patients, were obtained by the MALDI Biotyper using the manufacturer's methodology. An analysis of the same fungal isolates and conditions, but using two sample preparation procedures were evaluated (Figures 9.1B and D). The spectra are different for each fungus, and different protocols gave different spectra. The alternative sample preparation was based on the NIH-National Institutes of Health, USA protocol (Lau *et al.*, 2013), which uses absolute ethanol and 0.1-mm-diameter zirconia-silica beads to disrupt the fungal cell wall before MALDI-TOF MS analysis.

The modification of sample preparation protocols may have changed the fungal profiles and misidentifications were obtained by the alternative method, although

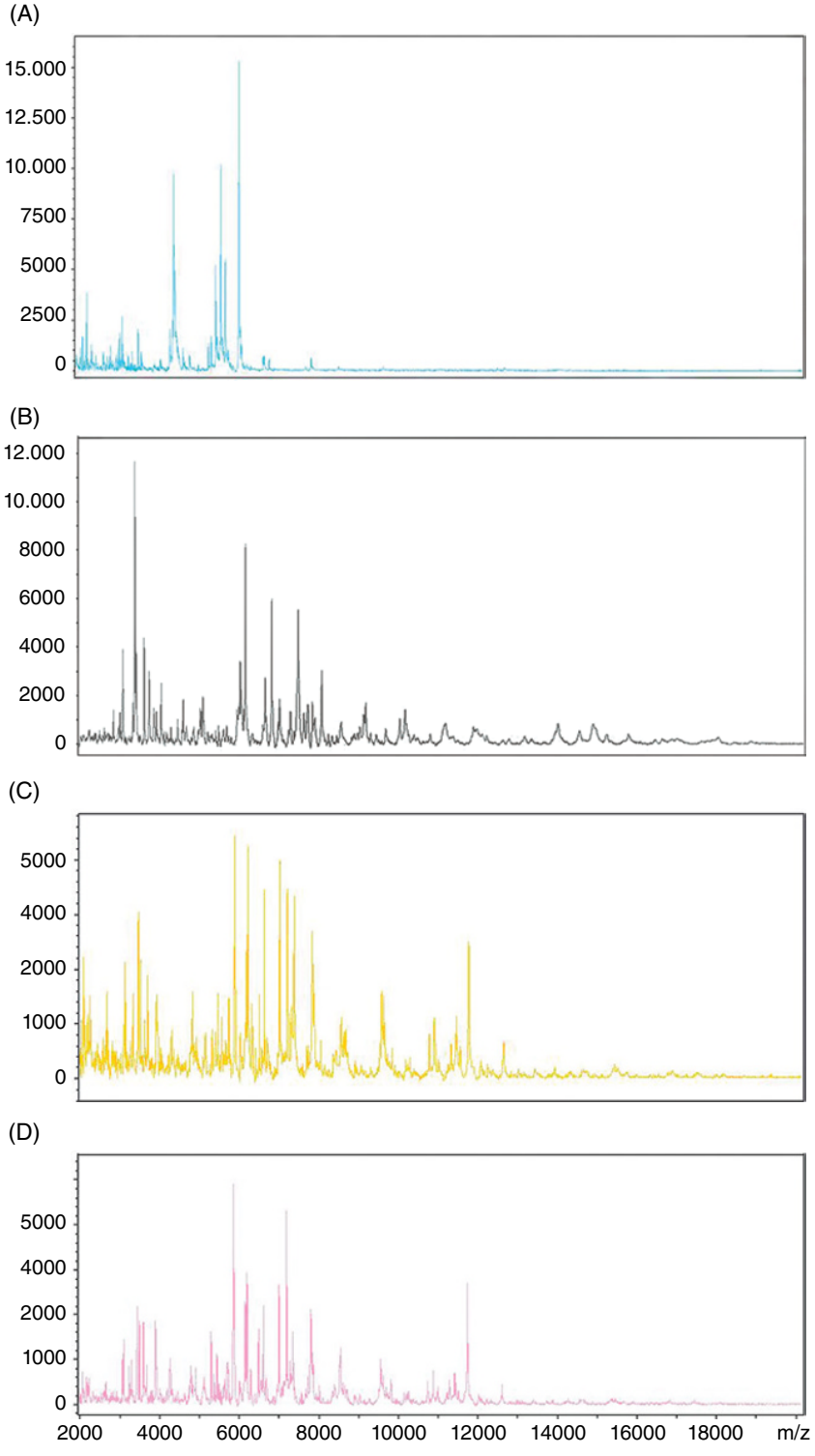


Figure 9.1 Spectra obtained for *Aspergillus udagawae* DMic 062879 (A and B) and *Trichosporon asahii* CBS 2479 (C and D) by the MALDI Biotyper® system. Spectra A and C were obtained by the MALDI Biotyper standard protocol, and spectra B and D were obtained by the NIH-National Institutes of Health protocol.

repetitions were required to obtain an accurate assessment. Spectral data produced with different methodologies and quality assurances have been shared among laboratories around the world. A data cleaning process, which involve a process of detecting, diagnosing and editing faulty data in in-house libraries that archive spectra obtained by different methodologies is mandatory, and a similar approach should be implemented with the commercial databases. More extensive data about the methodology associated with the archival spectral files in these databases is needed. It is suggested that a joint effort on data cleaning processes should be addressed by all data providers to avoid spectral data that does not represent real fungal species identity.

9.2.3 VITEK® MS

The VITEK MS system was developed with the experience gained on the Spectral ARchive And Microbial Identifications System (SARAMIS™), developed by AnagnosTec (Germany), with the contributions of researchers participating in the biennial SARAMIS Experts Meeting held in Potsdam, Germany, and once in Guimarães, Portugal (held by Micoteca da Universidade do Minho culture collection). The VITEK MS comprises a MALDI-TOF MS spectral database with a level of microbial diversity comparable to the MALDI Biotyper, in terms of strain origin and sites of isolation.

Stackebrandt *et al.* (2005) were pioneers in the use of MALDI-TOF MS and the SARAMIS software package for grouping bacteria on the basis of similarities: a set of Myxococci (*Corallococcus*) strains was evaluated, and the methodology based on this study was used in the SARAMIS software package for microbial identification, before the enterprise changed the sample preparation protocol. Briefly, a few micrograms of dried bacteria were placed onto a stainless steel target plate, and 1 µl of matrix [10 mg/ml 2,5-dihydroxybenzoic acid (DHB) in water/acetonitrile (1:1) with 0.03% trifluoroacetic acid (TFA)] was added immediately. Samples were air-dried at room temperature and analyzed by MALDI-TOF MS. Kallow *et al.* (2006) were able to separate the newly described species *Aspergillus ibericus* (Serra *et al.*, 2006) from the two closely related species *A. niger* and *A. carbonarius* within the section *Nigri* by MALDI-TOF MS using the SARAMIS software package. A similar protocol was used for sample preparation as described for bacteria with DHB and TFA.

Seyfarth *et al.* (2008) and Erhard *et al.* (2008) published the first protocols for analysis of clinically relevant filamentous fungi by MALDI-TOF MS to be identified using the SARAMIS software package. The method was similar to the protocol established by Stackebrandt *et al.* (2005) and Kallow *et al.* (2006), although with fresh cells rather than dried cell and 0.3 µl of DHB matrix and 1% TFA, instead 1 µl of DHB and 0.3% TFA. From this point, this new protocol was used, and the spectral information generated is still available and has been widely used for research and clinical identification. This protocol was optimized for use in commercial kits.

Currently, VITEK MS is a customized system, which makes the MALDI-TOF MS analysis expensive due to the stringent protocols and disposable consumables that need to be used. This optimized workflow involves a pipette tip to transfer a piece of fungal culture from an agar medium to a 2 ml microtube containing 50 µl TFA. Then, the sample is vortexed for a few seconds and allowed to stand for 30 min. The acidic suspension is then diluted 1/10 by adding 450 µl demineralized and deionized water to each microtube. Samples are vortexed for a few seconds, and then 1 µl of the suspension is spotted onto the disposable MALDI-TOF sample plate. Samples are air-dried, and then 1 µl of the customized VITEK MS-CHCA matrix is added (VITEK MS-CHCA 41107). The

sample is then immediately analyzed by the VITEK MS system, and the results are analyzed by the SARAMIS software using the automatic method and linear mode. The results are expressed as confidence values ranging from 0 to 99.9. Results of 99.9 have a strong match and may be reported. In contrast, results from 60 to 99.8 indicate a low discrimination capability and require further review. Values lower than 60 represent low-quality results with no associated identification.

The current VITEK MS system evolved from the Axima@SARAMIS software package due to the efforts of various researchers using clinical samples from different countries, but it was only used for research proposes (Dias *et al.*, 2011; Lima-Neto *et al.*, 2014; More and Rosselló-Móra 2011; Oliveira *et al.*, 2015; Santos *et al.*, 2011). Then, when bioMérieux bought the Axima@SARAMIS system, it was incorporated into the VITEK MS RUO (research use only). A pre-commercialized database called VITEK MS IVD (In Vitro Diagnostic Applications) version-1 was established, and continuously updated versions have been available to the costumers.

A direct blood culture test using culture-independent methods for fungal identification by VITEK MS was also developed by bioMérieux. The basis of the test is a customized kit modified to match the RUO sample processing method, where identification of pathogenic fungi in blood is directly analyzed by the VITEK MS system. The customized kit contents include the VITEK MS blood culture lysis and washing buffers. The procedure is similar to that of the MALDI Biotyper as the same reagents are used for sample preparation.

SARAMIS database works as part of a network where spectra archived by end users become available for participants in the network. A disadvantage therefore is that errors and/or low-quality spectra can be propagated in the network database. Furthermore, laboratories are unable to maintain confidentially.

9.2.4 MS LT2-ANDROMAS

The MS LT2-ANDROMAS is an integrated MALDI-TOF MS system with an MS database, and can be used for clinical yeast and filamentous fungal identifications. The system operates with the same sample preparation and form of analysis as the MALDI Biotyper and VITEK MS. However, operating information and the interface spectral acquisition, analysis and identification is only available in French, which restricts its use. It is possible to add spectra by using MS LT2-ANDROMAS to develop in-house libraries, which can be used to increase the number of reference strains.

It should be noted that in all of the above systems, it is essential that correctly identified fungi are employed using DNA sequencing wherever possible, to generate reliable reference data.

9.3 Effect of In-House Sample Preparation on Database Reliability

9.3.1 Yeast Identification in Pure Culture

Amiri-Eliasi and Fenselau (2001) were pioneers in the identification of yeast by MALDI-TOF MS. The authors could identify and characterize *Saccharomyces cerevisiae*, and described an optimal technique for yeast cell wall lysis for analysis which involved the

use of high concentrations of formic acid solutions. Identification of clinically relevant yeast by MALDI-TOF MS has been addressed frequently in the literature (Croxatto *et al.*, 2012; Lima-Neto *et al.*, 2014; Maier *et al.*, 2006; Oliveira *et al.*, 2015; Santos *et al.*, 2011; Sow *et al.*, 2015; Wang *et al.*, 2014). However, various sample preparation methods were reported among the different laboratories, or even within a laboratory or research group (Bidart *et al.*, 2015).

Modified protocols for sample preparation have been established by the main MALDI-TOF MS manufacturers for yeasts to improve spectral acquisition. These ranged from on-site plate inactivation with protein extraction using organic solvents plus the matrix to a full-scale protein extraction in microtubes with ethanol-based methods (Bernhard *et al.*, 2014; Clark *et al.*, 2013; De Carolis *et al.*, 2014b; Reich *et al.*, 2013).

However, the clinical yeast inter- and intraspecific diversity is much greater than the spectral information currently stored on the databases, and new spectra are continually being archived. The choice of best MALDI-TOF matrices and yeast cell disruption methods to release intracellular proteins have been evaluated, and this shows that the simplest sample preparation procedure is the direct smear protocol, for which different matrices are commercially available and have been widely used. The most commonly used matrices for yeast identification are the aforementioned DHB, CHCA plus sinapinic acid (SA). These are optimal matrices for detection of ions in the mass spectra range used for microbial identification (Bader, 2013). The most frequently used laser for microbiological applications is nitrogen, which works at 337 nm wavelength, and DHB, CHCA and SA are compatible. It is necessary to point out that the chosen matrices have to be in accordance with the laser wavelength that each equipment uses.

A standard procedure based upon yeast cell inactivation and further protein extraction is widely used. Although the full details are not given in the manuals for yeast identification of the main MALDI-TOF MS manufacturers, several in-house MALDI-TOF MS yeast libraries based on this sample preparation procedure have been established.

The method consists of suspending yeast cells in an aqueous ethanol solution (water-ethanol 1:3, v/v) followed by a centrifugation step. The pellet obtained is further vortexed with a solution composed of 70% aqueous solution of formic acid and an equal volume of acetonitrile. The sample is finally transferred onto a MALDI-TOF MS sample plate and overlaid with an appropriate matrix solution (Bader, 2013; Posteraro *et al.*, 2013). This method is laborious and time consuming and is considered a potential limitation in clinical laboratories (Chen *et al.*, 2013). In order to overcome this, alternative protocols have been evaluated (Bidart *et al.*, 2015).

Short 'on-target lysis', 'on-plate extraction' or 'fast formic acid' extraction involve an additional step consisting of yeast cell treatment with ca. 1 µl of 25%–70% formic acid on the smeared yeasts on the MALDI-TOF sample plate which is air-dried at room temperature. Ethanol can be added in order to aid cell lysis, after which the matrix solution is overlaid on the dried sample (Posteraro *et al.*, 2013).

Revisions of the manufacturer-recommended criteria for yeast identification are available. Cassagne *et al.* (2013) compared the impact of four pre-treatment procedures on the performance of MALDI-TOF-based identification: direct smear, fast formic acid extraction and two additional variants of a standard protocol involving complete formic acid/acetonitrile extractions. The authors concluded that the longer, complete extraction

procedures are better suited for yeast identification than the shorter approach. Smear and fast formic acid extraction procedures yielded lower (<40%) correct identification rates than the use of the two complete extraction procedures (>77%) of portions of colonies grown on Sabouraud and chromogenic media.

De Carolis *et al.* (2014a) evaluated the impact of sample preparation on database performance. These authors used an adapted sample preparation protocol based on a preliminary yeast cell extraction with a 10% aqueous solution of formic acid. Using this sample preparation procedure, an in-house MALDI-TOF MS library with spectra of 156 reference and clinical yeast isolates (48 species belonging to 11 genera) was created. After a retrospective validation study, the database was evaluated on 4232 yeasts routinely isolated during a 6-month period and analyzed by MALDI-TOF MS. A high level (95.8%) of correct yeast identification was obtained.

The establishment of a public database of spectra, in which researchers could deposit their spectra made in their in-house systems, has been suggested in different fora. However, to avoid mistakes, it will be necessary to obtain agreement on the most suitable culture growth conditions and protein extraction protocols. To ensure a high-quality, complete and accurate database, the spectral deposition should be monitored in compliance with quality assurance quality control guidelines (Vermeulen *et al.*, 2012).

However, reproducibility of yeast identification by MALDI-TOF MS may be unaffected by growth conditions such as culture media or temperature (Bernhard *et al.*, 2014; Clark *et al.*, 2013; De Carolis *et al.*, 2014b; Reich *et al.*, 2013). The main reason for this is that the ribosomal proteins, which are the target compounds analyzed by MALDI-TOF MS, do not vary with environmental conditions. This is contradicted by the observation that when MALDI-TOF MS is used for identification of closely related yeasts, standardization of growth conditions and sample preparation protocols remain crucial. For example, standardized MALDI-TOF MS analysis is able to distinguish between yeast complexes, such as *Candida parapsilosis*, *C. orthopsilosis* and *C. metapsilosis* (Santos *et al.*, 2011); *Cryptococcus neoformans* and *C. gattii* (Balážová *et al.*, 2014; Firacative *et al.*, 2012; McTaggart *et al.*, 2011; Posteraro *et al.*, 2012; Šedo *et al.*, 2013), and the *Sporothrix schenckii* complex (Oliveira *et al.*, 2015). Very closely related species such as C. *glabrata*, C. *bracarensis*, C. *nivariensis*, C. *albicans* and C. *dubliniensis* (Santos *et al.*, 2011) or phenotypically similar species such as *Candida famata* and *C. guillermondii* (Castanheira *et al.*, 2013; Ghosh *et al.*, 2015) can also be distinguished.

The authors obtained preliminary data on growth conditions of clinical yeasts and sample preparation that can affect MALDI-TOF MS identifications of closely related species. Figure 9.2 shows an IGS1 phylogenetic analysis and a statistical grouping obtained by using the MALDI Biotyper with spectra of clinical isolates of *Trichosporon faecale and T. asahii*. All the isolates tested were also identified using morphology and biochemistry methods according to Chagas-Neto *et al.* (2009). Two methods of protein extraction were tested for their accuracy to identify the *Trichosporon* strains. Figure 9.2B illustrates the protein spectra generated by the MALDI Biotyper using the manufacturer's protocol (ethanol-formic acid-acetonitrile) for extraction. In contrast, Figure 9.2C illustrates the protein spectra obtained by the MALDI Biotyper using the protocol with trypsin as an additional proteolytic treatment. Figures 9.2B and C show that changes in the protocol for MALDI-TOF MS analysis impacts the clustering of the strains. By adding trypsin to the protocol, the protein spectra changed significantly, as documented by the number of ions peaks and respective molecular weight generated by both methods.

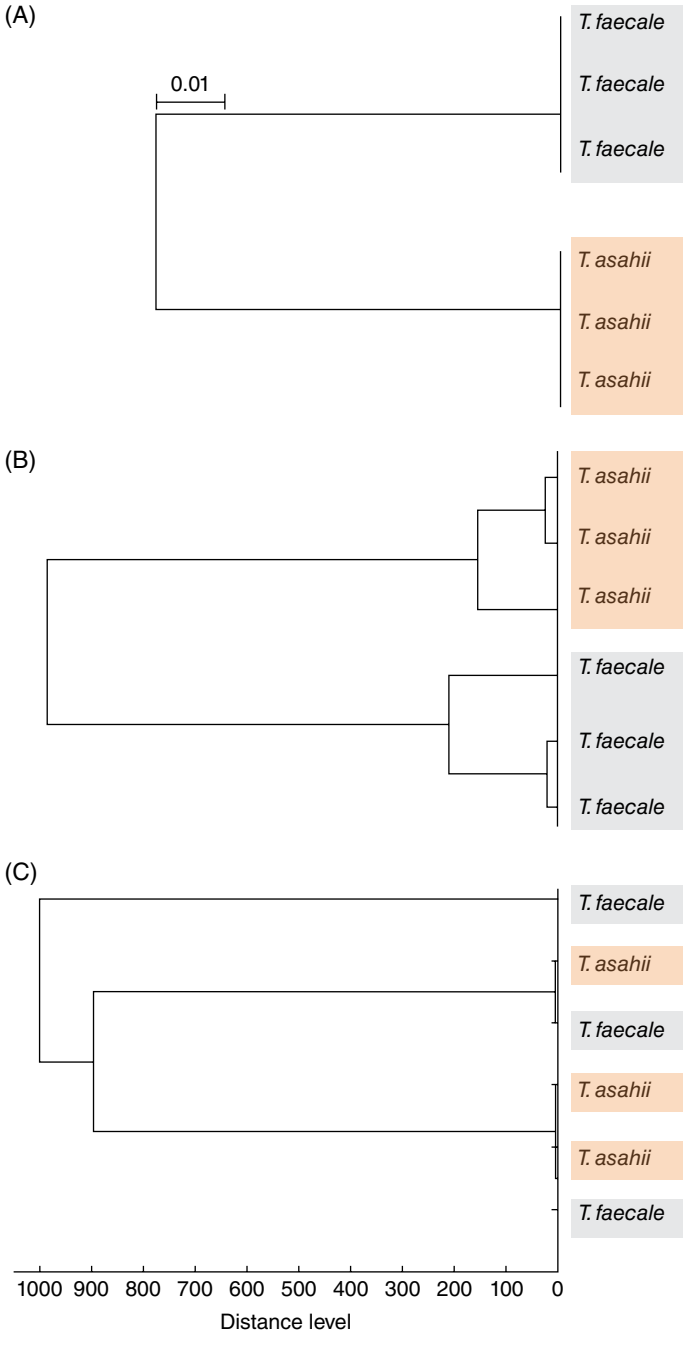


Figure 9.2 Dendrograms obtained for species of *Trichosporon* by (A) molecular biology through the IGS1 region sequencing; by MALDI Biotyper® using (B) the manufacturer's protocol (ethanol-formic acid-acetonitrile); and (C) using protein extraction by treatment of the sample with trypsin (Credits are due to Elaine Francisco (UNIFESP, Brazil] and Cledir Santos (Universidad de La Frontera, Chile].)

Consequently, the *Trichosporon* strains clustered differently as a result of the trypsin in the protein extraction protocol. However, more extensive work is required for final conclusions to be made.

9.3.2 Filamentous Fungi Identification

Filamentous fungi are more difficult than yeast to analyze by MALDI-TOF MS. Diverse spectral profiles of the same filamentous fungus strain can be obtained for reasons other than genuine taxonomic differences; for example, samples prepared with mycelia, conidiophores and/or spores may give different results. Similarly, (a) newly grown compared to preserve fungi and/or (b) the presence of secondary metabolites may affect results. The CHCA and DHB are the optimal matrices for filamentous fungi identification, where the principal objective is to achieve an optimal signal-to-noise ratio (Santos *et al.*, 2010; Simões *et al.*, 2013). Nonetheless, if all these are standardized, the results obtained become more comparable and reliable, if they are represented in the reference mass spectra database (Santos *et al.*, 2010). Finally, the ionization process in MALDI-TOF MS could be inhibited if there are high concentrations of melanin in the fungi (Buskirk *et al.*, 2011; Valentine *et al.*, 2002).

Welham *et al.* (2000) published a formative paper for the characterization of fungal spores of *Penicillium* spp., *Scytalidium dimidiatum* and *Trichophyton rubrum* using MALDI-TOF MS. In order to provide a source of protons to facilitate sample ionization, fungal spores were washed with 0.1% TFA aqueous solution. Samples were then vortexed and centrifuged, after which the fungal spores were re-suspended in methanol and transferred to the MALDI-TOF sample plate, covered with the matrix solution and analyzed.

This intact spore methodology allowed procurement of simple spectra with only a few peaks of the cellular material over the mass range 2 to 13 kDa that enabled differentiation of *Penicillium* strains, *Scytalidium dimidiatum* and *Trichophyton rubrum*. Additional studies were developed for filamentous fungi on the basis of their spores (Chen and Chen, 2005; Kemptner *et al.*, 2009; Valentine *et al.*, 2002). No databases are available based on poor spectra such as those generated from fungal spores. For reliable filamentous fungi species identification, more peaks are required, especially in the case of closely related clinical filamentous fungi.

Identification of clinically relevant filamentous fungi is based on the analysis of the whole fungal biomass of fungi such as dermatophytes. Pereira *et al.* (2014) evaluated the potential of MALDI-TOF MS for the typing and identification of clinical isolates of *Trichophyton rubrum*. Isolates were identified at the species level, and results were compared with sequencing of the internal transcribed spacers (ITS1 and ITS2) with the 5.8S rDNA region. Furthermore, in order to assess the intraspecific variability of *Trichophyton rubrum* strains, PCR fingerprinting analysis using primers M13, $(GACA)_4$ and $(AC)_{10}$ was performed.

The authors evaluated DHB and CHCA matrices. In addition, (1) different extraction protocols (including formic acid and sonication) and (2) previous suspension of cells in water or solvent mixtures were evaluated. These did not improve results compared to the smear application of cells onto the MALDI-TOF sample plate. Results provided new insights into mass spectra analysis and the potential for strain typing for MALDI-TOF MS. Statistical analysis of the mass spectral profiles (Figure 9.3A) and comparison with

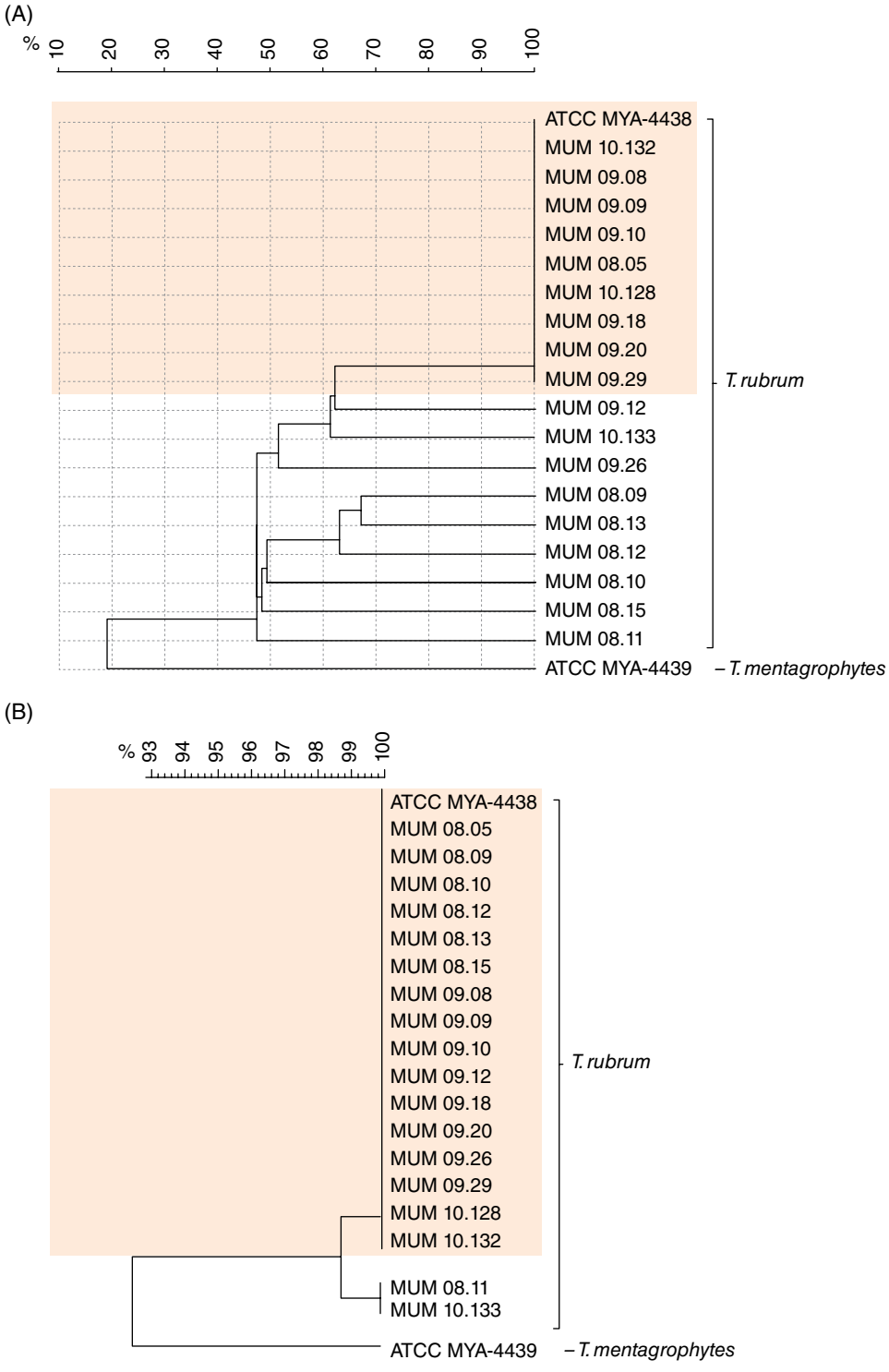


Figure 9.3 Analysis of the mass spectral profiles obtained by SARAMIS™ software package for (A) *T. rubrum* strains and (B) its comparison with molecular ITS sequence.

molecular ITS profiles (Figure 9.3B) provided the determination of phenotypic similarity and variability of *T. rubrum* strains. Moreover, MALDI-TOF MS was a valid alternative tool for intraspecific variability determination of *T. rubrum* at the strain level.

The MALDI-TOF spectra-based dendrogram was obtained by using the SARAMIS software package. The *T. mentagrophytes* strain is separate from the *T. rubrum* strains from MALDI-TOF and ITS. Interestingly, the MALDI-TOF MS profiles demonstrated that 10 strains (about 53%) were identical (Figure 9.3A), whereas 17 (about 90%) were identical by ITS (Figure 9.3B). The two strains which were dissimilar from the majority were MUM 08.11 and 10.133. The MALDI-TOF MS method indicated differences among those strains, where about 47% exhibited considerable spectrum variation between the strains and from the main cluster. Interestingly, strains MUM 08.11 and 10.133 were distinct from each other in Figure 9.3A. The clonal propagation of this fungus that is well adapted to its ecological niche could elucidate these results, and the sources from where each strain was isolated would be useful information to obtain.

The sample preparation is the most important step in the procedure (Chalupova *et al.*, 2013). However, due to the heterogeneity of some microbial taxa, it is very difficult to determine which strategy is optimal for these. Figure 9.4 shows a

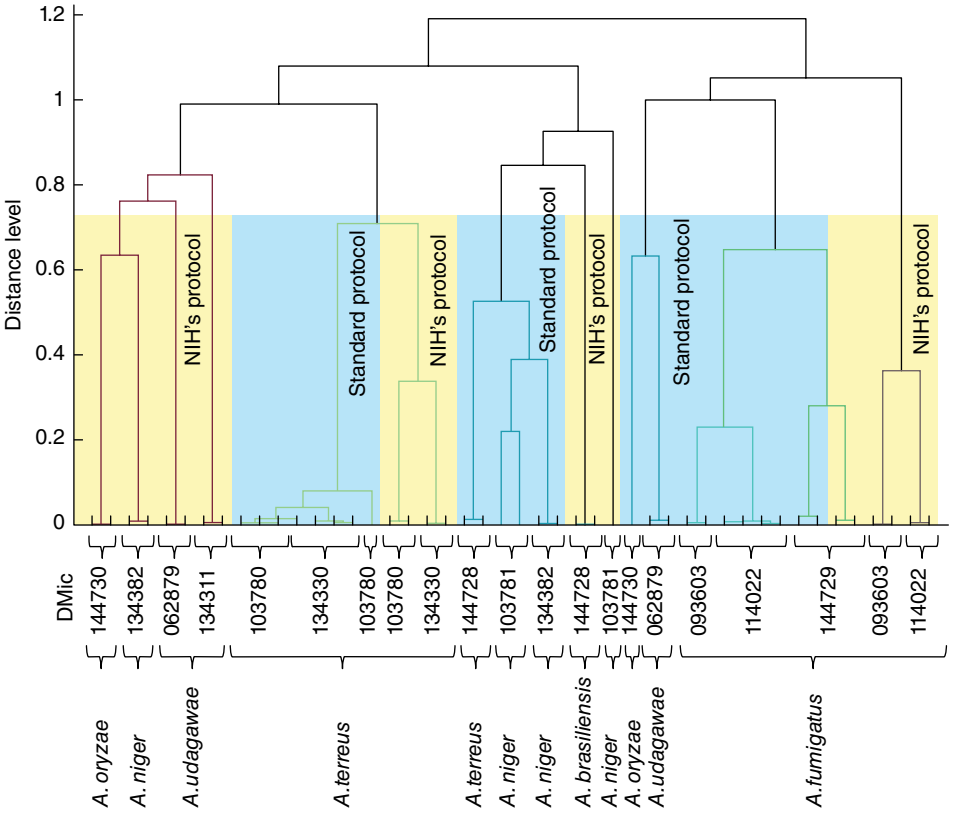


Figure 9.4 Dendrogram obtained for DMic strains with data generated by the MALDI Biotyper® from applying the manufacture's protocol with ethanol-formic acid-acetonitrile in blue and the NIH protocol with absolute ethanol and zirconia-silica beads in yellow.

comparative study developed with reference strains deposited at the Department of Mycology (DMic) culture collection of Instituto Nacional de Enfermedades Infecciosas of Dr. Carlos G. Malbrán, Buenos Aires, Argentina. Sample preparation was based upon two methodologies, mainly by applying the standard protocol with ethanol-formic acid-acetonitrile and by the US National Institutes of Health (NIH) protocol, which uses absolute ethanol and zirconia-silica beads to increase the extraction. This shows that sample preparation can have an effect on the final fungal groupings.

Some strains present further difficulties when analyzed, such as *A. udagawae* DMic134311 and *A. brasiliensis* DMic144728, where spectra were not generated by the protocols. When analyzed by the NIH protocol, *A. udagawae* DMic062879 gave a spectrum closer to *A. fumigatus* than the same strain analyzed by the manufacture's protocol (Figure 9.4). This is particularly important as *A. fumigatus* is a very important human pathogen where correct identification is essential.

Furthermore, 11 strains were evaluated in quadruplicate by both protocols in this study. Only *A. terreus* and *A. fumigatus* analyzed by the manufacture's protocol generated spectra for all tests. For the other strains, the manufacture's protocol was also the most effective in spectral generation when compared with the NIH's protocol. Spectral differences in a database can generate misidentification in clinically relevant filamentous fungi. The worst situation would be misidentifications in the clinical setting where a patient's health is at risk. The transfer of this type of data between clinical laboratories could propagate incorrect information across the entire network, leading to long-standing misidentifications.

9.4 Conclusion

MALDI-TOF MS is an accurate and rapid technique for identification of clinically important yeasts and filamentous fungi. It is inexpensive in terms of labour and consumables. However, general acceptance as the only procedure required for an identification will be achieved only when more diverse taxa are studied by employing extensive databases. The method adds an important additional range of characters for the generally recommended polyphasic approach essential for fungi. The limitations of (1) different manufacturers' equipment providing dissimilar results, (2) databases with limited coverage of representative fungal taxa and (3) diverse sample preparation protocols demonstrate the need to standardize results and extend research and development. The intrinsic variation in characters of different strains of the same fungal taxa also needs addressing, although this is a general requirement addressed by the polyphasic approach. The number of databases is increasing, but they will need to become more interpretative to cope with the slightly different spectra generated by different sampling methods. Further work will lead to construction of a global database based on a distributive network, to reduce data fragmentation and promote the use of MALDI-TOF MS for clinical fungi. With approximately 500 publications since 2010 (SCOPUS) concerning MALDI-TOF MS and fungal identifications, and with more general use of the method at the point of care in clinics, database evolution is required to fully realize the vast potential of this technology in the clinical mycology field.

References

Alexander, D., Kennedy, A. D., Desai, N., Kensicki, E., and Pappan, K. L. (2015). Metabolomics of food- and waterborne fungal pathogens. In R. R. M. Paterson and N. Lima (Eds.), *Molecular Biology of Food and Water Borne Mycotoxigenic and Mycotic Fungi* (Ch. 5, pp. 57–68). CRC Press, Boca Raton.

Amiri-Eliasi, B., and Fenselau, C. (2001). Characterisation of protein biomarkers desorbed by MALDI from whole fungal cells. *Anal. Chem.* 73: 5228–5231.

Bader, O. (2013). MALDI-TOF-MS-based species identification and typing approaches in medical mycology. *Proteomics.* 13: 788–799.

Balážová, T., Makovcová, J., Šedo, O., Slaný, M., Faldyna, M., and Zdráhal, Z. (2014). The influence of culture conditions on the identification of *Mycobacterium* species by MALDI-TOF MS profiling. *FEMS Microbiol. Lett.* 353: 77–84.

Becker, P. T., de Bel, A., Martiny, D., Ranque, S., Piarroux, R., Cassagne, C., Detandt, M., and Hendrickx, M. (2014). Identification of filamentous fungi isolates by MALDI-TOF mass spectrometry: Clinical evaluation of an extended reference spectra library. *Med. Mycol.* 52: 826–834.

Bernhard, M., Weig, M., Zautner, A. E., Groß, U., and Bader, O. (2014). Yeast on-target lysis (YOTL), a procedure for making auxiliary mass spectrum data sets for clinical routine identification of yeasts. *J. Clin. Microbiol.* 52: 4163–41637.

Bidart, M., Bonnet, I., Hennebique, A., Kherraf, Z. E., Pelloux, H., Berger, F., Cornet, M., Bailly, S., and Maubon, D. (2015). An in-house assay is superior to Sepsityper for direct matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry identification of yeast species in blood cultures. *J. Clin. Microbiol.* 53: 1761–1764.

Bridge, P. D., Hawksworth, D. L., Kozakiewicz, Z., Onions, A. H. S., Paterson, R. R. M., and Sackin, M. J. (1989a). A reappraisal of the terverticillate penicillia using biochemical, physiological and morphological features II. *Identification. J. Gen. Microbiol.* 135: 2967–2978.

Bridge, P. D., Hawksworth, D. L., Kozakiewicz, Z., and Sneath, P. H. A (1989b). A reappraisal of the terverticillate penicillia using biochemical, physiological and morphological features I. Numerical taxonomy. *J. Gen. Microbiol.* 135: 2941–2966.

Buskirk, A. D., Hettick, J. M., Chipinda, I., Law, B. F., Siegel, P. D., Slaven, J. E., Green, B. J., and Beezhold, D. H. (2011). Fungal pigments inhibit the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis of darkly pigmented fungi. *Anal. Biochem.* 411: 122–128.

Cain, T. C., Lubman, D. M., Weber, W. J. Jr., and Vertes, A. (1994). Differentiation of bacteria using protein profiles from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 8: 1026–1030.

Chalupova, J., Raus, M., Sedlářová, M., and Šebela, M. (2013) Identification of fungal microorganisms by MALDI-TOF mass spectrometry. *Biotechnol. Adv.* 32: 230–241.

Castanheira, M., Woosley, L. N., Diekema, D. J., Jones, R. N., and Pfaller, M. A. (2013). *Candida guilliermondii* and other species of *Candida* misidentified as *Candida famata*: assessment by vitek 2, DNA sequencing analysis, and matrix-assisted laser desorption ionization-time of flight mass spectrometry in two global antifungal surveillance programs. *J. Clin. Microbiol.* 51: 117–124.

Cassagne, C., Cella, A. L., Suchon, P., Normand, A. C., Ranque, S., and Piarroux, R. (2013). Evaluation of four pretreatment procedures for MALDI-TOF MS yeast identification in the routine clinical laboratory. *Med. Mycol.* 51: 371–377.

Chagas-Neto, T. C., Chaves, G. M., Melo, A. S., and Colombo, A. L. (2009). Bloodstream infections due to *Trichosporon* spp.: Species distribution, *Trichosporon asahii* genotypes determined on the basis of ribosomal DNA intergenic spacer 1 sequencing, and antifungal susceptibility testing. *J. Clin. Microbiol.* 47: 1074–1081.

Chen, H. Y., and Chen, Y. C. (2005). Characterization of intact *Penicillium* spores by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 19: 3564–3568.

Chen, Y. S., Liu, Y. H., Teng, S. H., Liao, C. H., Hung, C. C., Sheng, W. H., Teng, L. J., and Hsueh, P. R. (2013). Evaluation of the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry Bruker Biotyper for identification of *Penicillium marneffei*, *Paecilomyces* species, *Fusarium solani*, *Rhizopus* species, and *Pseudallescheria boydii*. *Front. Microbiol.* 6: 679, doi:10.3389/fmicb.2015.00679.

Clark, A. E., Kaleta, E. J., Donna, A. A., and Wolk, M. (2013). Matrix-assisted laser desorption ionization-time of flight mass spectrometry: A fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* 26: 547–603.

Croxatto, A., Prod'hom, G., and Greub, G. (2012). Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiol. Rev.* 36: 380–407.

De Carolis, E., Posteraro, B., Lass-Flörl, C., Vella, A., Florio, A. R., Torelli, R., Girmenia, C., Colozza, C., Tortorano, A. M., Sanguinetti, M., and Fadda, G. (2012). Species identification of *Aspergillus*, *Fusarium* and Mucorales with direct surface analysis by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Microbiol. Infect.* 18: 475–484.

De Carolis, E., Vella, A., Vaccaro, L., Torelli, R., Posteraro, P, Ricciardi, W., Sanguinetti, M., and Posteraro, B. (2014a). Development and validation of an in-house database for matrix-assisted laser desorption ionization-time of flight mass spectrometry-based yeast identification using a fast protein extraction procedure. *J. Clin. Microbiol.* 52: 1453–1458.

De Carolis, E., Vella, A., Vaccaro, L., Torelli, R., Spanu, T., Fiori, B., Posteraro, B., and Sanguinetti, M. (2014b). Application of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *J. Infect. Dev. Ctries.* 8: 1081–1088.

Dias, N., Santos, C., Portela, M., and Lima, N. (2011). Toenail onychomycosis in a Portuguese geriatric population. *Mycopathologia.* 172: 55–61.

Erhard, M., Hipler, U. C., Burmester, A., Brakhage, A. A., and Wöstemeyer, J. (2008). Identification of dermatophyte species causing onychomycosis and tinea pedis by MALDI-TOF mass spectrometry. *Exp. Dermatol.* 17: 356–361.

Firacative, C., Trilles, L., and Meyer, W. (2012). MALDI-TOF MS enables the rapid identification of the major molecular types within the *Cryptococcus neoformans*/*C. gattii* species complex. *PLoS ONE* 7: e37566, doi:10.1371/journal.pone.0037566.

Frisvad, J. C., and Filtenborg, O. (1983). Classification of terverticillate penicillia based on profiles of mycotoxins and other secondary metabolites. *Appl. Environ. Microbiol.* 46: 1301–1310.

Frisvad, J. C., Filtenborg, O., and Thrane, U. (1989). Analysis and screening for mycotoxins and other secondary metabolites in fungal cultures by thin-layer chromatography and high-performance liquid chromatography. *Arch. Environ. Contam. Toxicol.* 18: 331–335.

Frisvad, J. C., and Thrane, U. (1987). Standardized high-performance liquid chromatography of 182 mycotoxins and other fungal metabolites based on alkylphenone retention indices and UV-VIS spectra (diode array detection). *J. Chromatogr.* 404: 195–214.

Ghosh, A. K., Paul, S., Sood, P., Rudramurthy, S. M., Rajbanshi, A., Jillwin, T. J., and Chakrabarti, A. (2015). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry for the rapid identification of yeasts causing bloodstream infections. *Clin. Microbiol. Infect.* 21: 372–378.

Holland, R. D., Wilkes, J. G., Rafii, F., Sutherland, J. B., Persons, C. C., Voorhees, K. J., and Lay, J. O. Jr. (1996). Rapid identification of intact whole bacteria based on spectral patterns using matrix assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 10: 1227–1232.

Kallow, W., Santos, I. M., Erhard, M., Serra, R., Venâncio, A., and Lima, N. (2006). *Aspergillus ibericus*: a new species of section *Nigri* characterised by MALDI-TOF MS. In W. Meyer and C. Pearce (Eds.), *Proceedings of 8th International Mycological Congress* (pp. 189–193). Medimond S.r.l, Bologna.

Kemptner, J., Marchetti-Deschmann, M., Mach, R., Druzhinina, I. S., Kubicek, C. P., and Allmaier, G. (2009). Evaluation of matrix-assisted laser desorption/ionization (MALDI) preparation techniques for surface characterization of intact *Fusarium* spores by MALDI linear time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 23: 877–884.

Lau, A. F., Drake, S. K., Calhoun, L. B., Henderson, C. M., and Zelazny, A. M. (2013). Development of a clinically comprehensive database and a simple procedure for identification of molds from solid media by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 5: 828–834.

Lima-Neto, R., Santos, C., Lima, N., Sampaio, P., Pais, C., and Neves, R. P. (2014). Application of MALDI-TOF MS for requalification of a *Candida* clinical isolates culture collection. *Braz. J. Microbiol.* 45: 515–522.

Liu, D. (Ed.) (2011). *Molecular Detection of Human Fungal Pathogens*. CRC Press, Boca Raton.

Lohmann, C., Sabou, M., Moussaoui, W., Prévost, G., Delarbre, J. M., Candolfi, E., Gravet, A., and Letscher-Bru, V. (2013). Comparison between the Biflex III-Biotyper and the Axima-SARAMIS systems for yeast identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 51: 1231–1236.

Maier, T., Klepel, S., Renner, U., and Kostrzewa, M. (2006). Fast and reliable MALDI-TOF MS-based microorganism identification. *Nature Meth.* 3: doi:10.1038/nmeth870.

MALDI Biotyper® (2014). *Protocol Guide BDAL 06-2014*. 1818454. Edition 2.

Mancini, N., De Carolis, E., Infurnari, L., Vella, A., Clementi, N., Vaccaro, L., Ruggeri, A., Posteraro, B., Burioni, R., Clementi, M., and Sanguinetti, M. (2013). Comparative evaluation of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry systems for identification of yeasts of medical importance. *J. Clin. Microbiol.* 51: 2453–2457.

McTaggart, L. R., Lei, E., Richardson, S. E., Hoang, L., Fothergill, A., and Zhang, S. X. (2011). Rapid identification of *Cryptococcus neoformans* and *Cryptococcus gattii* by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 49: 3050–3053.

More, E., and Rosselló-Móra, R. (2011). MALDI-TOF MS: A return to phenotyping in microbial identification? *Syst. Appl. Microbiol.* 34: 1.

Morgenthaler, N. G., and Kostrzewa, M. (2015). Rapid identification of pathogens in positive blood culture of patients with sepsis: Review and meta-analysis of the performance of the Sepsityper kit. *Int. J. Microbiol.* 2015: 827416.

Nicolau, A., Sequeira, L., Santos, C., and Mota, M. (2014). Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) applied to diatom identification: influence of culturing age. *Aquat. Biol.* 20: 139–144.

Oliveira, M. M. E., Santos, C., Sampaio, P., Romeo, O., Almeida-Paes, R., Pais, C., Lima, N., and Zancope-Oliveira, R. M. (2015). Development and optimization of a new MALDI-TOF protocol for identification of the *Sporothrix* species complex. *Res. Microbiol.* 166: 102–110.

Passarini, M. R. Z., Santos, C., Lima, N., Berlinck, R. G. S., and Sette, L. D. (2013). Filamentous fungi from the Atlantic marine sponge *Dragmacidon reticulatum*. *Arch. Microbiol.* 195: 99–111.

Paterson, R. R. M. (1998). Chemotaxonomy of fungi by unsaponifiable lipids. In Frisvad, J. C., Bridge, P. D., and Arora, D. K. (Eds.), *Chemical Fungal Taxonomy* (pp. 183–217). Marcel Dekker, New York.

Paterson, R. R. M., and Bridge, P. (1994). *Biochemical Techniques for Filamentous Fungi*. IMI technical handbooks. Oxford University Press, USA.

Paterson, R. R. M., and Lima, N. (Eds.) (2015). *Molecular Biology of Food and Water Borne Mycotoxigenic and Mycotic Fungi*. CRC Press, Boca Raton.

Paterson, R. R. M., Venâncio, A., and Lima, N. (2006). A practical approach for identifications based on mycotoxin characters of *Penicillium*. *Rev. Iberoam. Micol.* 23: 155–159.

Pereira, L., Dias, D., Santos, C., and Lima, N. (2014). The use of MALDI-TOF ICMS as an alternative tool for *Trichophyton rubrum* identification and typing. *Enferm. Infecc. Microbiol. Clin.* 32: 11–17.

Pitt, J. I. (1973). An appraisal of identification methods for *Penicillium* species: novel taxonomic criteria based on temperature and water relations. *Mycologia.* 65: 1135–1157.

Posteraro, B., De Carolis, E., Vella, A., and Sanguinetti, M. (2013). MALDI-TOF mass spectrometry in the clinical mycology laboratory: Identification of fungi and beyond. *Expert Rev. Proteomics.* 10: 151–164.

Posteraro, B., Vella, A., Cogliati, M., De Carolis, E., Florio, A. R., Posteraro, P., Sanguinetti, M., and Tortorano, A. M. (2012). Matrix-assisted laser desorption ionization-time of flight mass spectrometry-based method for discrimination between molecular types of *Cryptococcus neoformans* and *Cryptococcus gattii*. *J. Clin. Microbiol.* 50: 2472–2476.

Rapper, K. B., and Fennell, D. I. (1965). *The Genus Aspergillus*. Williams and Wilkins Company, Baltimore, Maryland.

Reich, M., Bosshard, P. P., Stark, M., Beyser, K., and Borgmann, S. (2013) Species identification of bacteria and fungi from solid and liquid culture media by MALDI TOF mass spectrometry. *J. Bacteriol. Parasitol.* S5–002. doi:10.4172/2155-9597. S5-002.

Robert, V., Cardinali, G., Stielow, B., Vu, T. D., Borges dos Santos, F., Meyer, W., and Schoch, C. (2015). Fungal DNA barcoding. In Paterson, R. R. M. and Lima, N. (Eds.), *Molecular Biology of Food and Water Borne Mycotoxigenic and Mycotic Fungi* (Ch. 4, pp. 37–58). CRC Press, Boca Raton.

Rodrigues, A., Maciel, M., Santos, C., Machado, D., Sampaio, J., Lima, N., Carvalho, M. J., Cabrita, A., and Martins, M. (2014). Peritoneal dialysis infections: an opportunity for improvement. *Am. J. Infect. Control.* 42: 1016–1018.

Rodrigues, P., Santos, C., Venâncio, A., and Lima, N. (2011). Species identification of *Aspergillus* section *Flavi* isolates from Portuguese almonds using phenotypic, including MALDI-TOF ICMS, and molecular approaches. *J. Appl. Microbiol.* 111: 877–892.

Santos, C., Lima, N., Sampaio, P., and Pais, C. (2011). Matrix-assisted laser desorption/ionization time-of-flight intact cell mass spectrometry (MALDI-TOF ICMS) to detect emerging pathogenic *Candida* species. *Diag. Microbiol. Infec. Dis.* 71: 304–308.

Santos, C., Paterson, R. R. M., Venâncio, A., and Lima, N. (2010). Filamentous fungal characterisations by matrix-assisted laser desorption/ionisation time of flight mass spectrometry. *J. Appl. Microbiol.* 108: 375–385.

Schulthess, B., Ledermann, R., Mouttet, F., Zbinden, A., Bloemberg, G. V., Böttger, E. C., and Hombach, M. (2014). Use of the Bruker MALDI Biotyper for identification of molds in the clinical mycology laboratory. *J. Clin. Microbiol.* 52: 2797–2803.

Šedo, O., Vávrová, A., Vaďurová, M., Tvrzová, L., and Zdráhal, Z. (2013). The influence of growth conditions on strain differentiation within the *Lactobacillus acidophilus* group using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling. *Rapid Commun. Mass Spectrom.* 27: 2729–2736.

Serra, R., Cabañes, F. J., Perrone, G., Castellá, G., Venâncio, A., Mulè, G., and Kozakiewicz, Z. (2006). *Aspergillus ibericus*: A new species of section *Nigri* isolated from grapes. *Mycologia*. 98: 295–306.

Seyfarth, F., Ziemer, M., Sayer, H. G., Burmester, A., Erhard, M., Welker, M., Schliemann, S., Straube, E., and Hipler, U. C. (2008). The use of ITS DNA sequence analysis and MALDI-TOF mass spectrometry in diagnosing an infection with *Fusarium proliferatum*. *Exp. Dermatol.* 17: 965–971.

Silva, F. C., Chalfoun, S. M., Batista, L. R., Santos, C., and Lima, N. (2015). Use of a polyphasic approach including MALDI-TOF MS for identification of *Aspergillus* section *Flavi* strains isolated from food commodities in Brazil. *Ann. Microbiol.* 65: 2119–2129.

Simões, M. F., Pereira. L., Santos. C., and Lima. N. (2013). Polyphasic identification and preservation of fungal diversity: concepts and applications. In Malik, A., Grohmann, E., and Alves, M. (Eds.), *Management of Microbial Resources in the Environment* (Ch. 5, pp. 91–116). Springer, the Netherlands.

Sow, D., Fall, B., Ndiaye, M., Ba, B. S., Sylla, K., Tine, R., Lô, A. C., Abiola, A., Wade, B., Dieng, T., Dieng, Y., Ndiaye, J. L., Hennequin, C., Gaye, O., and Faye, B. (2015). Usefulness of MALDI-TOF mass spectrometry for routine identification of *Candida* species in a resource-poor setting. *Mycopathologia.* 180: 173–179.

Stackebrandt, E., Puker, O., and Erhard, M. (2005). Grouping Myxococci (*Corallococcus*) strains by matrix-assisted laser desorption ionization time-of-flight (MALDI TOF) mass spectrometry: comparison with gene sequence phylogenies. *Curr. Microbiol.* 50: 71–77.

Valentine, N., Wahl, J., Kingsley, M., and Wahl, K. (2002). Direct surface analysis of fungal species by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 16: 1352–1357.

Vermeulen, E., Verhaegen, J., Indevuyst, C., and Lagrou, K. (2012). Update on the evolving role of MALDI-TOF MS for laboratory diagnosis of fungal infections. *Curr. Fungal Infect. Rep.* 6: 206–214.

Wang, W., Xi, H., Huang, M., Wang, J., Fan, M., Chen, Y., Shao, H., and Li, X. (2014). Performance of mass spectrometric identification of bacteria and yeasts routinely isolated in a clinical microbiology laboratory using MALDI-TOF MS. *J. Thorac. Dis.* 6: 524–533.

Welham, K. J., Domin, M. A., Johnson, K., Jones, L., and Ashton, D. S. (2000). Characterization of fungal spores by laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 14: 307–310.

10

Development and Application of MALDI-TOF for Detection of Resistance Mechanisms

Stefan Zimmermann and Irene Burckhardt

Department of Infectious Diseases, University Hospital Heidelberg, Division of Medical Microbiology and Hygiene, Heidelberg, Germany

10.1 Attempts to Correlate Signature Mass Ions in MALDI-TOF MS Profiles with Antibiotic Resistance

Although the identification of microbes using MALDI-TOF MS is well documented, less well known are the early attempts to detect antimicrobial resistance markers of bacteria using MS. The field was stimulated by an enormous increase in bacterial resistance, not only in Europe, but worldwide. Multi-resistant phenotypes of bacteria can complicate the treatment of infections in critically ill patients and delay the start of appropriate and effective therapy (Livermore, 2012). Both multi-resistant and pan-resistant strains have been described (Nordmann *et al.*, 2011; Tzouvelekis *et al.*, 2012). The multi-resistant gram-negative bacteria are a more recent occurrence, whereas the resistant gram-positive species have been known for decades. For example, in Portugal, the increase in MRSA resulted in a higher rate of *methicillin-resistant Staphylococcus aureus* (MRSA) isolates from blood cultures compared to *methicillin-susceptible Staphylococcus aureus* (MSSA). Such developments result in a high rate of in-hospital mortality, especially in the elderly (Rebelo *et al.*, 2011).

As early as 2000, Gordon's group in Manchester typed 20 MRSA strains using intact cell mass spectrometry (ICMS) and defined specific peaks for MRSA (Edwards-Jones *et al.*, 2000). They detected peaks typical for MSSA and concluded that ICMS may have the potential for MRSA identification and typing. Two years later, Du *et al.* (2002) in Beijing published confirmatory data that the spectral profiles of MSSA and MRSA differ greatly from each other. They used a nuc and mecA gene-based PCR to control the results and drew the conclusion that comparison of MALDI-TOF MS spectral profiles of microorganisms could serve as a simple and rapid method for bacterial identification and antibiotic susceptibility analysis. In that study, seven mecA negative strains were misidentified as MRSA strains, a fact which was explained by the authors to be due to the heterogeneity of methicillin resistance in *S. aureus*. In the same year, Bernardo *et al.* (2002) stated that no uniform signatures for MRSA could be detected in

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

the mass spectra of the strains they studied. Szabados *et al.* (2012) later substantiated the lack of uniform signatures for MRSA and MSSA. They intensively analyzed an isogenic Staphylococcal Cassette Chromosome (SCCmec)-harbouring parent and a SCCmec-minus daughter strain with the same genetic background and unequivocally ruled out strain-specific protein peaks and could not show differences in the peak profile. They and others concluded that the protein encoded by the mecA gene, the penicillin binding protein 2A (PBP2A), is too large (76 kDA) and present in too low a concentration to be detected by mass spectrometry directly from cell extracts during routine measurements.

In 2011, Chatterjee *et al.* described a small peptide, the phenol soluble modulin (PSM), and its gene the psm-mec, which is linked to the mecA gene complex. It is not present in all SCCmec cassette types, but only in types II, III and VIII. Even in these cassette types, the expression of the psm-mec gene is strictly regulated by the RNAIII-independent Agr system. Josten *et al.* (2014) confirmed these findings a few years later. They analyzed their MRSA strain collection, finding that 95% of agr-positive MRSAs harbouring these cassette types can be identified by this signal. They defined a surrogate marker for the activity of this expression system using delta-toxin (m/z 3007) as an indicator, as it was present in 89% of their collection. The PSM peptide was detected in more than 95% of all relevant MRSA strains with a peak signal at 2415 m/z (2411–2419 m/z strain variance) (Figure 10.1). No additional or special measurements needed to be done for acquisition of this peak. A detailed software driven re-analysis of the relevant mass zone of the routine mass spectrum acquired for identification is sufficient and can provide an early indicator for routine identification that MRSA is present.

The most important antimicrobial drugs effective against MRSA are glycopeptides. Therefore, Majcherczyk focused not only on methicillin but also investigated teicoplanin resistance of *S. aureus* isolates. MALDI-TOF MS was found to show a superior discriminatory power compared to pulsed field gel electrophoresis (PFGE) or peptidoglycan muropeptide analysis for clonal strain analysis. Convincing test results in terms of teicoplanin resistance or susceptibility was not provided (Majcherczyk *et al.*, 2006). One study was published that demonstrated the ability of MALDI-TOF MS to distinguish between vancomycin-resistant enterococci and susceptible isolates. Griffin *et al.* (2012) analyzed 67 vanB-positive *Enterococcus faecium* strains and were able to

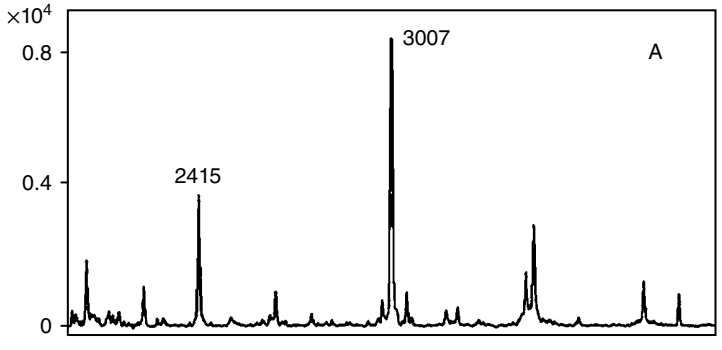


Figure 10.1 Mass spectrum of USA 100 MRSA strain. PSMmec peak (2415 m/z) as MRSA marker and the delta-toxin (3007 m/z) as surrogate marker for agr system (Josten *et al.*, 2014).

predict susceptibility or resistance to vancomycin through specific spectral peaks. The sensitivity of the assay was 97%, and the specificity was 98%. Up to now, this pattern analysis and these profiles were only used in one laboratory. Therefore, the possibility of the effects of clonal variants has not been ruled out.

10.2 Distribution and Spread of Carbapenems and Mass Spectrometry

All of these approaches share a common assumption; they infer the resistance of a specific bacterium towards a certain antibiotic from mass spectral peaks that are unrelated to the mechanism of resistance. However, these approaches need appropriate validation before the technique and its conclusions can be used in different laboratories.

It was pointed out in his review, 'Fourteen Years in Resistance' (Livermore, 2012), that the situation of multi-resistant gram-positive cocci had improved, but that of gram-negative rods had worsened over the last decade. In recent years, especially the percentage of carbapenem-resistant bacteria had increased at an alarming pace and became a major threat to patient recovery. This development culminated in the discovery of the NDM-1 carbapenemase (New Delhi Metallo-ß-lactamase), an enzyme degrading all ß-lactam antibiotics, not only in India and Pakistan, but also in Europe (Kumarasamy *et al.*, 2010). Bonomo's group in Ohio was the first to investigate the stability of carbapenems against OXA-1 carbapenemases using mass spectrometry (e.g. MALDI-TOF and ESI-TOF). They used a quadrupole time-of flight mass spectrometer with a nanospray source to measure binding between different ß-lactam antibiotics and bacterial carbapenemases of the OXA class to predict susceptibility and effectiveness (Bethel *et al.*, 2008). In more recent studies, mass spectrometry was also used to predict the structure of and antibiotic susceptibility for cephamycinase enzymes such as the widespread AmpC ß-lactamase, providing insights into the substrate specificity of this enzyme that is spreading worldwide (Lefurgy *et al.*, 2015). All resistant bacterial taxa mentioned above grow aerobically. However, rapid resistance determination of anaerobic bacteria is also of major clinical importance, especially in abdominal surgery and neurosurgery. *Bacteroides fragilis*, one of the most important species of the genus *Bacteroides*, is the leading anaerobic pathogen in human monobacterial or polymicrobial infections (Aldridge *et al.*, 2003; see Chapter 5). Members of the *Bacteroides*, including *B. fragilis*, often exhibit resistance to different drugs, such as ß-lactams, clindamycin and fluoroquinolones. However, the majority of strains are still susceptible to metronidazole and carbapenems.

It is said that up to 10% of *Bacteroides fragilis* strains might already be resistant to carbapenems. Elisabeth Nagy *et al.* (2011) from the Hungarian Anaerobe Reference Laboratory analyzed these strains and found that nearly all of them express a class B metallo-ß-lactamase, encoded by the *cfi*A gene, that confers resistance to all ß-lactam antibiotics. She reviewed the peaks in the mass spectrum of 40 strains of *Bacteroides fragilis* of known imipenem MICs and detected two specific peaks shifts from 4711 and 4817 to 4688 and 4826 Da, when the carbapenemase enzyme was expressed (see Chapter 5). The shift can be visualized in pseudo-gel view and supported by analytical software tools of the mass spectrometer. Johansson *et al.* (2014) later found out that this

method also works directly with positive blood cultures, which is important for rapid antibiotic sensitivity testing in septicaemia and other severe infections caused by anaerobes. Another Hungarian group demonstrated the usefulness of this protocol under routine conditions in a larger number of samples (Fenyvesi *et al.*, 2014).

10.3 Carbapenem-Resistant *Enterobacteriaceae*

As already mentioned, carbapenem-resistant *enterobacteriaceae* (CRE) have spread globally and represent a serious and growing threat to public health. One of the most frequent carbapenemases, which often induces hospital outbreaks and is difficult to contain by the infection control team, is *Klebsiella pneumoniae* carbapenemase (KPC) (Wendt *et al.*, 2010). Nowadays special assays are necessary to detect such resistance mechanisms (e.g. modified Hodge test or PCR), but it would be an enormous advantage for infection control to have a real-time tracking method for it while undertaking bacterial identification in the routine workflow. Lau *et al.* (2014) used a combination of proteomics and molecular techniques to identify a specific peak at 11,109 Da which is specific for the presence of a plasmid containing the ß-lactamase gene kpc (bla_{KPC}). Identification of resistance using this MALDI-TOF MS method was accomplished in as little as 10 min from single colonies and half an hour from positive blood cultures, proving the potential clinical utility for real-time plasmid tracking in an outbreak. The authors stated that although a pKpQIL_p019 MALDI-TOF MS peak was identified in all 18 re-cultured outbreak isolates studied, these did not constitute as a statistical validation set as they were all from a single outbreak. To confirm the concept, the same group analyzed 860 Enterobacteriaceae for the specific KPC peak (Figure 10.2). In a collection of outbreak strains, 100% of strains containing the resistance plasmid p019 were detected in MS analysis (n = 26), whereas in a larger retrospective investigation, 9 out of 720 isolates (1.3%) were positive (Youn *et al.*, 2016). The method was appropriate for reliable detection of the resistance mechanism and the transferable plasmid in a routine workflow automatically in real time and with high throughput. The authors emphasized that instrument tuning and calibration had a significant effect on assay sensitivity, highlighting important factors that must be considered as MALDI-TOF MS moves into applications beyond microbial identification.

10.4 MALDI-TOF MS Detection Based upon Changes in Antibiotic Structure due to Bacterial Degradation Enzymes

In spite of the accuracy of real-time methods such as the PSM peptide for MRSA or the 11,109 Da peak for the KPC plasmid, these are useful surrogate markers but not direct evidence for antibiotic degradation or modification. Two groups in Europe developed a very similar assay at the same time independently using a simple biochemical calculation of the degradation process. When a ß-lactam ring in an antibiotic substrate is degraded by a ß-lactamase, the ring is hydrolyzed, so water is added to it to give an additional mass of 18 Da. In the next step (at least for carbapenemases),

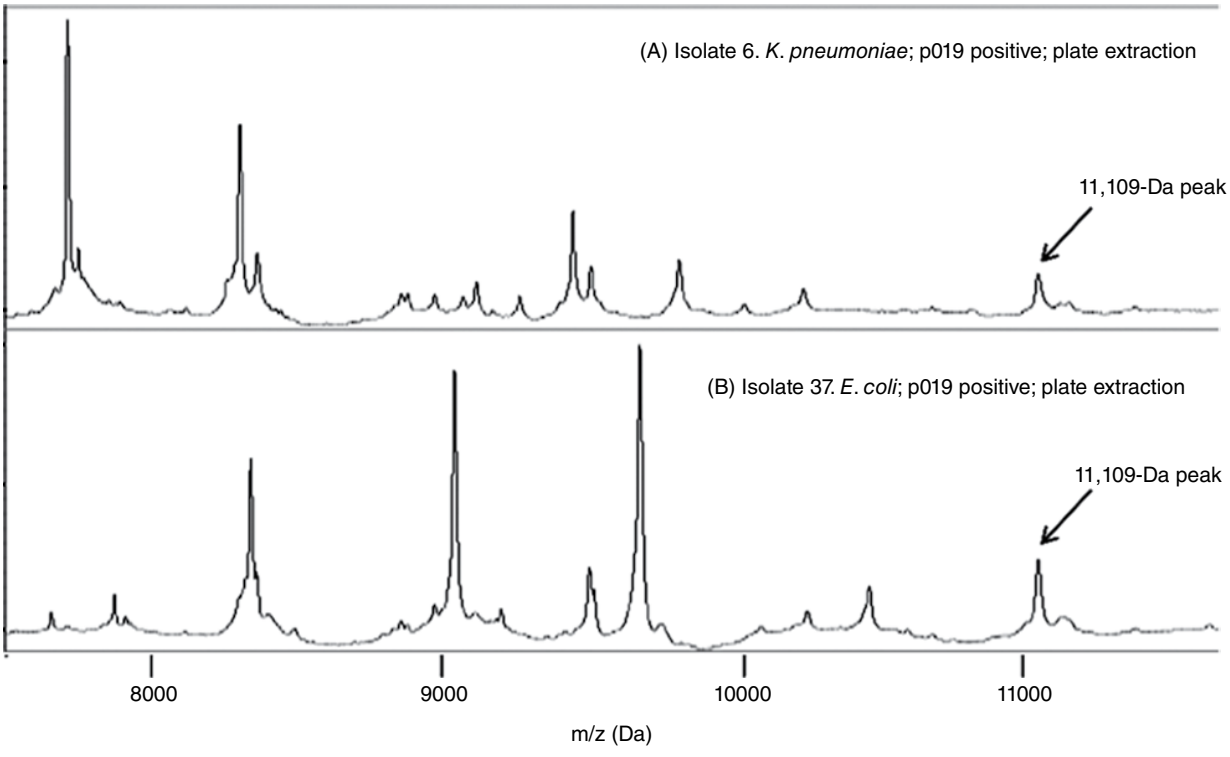


Figure 10.2 Representative spectra for p019 detection. Peak intensity is displayed in arbitrary units (10^4 A.U.). Presence of the ~11,109 Da peak in *K. pneumoniae* (A) and *E. coli* (B) p019-containing isolates (Youn *et al.*, 2016). Adapted from Youn et al., Poster at ICAAC 2014.

this unstable molecule is decarboxylated, resulting in a loss of 44 Da. So the total mass of the inactivated ß-lactam antibiotic is 26 Da lower than that of the base product. This mass difference can be measured in a MALDI-TOF mass spectrometer. Burckhardt and colleagues used ertapenem as a test substrate, because it is considered to be the most sensitive carbapenem to detect resistant phenotypes and is one of the most stable carbapenems (Burckhardt and Zimmermann, 2011). The ß-lactam was incubated with multi-resistant gram-negative rods harbouring carbapenemase enzymes such as NDM-1. At various timed intervals, bacteria were centrifuged, and the supernatant (with the carbapenem) was spotted onto a standard steel target and analyzed in a Bruker Microflex™ mass spectrometer. Ertapenem, with a molecular mass of 476 Da, is used for treatment as a sodium salt. The mass spectrometer depicts not only the pure substance at 476 Da, but also two sodium salts at 499 and 522 Da. After a short incubation period of only 1 h, the three mass peaks of ertapenem were degraded, and a degradation peak at 450 Da (minus 26 Da) was visible (Figure 10.3). This assay delivers rapid results for a number of carbapenemases in different bacteria (enterobacteriaceae, Pseudomonas, other gram-negative nonfermentative rods). Hrabak *et al.* (2011) simultaneously investigated 124 strains, including 30 carbapenemase producers, using meropenem as a carbapenem target. Both groups stated that the described method of carbapenem resistance detection has many advantages compared to agar-based methods or PCR. First, depending on the type of carbapenemase, results can be available within 1 h from the start of incubation. This is especially useful during an outbreak situation where the carbapenemase is already identified. Second, this method is comparatively easy to perform. Only readily available reagents are used, and because MALDI-TOF MS is increasingly used in routine microbiology laboratories for identification of bacterial strains, the hardware is already present in many laboratories. Third, the cost per determination is very low, less than 1 € per reaction. As the assay results are available in less than 3 h, this test can be implemented in infection control procedures identifying patients with CRE rapidly. In former years, rectal swabs or other CRE screening samples were streaked manually onto agar plates and incubated overnight. Overnight biochemical-based identification and conventional resistance screening (e.g. in an automated VITEK II instrument) was performed. If a CRE isolate was detected, it was necessary to confirm this by other manual tests, such as the modified Hodge test, which took another day. Thus, successful screening for multi-resistant microorganisms took 3 to 4 d in the laboratory, whereas today screening samples could be streaked with automatic devices on the first day.

10.5 Optimization of the Carbapenemase MALDI-TOF MS-Based Assay to Minimize the Time-to-Result

Suspicious colonies can be identified in less than an hour by mass spectrometry, and their carbapenem-degrading activity can be revealed in less than two additional hours, shortening the workflow to less than 24 h (Figure 10.4). Hrabak *et al.* (2014) suggested a similar workflow for the detection of life-threatening carbapenemases in his review concerning the usefulness of this promising technology in 2014. A major drawback in

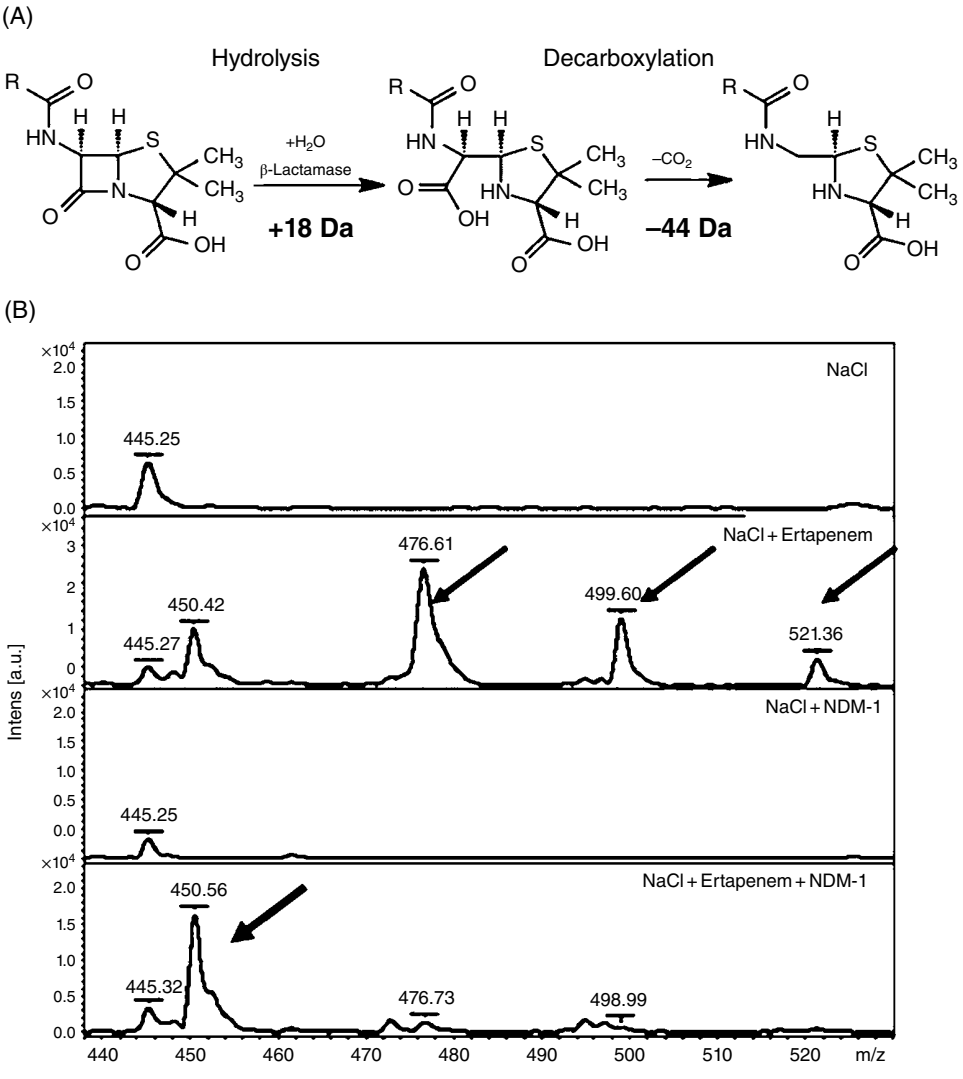


Figure 10.3 (A) Basic principle of the MS ß-lactam hydrolysis assay. The ß-lactam ring is hydrolyzed (+18 m/z), and the unstable reaction product is decarboxylated (–44 m/z). (B) Intact ertapenem (2nd panel) with 476 Da and its two sodium salts (499/522 Da), and the degraded product (4th panel) with 450 Da (–26 m/z) (Burckhardt and Zimmermann, 2011).

the beginning was the low sensitivity of the carbapenem hydrolysis assay. The main reason was the low degrading activity of carbapenemases of the OXA class with the buffer solutions used at the beginning. OXA-48 is the most common carbapenemase in Germany, which limited its routine use. Hrabak's laboratory optimized the assay conditions by adding ammonium hydrogen carbonate (NH_4HCO_3) and increasing the pH of the buffer (Hrabak *et al.*, 2014). This resulted in a significant increase of the carbapenemase activity, allowing standard assay times of less than 3 h and a sensitivity of up to 100% in the tested sample collection (Papagiannitsis *et al.*, 2015).

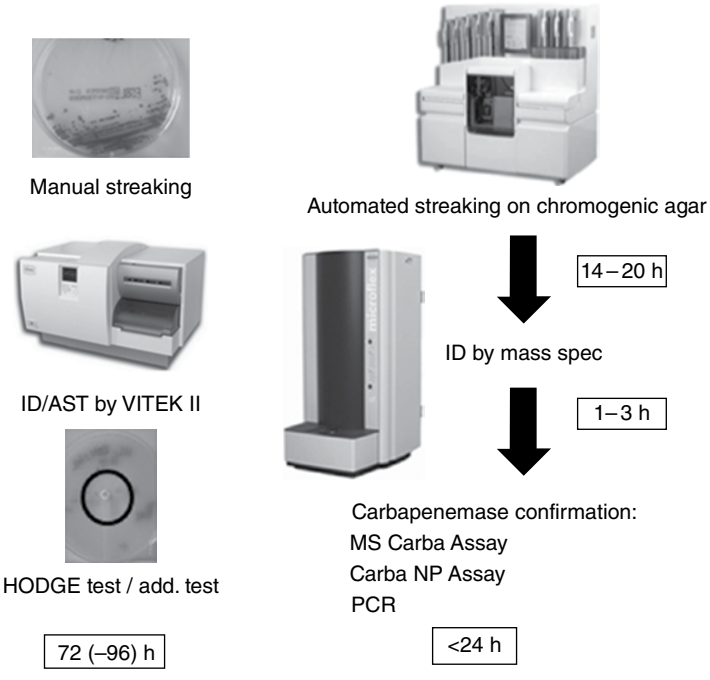


Figure 10.4 Workflow for conventional infection control screening for CRE (left panel) and with automated streaking and ID/AST determination by mass spectrometry (right panel).

10.6 Detection of Other Bacterial Enzymic Modifications to Antibiotic Structures

The principle of degradation product monitoring can be applied to other enzymatic resistance mechanisms. Green *et al.* (2010) provided evidence that not only can the degradation of ß-lactam antibiotics be depicted by mass spectrometry, but also enzymatic modifications resulting in resistance to aminoglycosides was feasible. Acetyltransferases transferring N-acylate amines to aminoglycosides are the most frequent resistance mechanisms for the inactivation of this antibiotic class. All clinically relevant aminoglycosides (gentamicin, tobramycin and amikacin) can be inactivated by N-acylate transfer in the presence of coenzyme A (CoA). Aminoglycoside-6′N acyetyltransferase (AAC6) from *Escherichia coli* was used to transfer N-propionyl to neomycin. This resistance-conferring modification could be visualized by a shift of the mass peak of neomycin from 615 to 671 Da (Green *et al.*, 2010). Variations in the usage of N-acylate-CoA already allowed a form of specificity for different aminoglycosides. Other groups confirmed these findings, i.e., that the acetyltransferase-based resistance against aminoglycosides can be detected in a rapid assay in a few hours by mass spectrometry (Burckhardt *et al.*, 2013). Both groups used electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) to demonstrate the modification of the antibiotic substrate, as aminoglycosides and their exact mass profile could not be sufficiently delineated with MALDI-TOF MS (see Chapter 15).

A few months after Hrabak and Burckhardt (2014) had published the hydrolysis assays focusing on carbapenems, Sparbier *et al.* (2012) overcame this limitation by offering a valid protocol for detecting the action of different ß-lactam molecules including penicillinases and cephalosporinases. They developed a method for the detection of ampicillin, piperacillin, cefotaxime, ceftazidime, ertapenem and meropenem and analyzed the different degradation products and sodium or potassium salt variants. This led to protocols for all groups of ß-lactam antibiotics now being available (Figure 10.5). A restriction of these new assays was that they could only detect degrading enzymes, but not other molecular mechanisms of bacterial resistance such as target modifications, efflux pumps or porin loss. Despite this fact, MALDI-TOF MS is a relevant tool for the detection of antibiotic resistance, but the application cannot replace standard susceptibility testing owing to these limitations (Hrabak *et al.*, 2013).

10.7 Isotopic Detection using MALDI-TOF MS

Jung and Kostrzewa (2014) recently described a new assay overcoming these limitations. The basic idea is to grow intact microorganisms in drug-containing stable isotope-labelled media and acquire mass spectra from these microorganisms. These spectra are compared with the mass spectra of the intact microorganism grown in non-labelled media without the drug present. Drug resistance is detected by characteristic mass shifts of one or more biomarker peaks. The method essentially is based on a publication by Demirev *et al.* (2013), who used a commercially available medium where almost all of the C atoms are ^{13}C and grew the bacteria in a medium in the presence of an antibiotic which provided evidence of its resistance against this drug. This evidence is visualized by peak shifts caused by the ^{13}C incorporation into the biomarkers. With this method, resistance detection by MALDI-TOF mass spectrometry can be expanded to antibiotics also, which show no enzymatic degradation. A disadvantage is that the completely labelled medium is costly. Sparbier *et al.* (2013) grew their bacteria in normal medium where only one lysine molecule was labelled with $^{13}C_6$ $^{15}N_2$-L-Lysine, which is more cost-effective. In this supplemented media, they grew *Klebsiella pneumoniae* strains with or without meropenem. After 2 h of incubation, the spectrum acquired from the resistant bacterium in the labelled medium with antibiotic was nearly identical to the spectrum from the organism grown in labelled medium without meropenem. In contrast, as the growth and thereby protein synthesis of the susceptible strain is inhibited, its spectrum after incubation in the labelled and antibiotic-containing medium is still similar to the "normal" spectrum (Jung *et al.*, 2014). Sparbier *et al.* (2013) used the new approach to distinguish MSSA strains from resistant MRSA strains. Cefoxitin and Oxacillin were used to investigate 20 *Staph. aureus* strains for their resistance phenotype by mass spectrometry, and the results were compared to conventional culture-based methods and PCR. Nineteen out of the 20 strains showed excellent correlation within 3 h. The strains were inoculated in three different broths: one with a non-labelled medium without antibiotics, one with antibiotics without heavy lysine and one with antibiotics and heavy-lysine-containing medium [MS-Resist™; Figure 10.6]. After the incubation period, the mass spectrum of susceptible strains showed a high similarity between the unlabelled medium and the labelled medium with antibiotic. As the strains

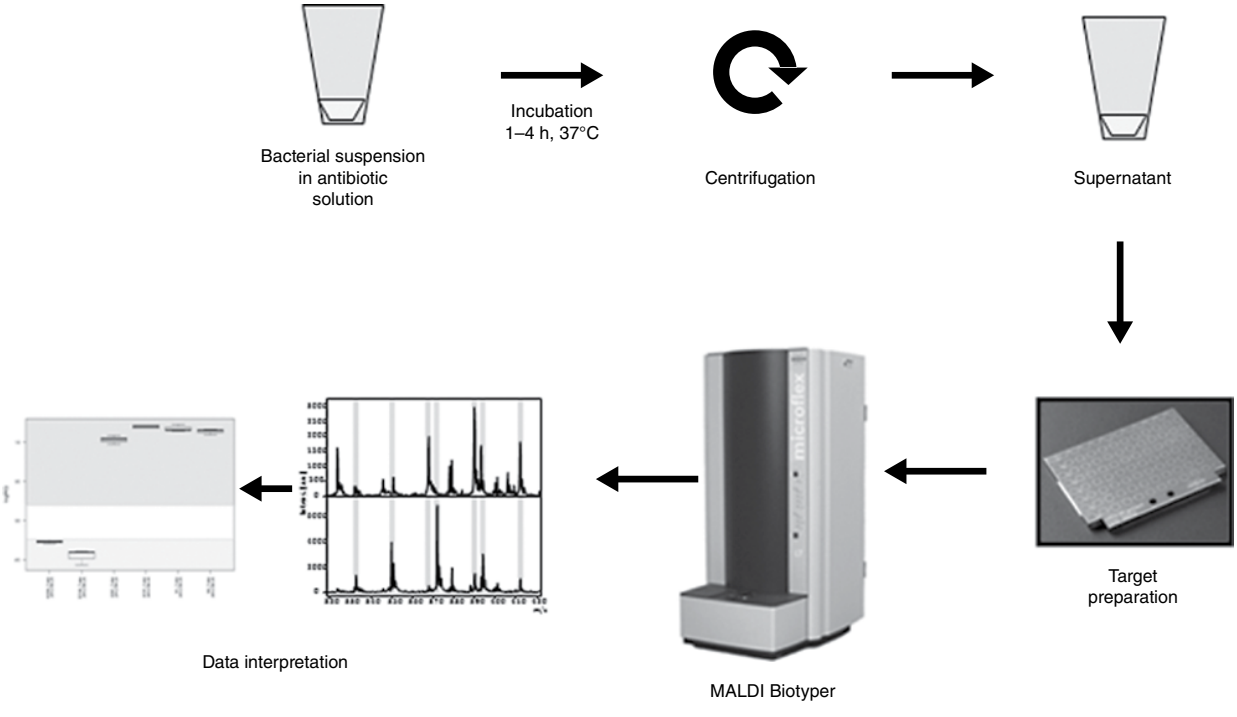


Figure 10.5 Workflow of the ß-lactam hydrolysis assay as proposed in Sparbier *et al.*, 2012.

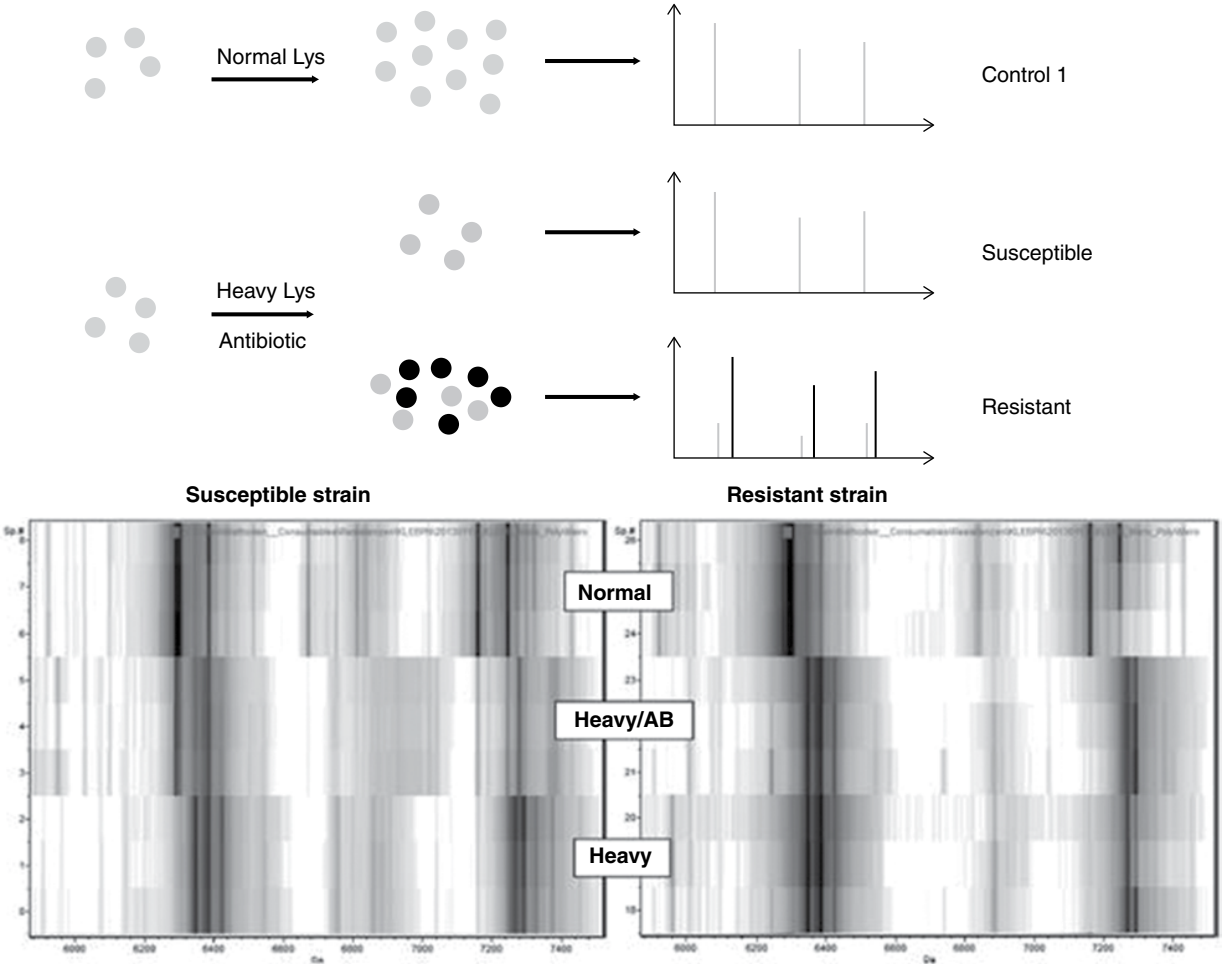


Figure 10.6 Basic principles of the MS-Resist™ assay using broth with heavy non-radioactive lysine as metabolic marker (upper panel). An oxacillin-susceptible *S. aureus* strain (left) and an MRSA strain (right) mass spectra (pseudo-gel view) after 3 h incubation (Sparbier *et al.*, 2013). Adapted from Sparbier, K. *et al.* (2013). MALDI Biotyper based rapid resistance detection by stable isotope labeling. *J Clin Microbiol* 51, 3741–3748.

cannot grow in the presence of oxacillin, no main changes in the mass spectra occurred. MRSA strains showed no growth inhibition with cefoxitin or oxacillin; the peaks were similar between the heavy isotope medium and the heavy isotope antibiotic approach. The time-to-result for a conventional antibiotic susceptibility test is at least 8–10 h, so less than one third of the time is needed with the new assay. As already mentioned, rapid testing is the precondition for appropriate therapy and effective antibiotic stewardship (Perez *et al.*, 2013).

10.8 Multi-Resistant *Pseudomonas aeruginosa*

Trecarichi and Tumbarello (2014) recently emphasized the important role of *Pseudomonas aeruginosa*, especially multi-resistant strains, in neutropenic patients. *Pseudomonas* spp. is the second most common isolated microbe in these patients and is linked to high mortality, especially when bloodstream infections occur. Jung *et al.* (2014) investigated the resistance of *Pseudomonas aeruginosa* using the new and rapid mass spectrometry technique based on stable isotopes. They were able to test 30 different pseudomonad strains with three different key antibiotics: ciprofloxacin, tobramycin and meropenem. All the strains showed good correspondence between the susceptibility interpretation of conventional test methods and the MS-Resist protocol. Kostrzewa *et al.* (2013) used a similar approach for the rapid detection of carbapenem resistance in *Enterobacteriaceae*. Heavy-lysine-labelled and -unlabelled *Klebsiella pneumoniae* strains were exposed to meropenem for only 2 h. The following mass spectrometry analysis could clearly differentiate between susceptible strains and CREs. Therefore, this new assay seems to be a promising tool for a multiplicity of antibiotic substances, providing both rapid and reliable results.

10.9 MALDI Biotyper Antibiotic Susceptibility Test Rapid Assay (MBT-ASTRA™)

A slight disadvantage of the MS-Resist assay is the high cost of the labelled lysine containing incubation broth and the relatively long hands-on-time. Therefore, the inventers of this test searched for a simpler and cheaper MS-based antibiotic susceptibility test. The new approach, the MALDI Biotyper antibiotic susceptibility test rapid assay (MBT-ASTRA), does not use stable amino acids but a quantitative MALDI-TOF MS approach. A standard broth (e.g. brain heart infusion broth) can be used, and some bacterial taxa only require 1 h incubation time. The cells are spiked with an internal standard, lysed and prepared for the mass spectrometer. The sample was normalized to the highest peak of the standard and the amount of peptides and proteins detected by the MALDI correlated with the bacterial growth. If the tested bacteria are resistant to the antibiotic of choice, the mass peaks are similar to the growth control. If they are susceptible, none or only a few peaks are seen. Lange *et al.* (2014) demonstrated this with different Klebsiella strains, focusing on meropenem as the antibiotic target (Figure 10.7). With this assay, antibiotic susceptibility testing by mass spectrometry reached a turning point, because now not only could the terms *susceptible* and *resistant*

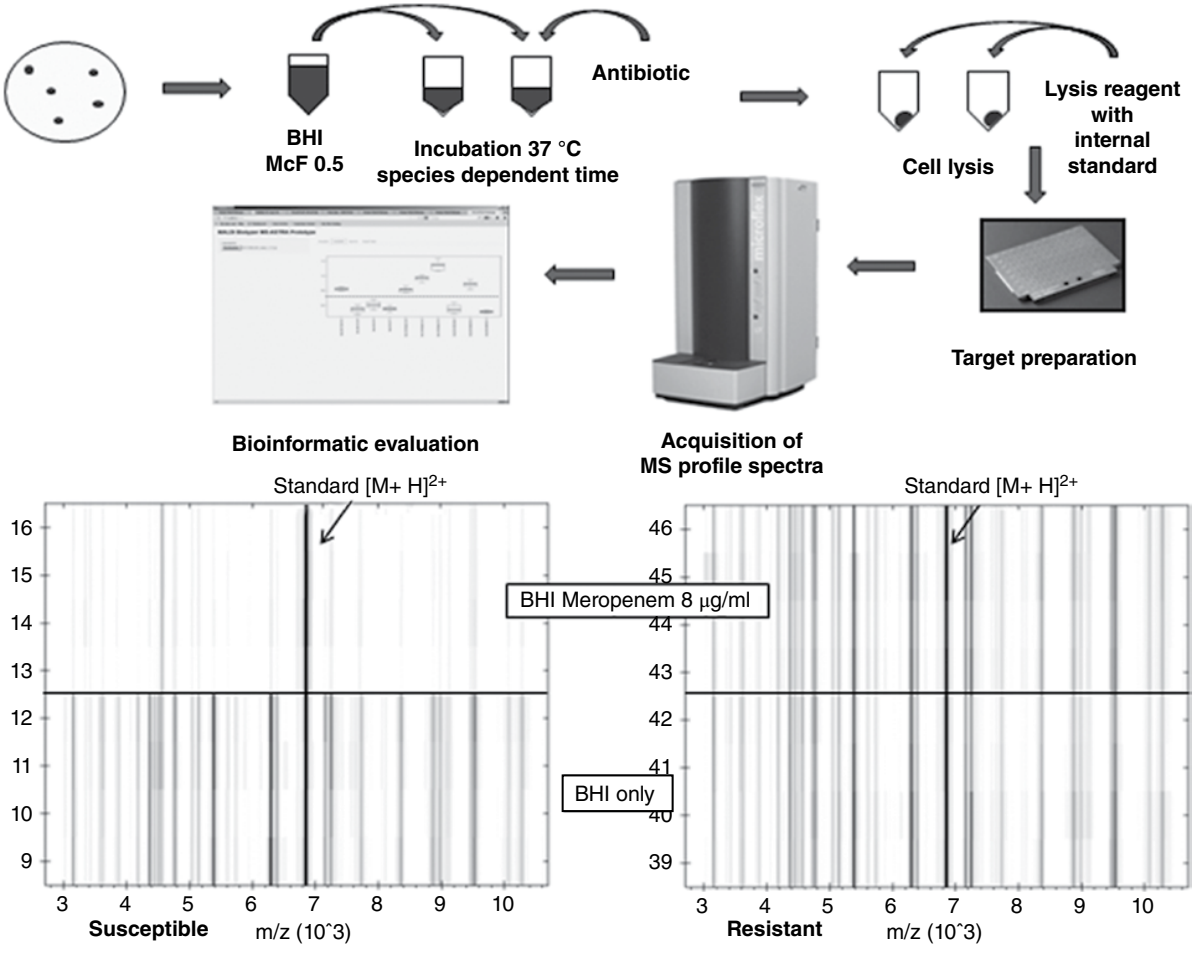


Figure 10.7 Workflow of the MBT STAR-ASTRA™ assay (upper panel) and examples with a meropenem-susceptible *K. pneumoniae* strain (left lower panel) and a resistant KPC strain (right lower panel; pseudo-gel view) as described in (Sparbier *et al.*, 2016).

be predicted, but a minimal inhibitory concentration could be determined. The method also worked directly from positive blood cultures with reliable and accurate results. One of the studies, which had proved the direct identification of bacteria and yeast by MALDI-TOF MS, demonstrated, in addition, that a significant percentage of blood cultures in septicaemia turned positive within 12 h and more than 90% of the relevant samples became positive within a day (Klein *et al.*, 2011). Using mass spectrometry in sepsis diagnostics implies that identification and antibiotic susceptibility results could be obtained in less than a day. Today no sepsis study has been published using these new promising methods. Taking into account Kumar's well-known septicaemia study (Kumar *et al.*, 2009) concerning the importance of early and appropriate antibiotic therapy in combination with the benefits that Perez *et al.* (2013) could demonstrate with the Sepsityper™ approach, it can be predicted that a significant improvement of clinical outcomes in septicaemia cases will be achieved in the near future. To reach this goal, it is necessary to show that not only can carbapenem susceptibility be determined but also other antibiotic classes of choice. Sparbier and her team recently demonstrated that all groups of beta-lactams (penicillins, cephalosporins and carbapenems), fluoroquinolones and aminoglycosides can be used with accurate and reliable results in this MALDI Biotyper antibiotic susceptibility test rapid assay (MBT-ASTRA). Preliminary results suggest slight advantages for bacteriostatic antibiotics when using the MS-Resist technology, whereas bactericidal antimicrobials give clearer results in the MBT-ASTRA (personal communication).

10.10 The Potential Use of Mass Spectrometry for Antibiotic Testing in Yeast

Yeasts are a major cause of septicaemia with 8% to 10% of nosocomial bloodstream infections being attributable to Candida, with candidemia increasing mortality rates between 20% and 49% (Koehler *et al.*, 2014). Direct identification of yeasts from positive blood culture by mass spectrometry was already described in 2010 (Ferroni *et al.*, 2010; see Chapter 9).

Marinach analyzed the response of *Candida albicans* to fluconazole and demonstrated significant alterations in the mass spectra that are dependent on the concentration. He defined the minimal profile change concentration (MPCC) as an endpoint for his determination of a yeast strain collection. Comparison of the MPCCs with an MIC determination according to the CLSI method revealed a 94% concordance (within +/– 1 drug dilution values), indicating accurate results (Marinach *et al.*, 2009). Echinocandins are currently the drug of choice for Candida bloodstream infection. Vella *et al.* (2013) investigated a MALDI-based method for the antifungal susceptibility test (AFST) to caspofungin, the most common drug of this class. He demonstrated reliable agreement with the CLSI reference method within 6 h assay time compared to a 20–24 h incubation of the conventional test. Kostrzewa *et al.* (2013) confirmed both results, highlighting the speed of these analyses, but also had some reservations, i.e., that the reproducibility and robustness of this method still needs to be proved under routine conditions in different laboratories.

The detection of antimicrobial resistances in bacteria and yeasts is a hallmark of clinical microbiological diagnostics. Though time consuming, culture-based methods for

antibiotic resistance testing are still the gold standard in diagnostic laboratories around the world. The introduction of MALDI-TOF MS in medical microbiology has already dramatically changed the approach to the identification of bacteria and fungi. Now that these first steps have been taken to exploit MALDI-TOF MS for the rapid detection of antibiotic resistances, this will no doubt encourage further imminent developments in the field.

References

Aldridge, K. E., Ashcraft, D., O'Brien, M., and Sanders, C. V. (2003). Bacteremia due to *Bacteroides fragilis* group: Distribution of species, beta-lactamase production, and antimicrobial susceptibility patterns. *Antimicrob Agents Chemother* 47, 148–153.

Bernardo, K., Pakulat, N., Macht, M., Krut, O., Seifert, H., Fleer, S., Hunger, F., and Kronke, M. (2002). Identification and discrimination of *Staphylococcus aureus* strains using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Proteomics* 2, 747–753.

Bethel, C. R., Distler, A. M., Ruszczycky, M. W., Carey, M. P., Carey, P. R., Hujer, A. M., Taracila, M., Helfand, M. S., Thomson, J. M., Kalp, M. *et al.* (2008). Inhibition of OXA-1 {beta}-Lactamase by Penems. *Antimicrob Agents Chemother* 52, 3135–3143.

Burckhardt, I., Pauker, V., Bode, K., and Zimmermann, S. (2013). Detecting Aminoglycoside Resistance by Mass Spectrometry. Paper presented at: *23rd ECCMID 2013* (Berlin: ESCMID).

Burckhardt, I., and Zimmermann, S. (2011). Using MALDI-TOF mass spectrometry to detect Carbapenem resistance within one to two and a half hours. *J Clin Microbiol* 49, 3321–3324.

Chatterjee, S. S., Chen, L., Joo, H. S., Cheung, G. Y., Kreiswirth, B. N., and Otto, M. (2011). Distribution and regulation of the mobile genetic element-encoded phenol-soluble modulin PSM-mec in methicillin-resistant *Staphylococcus aureus*. *PLoS ONE* 6, e28781.

Demirev, P. A., Hagan, N. S., Antoine, M. D., Lin, J. S., and Feldman, A. B. (2013). Establishing drug resistance in microorganisms by mass spectrometry. *J Am Soc Mass Spectrom* 24, 1194–1201.

Du, Z., Yang, R., Guo, Z., Song, Y., and Wang, J. (2002). Identification of *Staphylococcus aureus* and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Chem* 74, 5487–5491.

Edwards-Jones, V., Claydon, M. A., Evason, D. J., Walker, J., Fox, A. J., and Gordon, D. B. (2000). Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. *J Med Microbiol* 49, 295–300.

Fenyvesi, V. S., Urban, E., Bartha, N., Abrok, M., Kostrzewa, M., Nagy, E., Minarovits, J., and Soki, J. (2014). Use of MALDI-TOF/MS for routine detection of cfiA gene-positive *Bacteroides fragilis* strains. *Int J Antimicrob Agents* 44, 474–475.

Ferroni, A., Suarez, S., Beretti, J. L., Dauphin, B., Bille, E., Meyer, J., Bougnoux, M. E., Alanio, A., Berche, P., and Nassif, X. (2010). Real-time identification of bacteria and Candida species in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* 48, 1542–1548.

Green, K. D., Chen, W., Houghton, J. L., Fridman, M., and Garneau-Tsodikova, S. (2010). Exploring the substrate promiscuity of drug-modifying enzymes for the chemoenzymatic generation of N-acylated aminoglycosides. *Chembiochem* 11, 119–126.

Griffin, P. M., Price, G. R., Schooneveldt, J. M., Schlebusch, S., Tilse, M. H., Urbanski, T., Hamilton, B., and Venter, D. (2012). Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *J Clin Microbiol* 50, 2918–2931.

Hrabak, J., Chudackova, E., and Papagiannitsis, C. C. (2014). Detection of carbapenemases in Enterobacteriaceae: A challenge for diagnostic microbiological laboratories. *Clin Microbiol Infect* 20, 839–853.

Hrabak, J., Chudackova, E., and Walkova, R. (2013). Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for detection of antibiotic resistance mechanisms: From research to routine diagnosis. *Clin Microbiol Rev* 26, 103–114.

Hrabak, J., Walkova, R., Studentova, V., Chudackova, E., and Bergerova, T. (2011). Carbapenemase activity detection by matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry. *J Clin Microbiol* 49, 3222–3227.

Johansson, A., Nagy, E., and Soki, J. (2014). Instant screening and verification of carbapenemase activity in *Bacteroides fragilis* in positive blood culture, using matrix-assisted laser desorption ionization – time of flight mass spectrometry. *J Med Microbiol* 63, 1105–1110.

Josten, M., Dischinger, J., Szekat, C., Reif, M., Al-Sabti, N., Sahl, H. G., Parcina, M., Bekeredjian-Ding, I., and Bierbaum, G. (2014). Identification of agr-positive methicillin-resistant *Staphylococcus aureus* harbouring the class A mec complex by MALDI-TOF mass spectrometry. *Int J Med Microbiol* 304, 1018–1023.

Jung, J. S., Eberl, T., Sparbier, K., Lange, C., Kostrzewa, M., Schubert, S., and Wieser, A. (2014). Rapid detection of antibiotic resistance based on mass spectrometry and stable isotopes. *Eur J Clin Microbiol Infect Dis* 33, 949–955.

Klein, S., Zimmermann, S., Koehler, C., Mischnik, A., Alle, W., and Bode, K. (2011). Integration of MALDI-TOF mass spectrometry in blood-culture diagnostic. *A fast and effective approach. J Med Microbiol* 61, 323–331.

Koehler, P., Tacke, D., and Cornely, O. A. (2014). Our 2014 approach to candidaemia. *Mycoses* 57, 581–583.

Kostrzewa, M., Sparbier, K., Maier, T., and Schubert, S. (2013). MALDI-TOF MS: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. *Proteomics Clin Appl* 7, 767–778.

Kumar, A., Ellis, P., Arabi, Y., Roberts, D., Light, B., Parrillo, J. E., Dodek, P., Wood, G., Kumar, A., Simon, D. *et al.* (2009). Initiation of inappropriate antimicrobial therapy results in a fivefold reduction of survival in human septic shock. *Chest* 136, 1237–1248.

Kumarasamy, K. K., Toleman, M. A., Walsh, T. R., Bagaria, J., Butt, F., Balakrishnan, R., Chaudhary, U., Doumith, M., Giske, C. G., Irfan, S. *et al.* (2010). Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: A molecular, biological, and epidemiological study. *Lancet Infect Dis* 10, 597–602.

Lange, C., Schubert, S., Jung, J., Kostrzewa, M., and Sparbier, K. (2014). Quantitative MALDI-TOF MS for rapid resistance detection. *J Clin Microbiol* 52, 4155–4162.

Lau, A. F., Wang, H., Weingarten, R. A., Drake, S. K., Suffredini, A. F., Garfield, M. K., Chen, Y., Gucek, M., Youn, J. H., Stock, F. *et al.* (2014). A rapid matrix-assisted laser

desorption ionization-time of flight mass spectrometry-based method for single-plasmid tracking in an outbreak of carbapenem-resistant Enterobacteriaceae. *J Clin Microbiol* 52, 2804–2812.

Lefurgy, S. T., Malashkevich, V. N., Aguilan, J. T., Nieves, E., Mundorff, E. C., Biju, B., Noel, M. A., Toro, R., Baiwir, D., Papp-Wallace, K. M. *et al.* (2015). Analysis of the Structure and Function of FOX-4 Cephamycinase. *Antimicrob Agents Chemother* 60, 717–728.

Livermore, D. M. (2012). Fourteen years in resistance. *Int J Antimicrob Agents* 39, 283–294.

Majcherczyk, P. A., McKenna, T., Moreillon, P., and Vaudaux, P. (2006). The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus aureus*. *FEMS Microbiol Lett* 255, 233–239.

Marinach, C., Alanio, A., Palous, M., Kwasek, S., Fekkar, A., Brossas, J. Y., Brun, S., Snounou, G., Hennequin, C., Sanglard, D. *et al.* (2009). MALDI-TOF MS-based drug susceptibility testing of pathogens: The example of *Candida albicans* and fluconazole. *Proteomics* 9, 4627–4631.

Nagy, E., Becker, S., Soki, J., Urban, E., and Kostrzewa, M. (2011). Differentiation of division I (cfiA-negative) and division II (cfiA-positive) *Bacteroides fragilis* strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Med Microbiol* 60, 1584–1590.

Nordmann, P., Naas, T., and Poirel, L. (2011). Global spread of Carbapenemase-producing Enterobacteriaceae. *Emerg Infect Dis* 17, 1791–1798.

Papagiannitsis, C. C., Studentova, V., Izdebski, R., Oikonomou, O., Pfeifer, Y., Petinaki, E., and Hrabak, J. (2015). Matrix-assisted laser desorption ionization-time of flight mass spectrometry meropenem hydrolysis assay with NH4HCO3, a reliable tool for direct detection of carbapenemase activity. *J Clin Microbiol* 53, 1731–1735.

Perez, K. K., Olsen, R. J., Musick, W. L., Cernoch, P. L., Davis, J. R., Land, G. A., Peterson, L. E., and Musser, J. M. (2013). Integrating rapid pathogen identification and antimicrobial stewardship significantly decreases hospital costs. *Arch Pathol Lab Med* 137, 1247–1254.

Rebelo, M., Pereira, B., Lima, J., Decq-Mota, J., Vieira, J. D., and Costa, J. N. (2011). Predictors of in-hospital mortality in elderly patients with bacteraemia admitted to an Internal Medicine ward. *Int Arch Med* 4, 33.

Sparbier, K., Lange, C., Jung, J., Wieser, A., Schubert, S., and Kostrzewa, M. (2013). MALDI Biotyper based rapid resistance detection by stable isotope labeling. *J Clin Microbiol* 51, 3741–3748.

Sparbier, K., Schubert, S., and Kostrzewa, M. (2016). MBT-ASTRA: A suitable tool for fast antibiotic susceptibility testing? *Methods*.

Sparbier, K., Schubert, S., Weller, U., Boogen, C., and Kostrzewa, M. (2012). MALDI-TOF MS based functional assay for the rapid detection of resistance against ss-lactam antibiotics. *J Clin Microbiol* 50, 927–937.

Szabados, F., Kaase, M., Anders, A., and Gatermann, S. G. (2012). Identical MALDI TOF MS-derived peak profiles in a pair of isogenic SCCmec-harboring and SCCmec-lacking strains of *Staphylococcus aureus*. *J Infect* 65, 400–405.

Trecarichi, E. M., and Tumbarello, M. (2014). Antimicrobial-resistant Gram-negative bacteria in febrile neutropenic patients with cancer: Current epidemiology and clinical impact. *Curr Opin Infect Dis* 27, 200–210.

Tzouvelekis, L. S., Markogiannakis, A., Psichogiou, M., Tassios, P. T., and Daikos, G. L. (2012). Carbapenemases in *Klebsiella pneumoniae* and other Enterobacteriaceae: An evolving crisis of global dimensions. *Clin Microbiol Rev* 25, 682–707.

Vella, A., De, C. E., Vaccaro, L., Posteraro, P., Perlin, D. S., Kostrzewa, M., Posteraro, B., and Sanguinetti, M. (2013). Rapid antifungal susceptibility testing by matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis. *J Clin Microbiol* 51, 2964–2969.

Wendt, C., Schutt, S., Dalpke, A. H., Konrad, M., Mieth, M., Trierweiler-Hauke, B., Weigand, M. A., Zimmermann, S., Biehler, K., and Jonas, D. (2010). First outbreak of *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* in Germany. *Eur J Clin Microbiol Infect Dis* 29, 563–570.

Youn, J. H., Drake, S. K., Weingarten, R. A., Frank, K. M., Dekker, J. P., and Lau, A. F. (2016). Clinical performance of a matrix-assisted laser desorption ionization-time of flight mass spectrometry method for detection of certain blaKPC-containing plasmids. *J Clin Microbiol* 54, 35–42.

11

Discrimination of *Burkholderia* Species, *Brucella* Biovars, *Francisella tularensis* and Other Taxa at the Subspecies Level by MALDI-TOF Mass Spectrometry

Axel Karger

Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Greifswald – Insel Riems, Germany

11.1 Introduction

Identification of bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MADLI-TOF MS) has revolutionized bacterial diagnostics for several reasons (Seng *et al.*, 2009). The technique is rapid, high-throughput, sample costs are low and the information content of a spectrum is high. In addition, samples can be prepared from inactivated organisms and, therefore, tube testing with living pathogens can be avoided. As with the introduction of any new technology, the question is, how well can 'MALDI identification' (MI) perform in areas that are problematic. In a number of studies, MI of bacteria has been compared with classical bacteriological methods to demonstrate its usefulness as a diagnostic tool in clinical environments. Usually, the ability of MI to identify species with high accuracy has been confirmed. However, in cases where closely related species were to be discriminated, or subspecies-level discrimination was attempted, misidentification rates were higher or variations of the standard procedures were necessary to improve the accuracy of diagnosis. In the following, strategies to enable or to improve MALDI-TOF-MS-based identification of closely related organisms will be discussed.

11.2 Principles of MALDI-TOF MS-Based Identification of Bacteria

One of the main features of MI is simplicity. The standard procedure is very straightforward and comprises only a few steps. The bacterium is cultured on a solid medium to the size of a visible colony which, in the most simple variation of the workflow, is directly deposited onto a MALDI sample target (Bizzini *et al.*, 2010; McElvania *et al.*, 2013), overlaid with matrix and the spectrum is acquired. In most studies, however, the sample is prepared from the colony with a simple extraction protocol that improves the quality of the spectra (Alatoom *et al.*, 2011; Christner *et al.*, 2014). In a frequently used protocol (Sauer *et al.*, 2008), bacteria are precipitated from an aqueous suspension by addition of ethanol and then extracted with formic acid/acetonitrile. Extraction can also be performed on-target to simplify the protocol (Matsuda *et al.*, 2012; McElvania *et al.*, 2013; Schmitt

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

et al., 2013). After addition of a MALDI matrix, usually α-Cyano-4-hydroxycinnamic acid (HCCA), spectra are acquired in a mass range between 2,000 and 20,000 Da in the linear mode. Simple MALDI-TOF mass spectrometers offer sufficient resolution in this mass range so that there is no need for highly sophisticated and costly equipment. The species of the unknown sample is then identified by comparison of the spectrum with a library of reference spectra by statistical means. Similarity between the sample and reference spectra are calculated by pattern matching approaches (Jarman and Wahl, 2005) that are based on the whole spectrum or on extracted peak lists, using metric or non-metric distances measures, correlation coefficients or empirical similarity scores. The reference spectrum that is most similar to the sample spectrum will be taken as the result of the identification procedure. Most notably, there is no need for any prior information on the sample such as a tentative taxonomic classification. Moreover, MI does not require any sample-specific reagents such as antibodies, PCR primers or nucleic acid probes, but will be practicable in any laboratory that has access to a MALDI-TOF mass spectrometer and to a reference database that includes the spectrum of the organism of interest.

The usefulness of MI for genus and species identification has been proved in numerous studies in which the performance of MI in routine microbiological laboratories was investigated (Bizzini *et al.*, 2010; Luo *et al.*, 2015; Martiny *et al.*, 2014; Patel, 2013; Seng *et al.*, 2013). This is the basis for the implementation of MI in diagnostic laboratories worldwide and the increasing acceptance of this technology.

11.3 Generality versus Specificity

The properties of a MALDI-TOF mass spectrum, for example, of an unknown bacterium, will depend on a large number of parameters. Several factors such as culture conditions (Anderson *et al.*, 2012; Balazova *et al.*, 2014; Karger *et al.*, 2013; Sedo *et al.*, 2013; Veloo *et al.*, 2014), the extraction protocol (Alatoom *et al.*, 2011; Fournier *et al.*, 2012; Khot *et al.*, 2012; Meetani and Voorhees, 2005; Schulthess *et al.*, 2013; Veloo *et al.*, 2014), the use of different matrices (Dieckmann *et al.*, 2008; Meetani and Voorhees, 2005; Paauw *et al.*, 2014), and so on, have a noticeable influence on the reproducibility of the mass spectrum. As MI relies on pattern recognition algorithms, the experimental conditions for the construction of the reference databases must be tightly controlled and must match the protocols that are used for the preparation of the unknown samples (Wilen *et al.*, 2015). This is even more important with respect to the discrimination of closely related organisms, which may rely on very subtle differences between the spectra.

Although it is desirable to restrict the number of 'standard' preparation protocols, a certain degree of variation is inevitable in order to meet the requirements of the different bacteria. Not all bacteria will grow to sufficiently large colonies on the same medium within the same time, and some, such as mycobacteria, have to be extracted with special protocols in order to obtain spectra of sufficient quality (Mather *et al.*, 2014; Wilen *et al.*, 2015, see Chapter 4).

If standard procedures fail to resolve spectra from closely related organisms, it is straightforward to take advantage of the flexibility MALDI-TOF MS and to optimize the experimental conditions with respect to the properties of the samples. In this way, discrimination of problematic organisms can often be improved or achieved in the first place, albeit at a certain cost. Highly sample-specific experimental conditions may require dedicated databases and the implementation of tailor-made algorithms for the

identification procedure, which may reduce the portability of the approach to other laboratories. Thus, the higher discriminatory power, to a certain extent, is traded in for simplicity and velocity, which are two hallmarks of MI.

In numerous studies, more or less drastic variations of the 'standard' sample preparation protocols were introduced to improve identification of closely related organisms by MALDI-TOF MS. For example, Schmidt *et al.* (2009) devised a workflow termed 'shotgun mass mapping' (SMM) for the discrimination of Lactobacillus subspecies. The major change in comparison to conventional protocols was that the bacterial colonies were digested with trypsin, and the resulting peptide mixture was measured with a MALDI-TOF mass spectrometer that was operated in the reflector mode. In this way, the low mass range (900–4000 Da) could be analyzed with high precision, and the application of a similarity search approach allowed the discrimination of Lactobacillus subspecies with excellent resolution. Krasny *et al.* (2014) have applied trypsin digestion for the differentiation of Cronobacter subspecies, albeit by analysis of a preparation of cytoplasmic proteins, using the linear mode of the instrument in a mass range from 1 to 15 kDa and MALDI Biotyper (Bruker) for the evaluation. Likewise, extension of the mass range to higher values has been shown to be useful, if appropriate matrices such as ferulic acid or Sinapinic acid were used. This approach was successful, for example, for the discrimination of differentially pathogenic *Vibrio cholera* strains (Paauw *et al.,* 2014), or for the identification of Salmonella subspecies (Dieckmann *et al.*, 2008). Combining ferulic acid with nonionic detergents, Meetani and Voorhees (2005) were able to expand the useful mass range up to 140 kDa. These examples demonstrate the versatility of MALDI-TOF-MS-based identification of bacteria and the excellent analytical performance that can be achieved by adaptation of the experimental conditions.

On the other hand, a number of studies have shown that subspecies-level identification can also be accomplished on the basis of spectra that resulted from conventional MI procedures if only the statistical evaluation was refined. This approach seems equally attractive, as the final identification of the sample is attempted by additional calculations which are appended to the routine procedure for MS-based diagnosis. In this way, spectra of a first-line diagnosis can be used for further evaluations, and delays due to an additional round of culturing, extraction and spectrum acquisition are avoided. Available general-purpose databases can be used, for example, to select appropriate sets of reference spectra, and, vice versa, new spectra of field isolates can be readily be integrated into existing reference databases.

In the following, successful approaches for the identification of closely related organisms by refined statistical analysis of their MALDI-TOF mass spectra will be described and discussed. The emphasis will be on the discrimination of select organisms such as subspecies of *Francisella tularensis*, different species of the Brucella genus, *Burkholderia mallei* and *Burkholderia pseudomallei* and Shiga-toxin-producing serotypes of *Escherichia coli*.

11.4 Shigatoxin-Producing and Enterohemorrhagic *Escherichia coli* (STEC and EHEC)

STEC are a group of zoonotic enteric pathogens that occur in some serotypes of *E. coli* (Nataro and Kaper, 1998). Infections of humans with certain STEC strains can lead to hemorrhagic diarrhoea, hence the designation EHEC for this subgroup. In more severe

infections, a haemolytic-uremic syndrome (HUS) can evolve which is potentially life-threatening. In May 2011, occurrences of illness with HUS and bloody diarrhoea were reported in Germany that was caused by an infection with EHEC of the serotype O104:H4. Before the source of the pathogen was clearly identified, the disease had spread from its origin in Hamburg to nearly all other German federal states and to France. Overall, 3842 cases of illness were reported (855 cases of HUS and 2987 cases of acute gastroenteritis), and 53 patients died of the infection (Robert-Koch-Institut, 2011).

Isolates that had been collected from affected individuals in the course of this outbreak were analyzed in a thorough study by Christner *et al.* (Christ, 2014). Protein profiles of outbreak-related isolates were compared to *E. coli* isolates that had been collected before the 2011 EHEC outbreak. By statistical analysis, two marker ions could be derived that allowed the reliable discrimination of spectra from outbreak-related and pre-outbreak isolates. On the basis of these masses, the classification into outbreak and non-outbreak strains was correct for 99.7% of all samples that had been prepared by formic acid extraction. Remarkably, classification that was based on spectra measured after direct deposition of the sample on the target was only slightly less accurate (99%), demonstrating the robustness of MI and its potential in an epidemic scenario.

The authors also evaluated the classification results that were obtained on the basis of whole spectrum comparison, which is a more likely procedure for ad-hoc diagnosis in the case of an acute epidemic when serotype-specific or strain-specific markers may not have been developed yet. With accuracies of up to 98%, classification was still very good if binary distance measures had been used for the calculation of the similarity of the spectra. Rates of correct identifications slightly dropped if unweighted metric distance measures, for example, the Euclidean distance, were used. This study is not only an excellent example for the potential usefulness of MI in a scenario of an acute outbreak, but also demonstrates the high precision that can be achieved for the subspecies-level differentiation of *E. coli* isolates if MI is used together with a dedicated identification algorithm that is based on reliable marker ions.

Unfortunately, prominent and reliable group-specific markers do not occur in every serotype of *E. coli*, and the search for specific markers and the implementation of classification algorithms will become more complicated as more different serotypes are to be differentiated, and more markers may have to be considered. Much work has focused on the discrimination of other serotypes that are associated with human disease, namely, O157, O26, O103, O91, O145, O128 and O111. Of these, O157 is considered the most virulent serotype aside from O104 (Preussel, Hohle, Stark, & Werber, 2013).

In an early study, Bright and colleagues (Bright *et al.*, 2002) had already observed that the O157 antigen had a strong influence on the MALDI-TOF spectra. Analysis of the spectrum-based distances showed that the O157 isolates separated well from a panel of *E. coli* strains carrying other O-antigens. In a screen for O157:H7 serotype-specific features of MALDI-TOF mass spectra, Mazzeo *et al.* (2006) could specify a mass of 9060 Da as a negative marker for this serotype. The protein behind the mass was later identified as the acid stress chaperone-like protein (HdeB) (Fagerquist *et al.*, 2010). In a recent publication, Ojima-Kato and colleagues have presented an algorithm for the discrimination of O157, O26 and O111 serotypes which is based on four markers that were identified as the ribosomal proteins S15 and L25, the DNA-binding protein H-NS and HdeB (Ojima-Kato *et al.*, 2014). The classification scheme, a simple decision tree, accurately predicted the serotypes of 83 strains that were blind-tested to validate the model.

Data reduction by feature selection has also been a powerful strategy to improve the differentiation of *E. coli* serotypes O26, O156 and O165 (Karger *et al.*, 2011). These serotypes are less frequently associated with human disease than, for example, O157:H7 (Bielaszewska *et al.*, 2007; Mellmann *et al.*, 2008), but can still cause severe infections in humans. Ruminants, especially beef herds, are large reservoirs for EHEC isolates of these serotypes and therefore, their identification and discrimination are of interest not only in veterinary medicine. As the search for single serotype-identifying markers failed for these serotypes, more sophisticated algorithms for the classification had to be developed. In this case, feature selection was performed by systematic exclusion of masses that were either not reliably detected within one serotype or that did not contribute to group discrimination as they were equally present in all three groups. The first parameter was expressed by the relative frequency of occurrence of a mass within a group, the second as the p-value of a Fisher's t-test that was performed for every mass across the serotypes. After gradual removal of masses that were not highly reproducible or did not contribute to group discrimination, the remaining masses were used to calculate synthetic 'prototype' spectra, which were then used for classification. Under optimized conditions, the accuracy of the classification was 99.3%, compared to only 69% when no mass selection had been performed for the construction of the prototype spectrum. Interestingly, very stringent removal of masses that were not highly reproducible or not highly group-specific was shown to be counterproductive, as the misclassification rate increased under these conditions. An attempt to use the described approach to discriminate smaller clusters of genetically related isolates within the given serotypes (Geue *et al.*, 2006; Geue *et al.*, 2009; Geue *et al.*, 2010) failed, indicating that group-specific markers or mass signatures may be difficult to find below the serotype level.

11.5 *Francisella tularensis*

Tularemia is a zoonotic disease caused by the bacterium *Francisella tularensis*. Four subspecies have been characterized that differ in virulence and in geographic distribution. Most infections in humans and in animals (frequently hares) are caused by two subspecies, the highly virulent *F. tularensis* ssp. *tularensis* (also designated as type A), which is endemic in North America, and the less virulent *F. tularensis* ssp. *holarctica* (type B), which is found predominantly in Europe. For the discrimination of *F. tularensis* subspecies, PCR-based tests are available that target genetic markers like the variable number of tandem repeats (Johansson *et al.*, 2004) or single-nucleotide polymorphisms (Vogler *et al.*, 2009).

Due to its high virulence, the low infection doses and the possible transmission by aerosol, *F. tularensis* has been listed as a potential 'category A' bioterrorism agent by the Centers for Disease Control and Prevention (CDC) (http://www.bt.cdc.gov/agent/agentlist-category.asp). As in any case of an intended or unintended release of highly hazardous pathogens, there is an urgent need for quick and reliable identification, and MI of select organisms such as *F. tularensis* and others has been studied intensively. In this context, inactivation of the sample is, of course, of paramount importance (Couderc *et al.*, 2012; Cunningham and Patel, 2015; Drevinek *et al.*, 2012; Lasch *et al.*, 2008; Lasch *et al.*, 2015; Tracz *et al.*, 2013). The dilemma with inactivation is that, on one hand, it must be absolute so that any exposure of laboratory personnel can be excluded. On the

other hand, chemical or physical damage or modification of the sample must be avoided as far as possible in order to allow the acquisition of high-quality spectra that are required for reliable identification. Lasch *et al.* (2015) have shown that gamma-irradiation reliably inactivated samples of *F. tularensis* and a broad panel of other highly pathogenic microorganisms. After gamma inactivation, the samples could readily be used for successful MI. However, when spectra of irradiated and trifluoro acidic acid-inactivated (Lasch *et al.*, 2008) samples were compared, a slight loss of quality was observed which came with irradiation, confirming an earlier publication (Tracz *et al.*, 2013). Also, use of oxidizing agents such as NaOCl was shown to cause satellite peaks of +16 Da, most likely induced by the oxidation of amino acids. These results indicate that different inactivation protocols may have subtle and differential effects on the spectra which should be taken into account if reference databases of highly pathogenic microorganisms are constructed.

The potential of mass spectrometry to discriminate subspecies of *F. tularensis* was first demonstrated by Lundquist *et al.* (2005) using SELDI (surface-enhanced laser desorption ionization) MS, an MS technique that is related to MALDI. The SELDI MS approach was elaborated by Seibold *et al.* (2007), who applied 'classification and regression tree' (CART) analysis (Breiman *et al.*, 1984) to the spectra of *F. philomiragia* and the four subspecies of *F. tularensis*. CART is a recursive partitioning tool that uses decision rules to predict an outcome of a classification experiment. In this case, decision rules were based on the intensities of only three masses but still allowed the correct classification of all samples.

The positive results that were obtained with SELDI MS were confirmed in a large MALDI-TOF MS study of francisellae that were analyzed with MALDI Biotyper software (Seibold *et al.*, 2010). The reference library was extended by spectra of *F. philomiragia* and the subspecies of *F. tularensis* (*F. tularensis* ssp. *tularensis*, *F. tularensis* ssp. *holarctica*, *F. tularensis* ssp. *mediasiatica* and *Francisella tularensis* ssp. *novicida*), and the identification experiments were then carried out with blinded samples of 45 field and reference strains. Results were 100% accurate using the scoring algorithm of the MALDI Biotyper software, although cluster analysis did not fully reproduce the phylogenetic relationships of the samples. Additional experiments aimed to challenge the robustness of the identification procedure. Biological replicates were as well tested as different growth media and cultivation periods, and one isolate was passaged 30 times before measurement, but still the species and subspecies identification was correct, underlining the power and robustness of MI for the discrimination of *F. tularensis* subspecies. However, to my knowledge, MI has not been successfully used to characterize *F. tularensis* isolates below the subspecies level. In a phylogeographic study with 52 *F. tularensis* ssp. *holarctica* strains that had been isolated from European brown hares in Germany (Muller *et al.*, 2013), genetic and biochemical traits were used to analyze spatial and temporal relations of these isolates. In parallel, MALDI spectra were measured with the aim to correlate spectral features with genetic or biochemical markers. Within *F. tularensis* ssp. *holarctica*, the discrimination of clades B.I and B.II was of greatest interest, as these differed in erythromycin resistance and were spatially separated roughly into the eastern and western half of Germany. But all efforts to correlate spectral features with erythromycin resistance or any of the 14 genetic markers that had been assessed failed, and, therefore, remained unpublished. However, the authors confirmed the MALDI-TOF-MS-based separation of *F. tularensis* subspecies.

11.6 The Genus Brucella

The genus Brucella is composed of the six classical species *B. abortus*, *B. melitensis*, *B. suis*, *B. canis*, *B. ovis* and *B. neotomae* (Corbel and Brinley-Morgan, 1984), and the four more recently described species *B. pinnipedialis*, *B. ceti*, *B. microti* and *B. inopinata* (Foster *et al.*, 2007; Scholz *et al.*, 2008; Scholz *et al.*, 2010). Brucellosis is a frequent zoonotic infection worldwide (Pappas *et al.*, 2006), and brucellae are classified as 'category B' agents by the CDC (http://www.bt.cdc.gov/agent/agentlist-category.asp). Thus, the prompt diagnosis of brucellosis is highly desirable. At the same time, the high numbers of laboratory infections with brucellae (Yagupsky and Baron, 2005) also suggest that the substitution of conventional microbiological tube testing by a technique that carries less risk for the laboratory staff may be of great benefit. As only four of the Brucella species cause human disease (*B. abortus*, *B. melitensis*, *B. canis* and *B. suis*), the reliable identification of the species is of great importance. If an infection with *B. suis* has been diagnosed, the differentiation of the five known biovars is additionally of interest as they show differential pathogenicity for humans (Godfroid *et al.*, 2011).

Due to the high homologies between the genomes of the brucellae (Verger *et al.*, 1985), the organization of the Brucella genus is being discussed (Moreno *et al.*, 2002; Verger *et al.*, 1985; Whatmore, 2009). Thus, it was not surprising that the investigation of the spectral phenotypes of the ten Brucella species, including different biovars of *B. abortus*, *B. melitensis* and *B. suis*, revealed major disagreements between the spectrum-based distances and the current taxonomic organization of the Brucella genus (Karger *et al.*, 2013). In this study, reference spectra of 104 field isolates and 33 reference and vaccine strains were generated and evaluated with MALDI Biotyper software. Although the overall rate of accurate species identifications was high (90% and 95% in the two participating laboratories), misidentifications accumulated within certain groups of species. Specifically, erroneous identifications occurred between *B. abortus* and *B. melitensis*, between the species from marine mammals, *B. ceti* and *B. pinnipedialis*, and between *B. ceti* and *B. canis*. Individual isolates of *B. suis* were misidentified as *B. canis*, *B. abortus* and as *B. melitensis*. These results could be partly explained by cluster analysis of the spectra on the basis of their Euclidean distances. The spectrum-based distances between some established species were only small (e.g. between *B. canis* and *B. ovis*), and for *B. abortus* and *B. melitensis* a continuum of spectral phenotypes was observed rather than two separate clusters. Surprisingly, spectrum-based distances between the *B. suis* biovars were larger than the distances between some established Brucella species (e.g. *B. abortus* and *B. melitensis* or *B. canis* and *B. ovis*). Moreover, *B. suis* biovars 3 and 4 were closer to *B. ovis* and *B. canis* than to any of the other *B. suis* biovars 1, 2 or 5. Nevertheless, it was possible to establish statistical models using ClinProTools software (Bruker), which resolved most of the misidentifications. Thus, a two-step procedure was required to achieve reliable identification of Brucella species and *B. suis* biovars. First, analysis with MALDI Biotyper allowed the unambiguous identification of *B. microti*, *B. inopinata* and of *B. suis* biovars 1 and 5 (including the identification of the biovar). For MALDI Biotyper results indicating *B. abortus*, *B. melitensis*, *B. canis*, *B. ovis* and *B. suis* biovars 2, 3 and 4, an additional evaluation on the basis of the developed statistical models ensured a correct identification.

The finding that spectrum-based clustering did not reflect the current taxonomy is not only interesting with respect to the organization of the genus Brucella (Moreno *et al.*, 2002; Whatmore, 2009), but it may also explain why the authors failed to find single species-identifying markers for each of the species within this genus. Spectra of some of the species (e.g. *B. canis* and *B. ovis*) probably were too similar to identify species-specific masses, whereas the biovars within *B. suis* might have been too heterogeneous to find a joint peak that would at the same time not be found in any of the other species.

Lista and colleagues (Lista *et al.*, 2011) have refined the MI of species from the Brucella genus by the construction of a MALDI Biotyper reference database that was based on the results of multilocus variable-number tandem repeat analysis (MLVA). A large set of Brucella isolates was genetically characterized by MLVA, cluster analysis was performed on basis of the data and for each of the seventeen MVLA clusters one or two representative isolates were selected to serve as references for the analysis with MALDI Biotyper software. In a test with spectra from 152 Brucella isolates of mostly clinical origin, this reference spectra set was shown to allow highly accurate species identification, as only one of the isolates was misidentified. Thus, careful selection of reference spectra can be useful to improve MI, as will be also shown in the following section on Burkholderia.

11.7 The Genus Burkholderia

Burkholderia mallei and *B. pseudomallei* are closely related zoonotic pathogens that cause glanders and melioidosis, respectively. Both species represent pathovars of a single genomospecies which was divided in two separate species due to their clinical impact and host tropism. Together with another closely related species which exhibits markedly lower pathogenicity, *B. thailandensis*, *B. mallei* and *B. pseudomallei* form the so-called 'Pseudomallei complex'. Aside from the differentiation of the Pseudomallei complex, the discrimination of *B. pseudomallei* from species of the Burkholderia cepacia complex, which are frequently isolated from cystic fibrose patients, is of high clinical relevance. As species differentiation within the burkholderiae with other methods remains difficult (Ho *et al.*, 2011; Lowe *et al.*, 2013), the identification of members of this genus by MALDI-TOF MS has been investigated in a number of studies (Degand *et al.*, 2008; Desai *et al.*, 2012; Fehlberg *et al.*, 2013; Fernandez-Olmos *et al.*, 2012; Inglis, Healy *et al.*,2012; Karger *et al.*, 2012; Lambiase *et al.*, 2013; Miñán *et al.*, 2009; Vanlaere *et al.*, 2008).

It has been noted in the literature that the unequivocal identification of species from the Pseudomallei complex strongly depends on the availability of appropriate reference spectra. Lau and colleagues (Lau *et al.*, 2012) had observed that samples of *B. thailandensis* had been recognized as *B. pseudomallei* using a reference database with only one *B. thailandensis* isolate. Addition of another isolate of *B. thailandensis* to the reference set solved the problem, indicating that the intraspecies variation of *B. thailandensis* may require an adequate representation in the reference database for successful identification. Cunningham and Patel (2013) reported that two *B. pseudomallei* isolates that were analyzed with MALDI Biotyper and a database that lacks *B. pseudomallei* references produced maximum scores of 1.954 and 1.962, respectively, with a *B. thailandensis* reference spectrum. As score values below 2.000 must be interpreted as correct genus identification, but not as correct species identification, the result of the query was

correct, but still the species was not identified for lack of an appropriate reference. Supplementation of the reference database with spectra from *B. pseudomallei* and *B. mallei* resulted in scores above the level for species identification; however, the scores that were obtained with reference spectra from both species were so similar that an unambiguous identification was still not possible. Similar observations had been made in a study that focused on the potential of MI to discriminate *B. mallei* and *B. pseudomallei* isolates (Karger *et al.*, 2012). Here, MALDI Biotyper was used with a custom reference spectrum set of 10 *B. mallei* and 17 *B. pseudomallei* isolates. It was noticed that the *B. pseudomallei* isolates exhibited a broader variation of spectral phenotypes than *B. mallei*, which is in agreement with genetic data (Godoy *et al.*, 2003). Within the constructed reference database, correct species identification of all *B. mallei* and *B. pseudomallei* isolates was possible, but the highest score values that some *B. pseudomallei* isolates produced with other representatives of their own species only very slightly exceeded the maximum score values they produced with representatives of *B. mallei*. Moreover, all *B. pseudomallei* isolates produced scores with *B. mallei* isolates that well exceeded the cut-off value of 2.000 that is recommended for reliable species identification, and vice versa, underlining again the close relatedness of these species

When the custom reference database was challenged with isolates that had been prepared in a second laboratory, some misidentifications occurred. Analyzing the MALDI Biotyper results in detail showed that misidentifications were caused by a restricted number of reference spectra. Systematic variation of the reference database allowed the definition of a smaller set of references with only two *B. mallei* and three *B. pseudomallei* isolates which produced correct results for the set of isolates that was tested. It is noteworthy that one of the type strains, *B. pseudomallei* ATCC23343, reproducibly exhibited two striking peak series with mass increments of 14 Da that were found in no other Burkholderia isolate and most likely represented proteins that had been modified by extensive methylation. Spectrum-based cluster analysis showed that *B. pseudomallei* ATCC23343 was an outlier in its own species and, therefore, not included in the reduced reference set.

In this study, two single masses were identified that are useful for the discrimination of certain groups within the Burkholderia genus. Specifically, a marker at 9713 Da was present in all isolates of the Pseudomallei complex (*B. mallei*, *B. pseudomallei* and *B. thailandensis*) but absent in the six other Burkholderia species (*B. ambifaria*, *B. cenocepacia*, *B. dolosa*, *B. glathei*, *B. multivorans* and *B. stabilis*) that had been analyzed in parallel. A mass of 6551 Da discriminated *B. mallei* and *B. pseudomallei* isolates from all other species, including *B. thailandensis*. Both masses have recently been confirmed in an independent study (Niyompanich *et al.*, 2014). Systematic qualitative differences between the spectra of *B. mallei* and *B. pseudomallei* were not observed. However, intensities of masses 5794 and 7553 differed significantly between both species and thus can be used for discrimination.

11.8 Studying Closely Related Organisms by MALDI-TOF MS

There is no general guide to the successful implementation of a procedure for the discrimination of closely related organisms on basis of MALDI-TOF MS. The success of any discrimination procedure will always depend on the specific conditions like the

presence of prominent group-specific features, the number of isolates that is available, the number of groups that are to be discriminated in parallel, the quality of the spectra, and more factors. The following considerations are neither meant to give comprehensive and detailed guidance, nor to replace a textbook on statistics, but rather to summarize our experience and the experiences of others in this field over the past years and to indicate some pitfalls that we have encountered.

11.8.1 Sample Selection

As with the design of any diagnostic test, some basic preconditions must be met before a MALDI-TOF MS study with closely related organisms can be considered. For the statistical evaluation, a sufficient number of isolates within every group (e.g. subspecies, biovar, serotype) is required. All samples that are included in the study should have been assigned to one of the groups using an accepted method. Isolates should be selected with respect to relevant features like genetic markers, origin, date of isolation, host species and other factors in order to ensure that the natural variation of the isolates within a group is represented in the study in a balanced way. It must be kept in mind that MI relies on pattern recognition, so that even excellent algorithms will not find what is not in the database. Also, other available data and metadata should be at hand as they can help explain unexpected clustering behaviour of the mass spectra in the course of the cluster analysis.

11.8.2 Spectrum Processing

Raw spectra are usually transformed into peak lists by three main steps. The raw spectrum is smoothed in order to reduce noise, the baseline is corrected and finally the peaks that meet the minimum quality parameters of the peak search algorithm are annotated. As spectra from closely related organisms may differ only slightly, it is important to adjust the spectrum processing parameters with care. Standard settings may not be suitable. Less intense peaks may be lost if too rigid smoothing is performed or if too stringent values for the minimum intensity or the minimum signal-to-noise (S/N) ratio of valid peaks are chosen. On the other hand, too permissive conditions for the peak search must be avoided in order to exclude noise from the peak list.

But, fortunately, the fact that all samples are very similar can also be taken as a mass-spectrometric advantage. If, for example, the spectra of two subspecies within a species are studied, it most certainly will be possible to establish species-specific marker ions which can be used for the internal mass calibration of spectra from both subspecies. Internal mass calibration significantly improves the mass accuracy and, therefore, smaller m/z tolerances can be applied for calculations like the peak alignments, or the search for group-specific markers. This is of great importance for subspecies-level classification, as it may rely on very subtle differences between the spectra. Choosing unnecessarily large m/z tolerances may lead to the loss of markers with only small mass shifts.

11.8.3 Choosing Software for Statistical Calculations

Manufacturers of hardware offer software tools for the refined analysis of MS data, which are designed for non-statisticians. Although these usually will be the tools of choice for reasons of convenience, the transfer of the peak lists to independent software

packages like the statistical programming languages Matlab® (MathWorks) or R (R_Development_Core_Team 2011) may be considered. R is freely available and offers a number of very powerful options, for example, for multivariate data analysis. Additional packages are available to expand the base package with, for example, implementations of algorithms for discriminant analysis (Sanchez, 2013) or with tools for the processing of MALDI-TOF mass spectra (Gibb and Strimmer, 2012).

11.8.4 Search for Taxon-Specific Markers

Irrespective of the environment that is used, the search for group-specific markers is a consequent next step. Under favourable conditions, one or more reliable, qualitative group-specific markers may be found, and an easy classification can be set up based on the presence or absence of these mass peaks. In this case, however, the parameters that rule peak detection (see above) are of great importance and must be chosen with care.

11.8.5 Spectrum-Based Cluster Analysis

If reliable qualitative taxon-identifying marker peaks cannot be identified, it is helpful to perform a cluster analysis and to visualize the distance relations between the individual isolates and the groups. Towards this end, spectrum-based distances between all isolates can be calculated and represented, for example, in a dendrogram, or, after multidimensional scaling, as a scatterplot (Sammon, 1969). Alternatively, principal component analysis (PCA) can be performed. Visual inspection of the resulting graphs will give a first impression of how well the different groups are separated on basis of the spectra as they are. Graphical representations of the spectrum-based distances will also give a hint on the homogeneity of the isolates within the groups, which is also an important point. If unexpected clusters within a group occur, it should be verified if these coincide with certain features of the samples (e.g. provenience, host species or a genetic marker) or could be related to a technical problem that affected a part of the samples. Also, the presence of singletons ('outliers') should be examined. Interestingly, we have observed in two studies that reference or type strains were indeed outliers in their own groups and, therefore, inappropriate as references for the MS-based identification of field strains (Karger *et al.*, 2012; Schafer *et al.*, 2014).

11.8.6 Statistical Models for Classification

If the graphical representations of the distances between the isolates or of the PCA have not displayed at least a rudimental separation of the different groups, but rather resulted in a random distribution of all isolates, it will be difficult or even impossible to achieve group separation. If further analysis seems promising, multivariate data analysis offers a wealth of different approaches that can be used for the analysis of mass spectra and the implementation of classification algorithms, that is, procedures that allow the prediction of the classes (e.g. subspecies) of unknown samples on basis of statistical models that have been derived from spectra of samples with known classes. A discussion of multivariate data analysis in the context of mass spectrometry would be beyond the scope of this review, but, to mention a few examples, artificial neural networks (ANNs) have been successfully applied for the characterization of Yersinia (Lasch *et al.*, 2010) and *Bacillus subtilis* (Lasch *et al.*, 2009) strains, but also for the tracing of methicillin

resistance in *Staphylococcus aureus* (Shah *et al.*, 2011). The support vector machine (SVM) algorithm has been useful to improve the identification of, among others, species from the Bacillus and Brevibacillus genera (AlMasoud *et al.*, 2014).

As for the design of any other diagnostic test, it is important at this stage that the classification algorithms are cross-validated in order to assess the prediction error and the influence of sample selection. Cross-validation is usually achieved by the implementation of 'leave-one-out' strategies or by repeated random sub-sampling of isolates from each group, which are then used to classify the remaining isolates on the basis of the statistical model. It is a common pitfall to add too many parameters (i.e. masses, in the case of mass spectrometry) to a statistical model in an attempt to optimize the performance, as very complex models are likely to fail with independent samples due to an 'overfitting' effect.

11.8.7 External Validation

Finally, once a procedure for the classification of closely related isolates has been established, it is advisable to test the performance with independent sample spectra from a second laboratory, especially if the workflow is intended to be transferred to other laboratories.

11.9 Conclusion

MALDI-TOF MS has become a routine technology for the species identification of bacteria in microbiological laboratories. Commercial solutions like the MALDI Biotyper (Bruker) and the Vitek MS (bioMérieux) systems have been shown to be very reliable for the identification of the genus and the species of unknown samples. However, already at the species level, some bacteria remain difficult to differentiate, like some species of the Brucella and of the Burkholderia genus that have been mentioned above. In contrast to that, the successful discrimination of subspecies of *Francisella tularensis* or biovars of *Brucella suis* using MALDI Biotyper demonstrates that the capability of established MI procedures to reliably discriminate related bacteria is not generally restricted to organisms above a certain taxonomic rank.

For the further development of MS-based identification of bacteria, two main strategies seem to develop. On the one hand, new experimental variants of MI or even completely different experimental setups like the proteotyping approach have been introduced. On the other hand, the conventional procedures for the MALDI-TOF-MS-based identification of bacteria can be markedly improved by refined statistical analysis. This has been shown for the differentiation of several *E. coli* serotypes including serotypes that can cause severe human disease like the serotypes O157:H7 and O104:H4. Essential improvements of the MS-based diagnosis of select organisms have been described, for example, for the discrimination of *Burkholderia mallei*, *Burkholderia pseudomallei* and *Burkholderia thailandensis*, and for the identification of Brucella species and the biovars of *Brucella suis*. Yet, we have also experienced limitations of MALDI-TOF MS for the discrimination of genetically defined groups of isolates below the level of subspecies, like the clusters of isolates with common genetic markers within *Francisella tularensis* ssp. *holarctica* or within *E. coli* serotypes.

References

Alatoom, A. A., Cunningham, S. A., Ihde, S. M., Mandrekar, J., & Patel, R. (2011). Comparison of direct colony method versus extraction method for identification of gram-positive cocci by use of Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*, 49(8), 2868–2873. doi: 10.1128/jcm.00506-11

AlMasoud, N., Xu, Y., Nicolaou, N., & Goodacre, R. (2014). Optimization of matrix assisted desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) for the characterization of *Bacillus* and *Brevibacillus* species. *Anal Chim Acta*, 840, 49–57. doi: 10.1016/j.aca.2014.06.032

Anderson, N. W., Buchan, B. W., Riebe, K. M., Parsons, L. N., Gnacinski, S., & Ledeboer, N. A. (2012). Effects of solid-medium type on routine identification of bacterial isolates by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*, 50(3), 1008–1013. doi: 10.1128/jcm.05209-11

Balazova, T., Makovcova, J., Sedo, O., Slany, M., Faldyna, M., & Zdrahal, Z. (2014). The influence of culture conditions on the identification of Mycobacterium species by MALDI-TOF MS profiling. *FEMS Microbiol Lett*, 353(1), 77–84. doi: 10.1111/1574-6968.12408

Bielaszewska, M., Köck, R., Friedrich, A. W., von Eiff, C., Zimmerhackl, L. B., Karch, H., & Mellmann, A. (2007). Shiga toxin-mediated hemolytic uremic syndrome: Time to change the diagnostic paradigm? *PLoS ONE*, 2(10). doi: 10.1371/journal.pone.0001024

Bizzini, A., Durussel, C., Bille, J., Greub, G., & Prod'hom, G. (2010). Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. *J Clin Microbiol*, 48(5), 1549–1554. doi: 10.1128/jcm.01794-09

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth. *Belmont, CA*.

Bright, J. J., Claydon, M. A., Soufian, M., & Gordon, D. B. (2002). Rapid typing of bacteria using matrix-assisted laser desorption ionisation time-of-flight mass spectrometry and pattern recognition software. *J Microbiol Methods*, 48(2–3), 127–138.

Christner, M., Trusch, M., Rohde, H., Kwiatkowski, M., Schluter, H., Wolters, M.,… Hentschke, M. (2014). Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-Toxigenic *Escherichia coli*. *PLoS ONE*, 9(7), e101924. doi: 10.1371/journal.pone.0101924

Corbel, M. J., & Brinley-Morgan, W. J. (1984). Genus Brucella Meyer and Shaw 1920, 173AL. *Bergey's Manual of Systematic Bacteriology*, 1, 377–388.

Couderc, C., Nappez, C., & Drancourt, M. (2012). Comparing inactivation protocols of Yersinia organisms for identification with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*, 26(6), 710–714. doi: 10.1002/rcm.6152

Cunningham, S. A., & Patel, R. (2013). Importance of using Bruker's security-relevant library for Biotyper identification of *Burkholderia pseudomallei, Brucella* species, and *Francisella tularensis*. *J Clin Microbiol*, 51(5), 1639–1640. doi: 10.1128/jcm.00267-13

Cunningham, S. A., & Patel, R. (2015). Standard matrix-assisted laser desorption ionization-time of flight mass spectrometry reagents may inactivate potentially hazardous bacteria. *J Clin Microbiol*, 53(8), 2788–2789. doi: 10.1128/jcm.00957-15

Degand, N., Carbonnelle, E., Dauphin, B., Beretti, J. L., Le Bourgeois, M., Sermet-Gaudelus, I.,… Ferroni, A. (2008). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of nonfermenting gram-negative bacilli isolated from cystic fibrosis patients. *J Clin Microbiol*, 46(10), 3361–3367. doi: 10.1128/jcm.00569-08

Desai, A. P., Stanley, T., Atuan, M., McKey, J., Lipuma, J. J., Rogers, B., & Jerris, R. (2012). Use of matrix assisted laser desorption ionisation-time of flight mass spectrometry in a paediatric clinical laboratory for identification of bacteria commonly isolated from cystic fibrosis patients. *J Clin Pathol*, 65(9), 835–838. doi: 10.1136/jclinpath-2012-200772

Dieckmann, R., Helmuth, R., Erhard, M., & Malorny, B. (2008). Rapid classification and identification of salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl Environ Microbiol*, 74(24), 7767–7778. doi: 10.1128/aem.01402-08

Drevinek, M., Dresler, J., Klimentova, J., Pisa, L., & Hubalek, M. (2012). Evaluation of sample preparation methods for MALDI-TOF MS identification of highly dangerous bacteria. *Lett Appl Microbiol*, 55(1), 40–46. doi: 10.1111/j.1472-765X.2012.03255.x

Fagerquist, C. K., Garbus, B. R., Miller, W. G., Williams, K. E., Yee, E., Bates, A. H.,… Mandrell, R. E. (2010). Rapid identification of protein biomarkers of Escherichia coli O157:H7 by matrix-assisted laser desorption ionization-time-of-flight-time-of-flight mass spectrometry and top-down proteomics. *Anal Chem*, 82(7), 2717–2725. doi: 10.1021/ac902455d

Fehlberg, L. C., Andrade, L. H., Assis, D. M., Pereira, R. H., Gales, A. C., & Marques, E. A. (2013). Performance of MALDI-ToF MS for species identification of Burkholderia cepacia complex clinical isolates. *Diagn Microbiol Infect Dis*, 77(2), 126–128. doi: 10.1016/j.diagmicrobio.2013.06.011

Fernandez-Olmos, A., Garcia-Castillo, M., Morosini, M. I., Lamas, A., Maiz, L., & Canton, R. (2012). MALDI-TOF MS improves routine identification of non-fermenting Gram negative isolates from cystic fibrosis patients. *J Cyst Fibros*, 11(1), 59–62. doi: 10.1016/j.jcf.2011.09.001

Foster, G., Osterman, B. S., Godfroid, J., Jacques, I., & Cloeckert, A. (2007). Brucella ceti sp. nov. and *Brucella pinnipedialis* sp. nov. for *Brucella* strains with cetaceans and seals as their preferred hosts. *International Journal of Systematic and Evolutionary Microbiology*, 57(11), 2688–2693. doi: 10.1099/ijs.0.65269-0

Fournier, R., Wallet, F., Grandbastien, B., Dubreuil, L., Courcol, R., Neut, C., & Dessein, R. (2012). Chemical extraction versus direct smear for MALDI-TOF mass spectrometry identification of anaerobic bacteria. *Anaerobe*, 18(3), 294–297. doi: 10.1016/j.anaerobe.2012.03.008

Geue, L., Klare, S., Schnick, C., Mintel, B., Meyer, K., & Conraths, F. J. (2009). Analysis of the clonal relationship of serotype O26:H11 enterohemorrhagic *Escherichia coli* isolates from cattle. *Appl Environ Microbiol*, 75(21), 6947–6953. doi: 10.1128/AEM.00605-09

Geue, L., Schares, S., Mintel, B., Conraths, F. J., Müller, E., & Ehricht, R. (2010). Rapid microarray-based genotyping of enterohemorrhagic *Escherichia coli* serotype O156:H25/H-/Hnt isolates from cattle and clonal relationship analysis. *Appl Environ Microbiol*, 76(16), 5510–5519. doi: 10.1128/AEM.00743-10

Geue, L., Selhorst, T., Schnick, C., Mintel, B., & Conraths, F. J. (2006). Analysis of the clonal relationship of Shiga toxin-producing Escherichia coli serogroup O165:H25 isolated from cattle. *Appl Environ Microbiol*, 72(3), 2254–2259. doi: 10.1128/AEM.72.3.2254-2259.2006

Gibb, S., & Strimmer, K. (2012). MALDIquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics*, 28(17), 2270–2271. doi: 10.1093/bioinformatics/bts447

Godfroid, J., Scholz, H. C., Barbier, T., Nicolas, C., Wattiau, P., Fretin, D.,…Letesson, J. J. (2011). Brucellosis at the animal/ecosystem/human interface at the beginning of the 21st century. *Prev Vet Med*, 102(2), 118–131. doi: 10.1016/j.prevetmed.2011.04.007

Godoy, D., Randle, G., Simpson, A. J., Aanensen, D. M., Pitt, T. L., Kinoshita, R., & Spratt, B. G. (2003). Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J Clin Microbiol*, 41(5), 2068–2079.

Ho, C. C., Lau, C. C., Martelli, P., Chan, S. Y., Tse, C. W., Wu, A. K.,…Woo, P. C. (2011). Novel pan-genomic analysis approach in target selection for multiplex PCR identification and detection of *Burkholderia pseudomallei, Burkholderia thailandensis*, and *Burkholderia cepacia* complex species: A proof-of-concept study. *J Clin Microbiol*, 49(3), 814–821. doi: 10.1128/jcm.01702-10

Inglis, T. J., Healy, P. E., Fremlin, L. J., & Golledge, C. L. (2012). Use of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis for rapid confirmation of *Burkholderia pseudomallei* in septicemic melioidosis. *Am J Trop Med Hyg*, 86(6), 1039–1042. doi: 10.4269/ajtmh.2012.11-0454

Jarman, K. H., & Wahl, K. L. (2005). Development of spectral pattern-matching approaches to matrix-assisted laser desorption/ionization mass spectrometry for bacterial identification. In C. L. Wilkins & J. O. Lay (Eds.), *Identification of Microorganisms by Mass Spectrometry* (pp. 153–160): Wiley.

Johansson, A., Farlow, J., Larsson, P., Dukerich, M., Chambers, E., Bystrom, M.,…Keim, P. (2004). Worldwide genetic relationships among *Francisella tularensis* isolates determined by multiple-locus variable-number tandem repeat analysis. *J Bacteriol*, 186(17), 5808–5818. doi: 10.1128/jb.186.17.5808-5818.2004

Karger, A., Melzer, F., Timke, M., Bettin, B., Kostrzewa, M., Nockler, K.,…Al Dahouk, S. (2013). Interlaboratory comparison of intact-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry results for identification and differentiation of *Brucella* spp. *J Clin Microbiol*, 51(9), 3123–3126. doi: 10.1128/jcm.01720-13

Karger, A., Stock, R., Ziller, M., Elschner, M. C., Bettin, B., Melzer, F.,…Tomaso, H. (2012). Rapid identification of *Burkholderia mallei* and *Burkholderia pseudomallei* by intact cell Matrix-assisted Laser Desorption/Ionisation mass spectrometric typing. *BMC Microbiol*, 12, 229. doi: 10.1186/1471-2180-12-229

Karger, A., Ziller, M., Bettin, B., Mintel, B., Schares, S., & Geue, L. (2011). Determination of serotypes of Shiga toxin-producing *Escherichia coli* isolates by intact cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl Environ Microbiol*, 77(3), 896–905. doi: 10.1128/AEM.01686-10

Khot, P. D., Couturier, M. R., Wilson, A., Croft, A., & Fisher, M. A. (2012). Optimization of matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis for bacterial identification. *J Clin Microbiol*, 50(12), 3845–3852. doi: 10.1128/jcm.00626-12

Krasny, L., Rohlova, E., Ruzickova, H., Santrucek, J., Hynek, R., & Hochel, I. (2014). Differentiation of *Cronobacter* spp. by tryptic digestion of the cell suspension followed by MALDI-TOF MS analysis. *J Microbiol Methods*, 98, 105–113. doi: 10.1016/j.mimet.2014.01.008

Lambiase, A., Del Pezzo, M., Cerbone, D., Raia, V., Rossano, F., & Catania, M. R. (2013). Rapid identification of *Burkholderia cepacia* complex species recovered from cystic fibrosis patients using matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *J Microbiol Methods*, 92(2), 145–149. doi: 10.1016/j.mimet.2012.11.010

Lasch, P., Beyer, W., Nattermann, H., Stämmler, M., Siegbrecht, E., Grunow, R., & Naumann, D. (2009). Identification of *Bacillus anthracis* by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. *Appl Environ Microbiol*, 75(22), 7229–7242. doi: 10.1128/aem.00857-09

Lasch, P., Drevinek, M., Nattermann, H., Grunow, R., Stammler, M., Dieckmann, R.,… Naumann, D. (2010). Characterization of *Yersinia* using MALDI-TOF mass spectrometry and chemometrics. *Anal Chem*, 82(20), 8464–8475. doi: 10.1021/ac101036s

Lasch, P., Nattermann, H., Erhard, M., Stämmler, M., Grunow, R., Bannert, N.,…Naumann, D. (2008). MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. *Anal Chem*, 80(6), 2026–2034. doi: 10.1021/ac701822j

Lasch, P., Wahab, T., Weil, S., Palyi, B., Tomaso, H., Zange, S.,…Jacob, D. (2015). Identification of Highly pathogenic microorganisms using MALDI-TOF mass spectrometry – results of an inter-laboratory ring trial. *J Clin Microbiol*. doi: 10.1128/jcm.00813-15

Lau, S. K., Tang, B. S., Curreem, S. O., Chan, T. M., Martelli, P., Tse, C. W.,…Woo, P. C. (2012). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of *Burkholderia pseudomallei*: Importance of expanding databases with pathogens endemic to different localities. *J Clin Microbiol*, 50(9), 3142–3143. doi: 10.1128/jcm.01349-12

Lista, F., Reubsaet, F. A., De Santis, R., Parchen, R. R., de Jong, A. L., Kieboom, J.,…Paauw, A. (2011). Reliable identification at the species level of *Brucella* isolates with MALDI-TOF-MS. *BMC Microbiol*, 11, 267. doi: 10.1186/1471-2180-11-267

Lowe, W., March, J. K., Bunnell, A. J., O'Neill, K. L., & Robison, R. A. (2013). PCR-based Methodologies Used to detect and differentiate the *Burkholderia pseudomallei* complex: *B. pseudomallei, B. mallei*, and *B. thailandensis*. *Curr Issues Mol Biol*, 16(2), 23–54.

Lundquist, M., Caspersen, M. B., Wikstrom, P., & Forsman, M. (2005). Discrimination of *Francisella tularensis* subspecies using surface enhanced laser desorption ionization mass spectrometry and multivariate data analysis. *FEMS Microbiol Lett*, 243(1), 303–310. doi: 10.1016/j.femsle.2004.12.020

Luo, Y., Siu, G. K., Yeung, A. S., Chen, J. H., Ho, P. L., Leung, K. W.,…Yam, W. C. (2015). Performance of the VITEK MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system for rapid bacterial identification in two diagnostic centres in China. *J Med Microbiol*, 64(Pt 1), 18–24. doi: 10.1099/jmm.0.080317-0

Martiny, D., Cremagnani, P., Gaillard, A., Miendje Deyi, V. Y., Mascart, G., Ebraert, A.,… Vandenberg, O. (2014). Feasibility of matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) networking in university hospitals in Brussels. *Eur J Clin Microbiol Infect Dis*, 33(5), 745–754. doi: 10.1007/s10096-0132006-6

Mather, C. A., Rivera, S. F., & Butler-Wu, S. M. (2014). Comparison of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight mass

spectrometry systems for identification of mycobacteria using simplified protein extraction protocols. *J Clin Microbiol*, 52(1), 130–138. doi: 10.1128/jcm.01996-13

Matsuda, N., Matsuda, M., Notake, S., Yokokawa, H., Kawamura, Y., Hiramatsu, K., & Kikuchi, K. (2012). Evaluation of a simple protein extraction method for species identification of clinically relevant staphylococci by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*, 50(12), 3862–3866. doi: 10.1128/jcm.01512-12

Mazzeo, M. F., Sorrentino, A., Gaita, M., Cacace, G., Di Stasio, M., Facchiano, A.,… Siciliano, R. A. (2006). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the discrimination of food-borne microorganisms. *Appl Environ Microbiol*, 72(2), 1180–1189. doi: 10.1128/aem.72.2.1180-1189.2006

McElvania Tekippe, E., Shuey, S., Winkler, D. W., Butler, M. A., & Burnham, C. A. (2013). Optimizing identification of clinically relevant Gram-positive organisms by use of the Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system. *J Clin Microbiol*, 51(5), 1421–1427. doi: 10.1128/jcm.02680-12

Meetani, M. A., & Voorhees, K. J. (2005). MALDI mass spectrometry analysis of high molecular weight proteins from whole bacterial cells: Pretreatment of samples with surfactants. *J Am Soc Mass Spectrom*, 16(9), 1422–1426. doi: 10.1016/j.jasms.2005.04.004

Mellmann, A., Bielaszewska, M., Köck, R., Friedrich, A. W., Fruth, A., Middendorf, B.,… Karch, H. (2008). Analysis of collection of hemolytic uremic syndrome-associated enterohemorrhagic *Escherichia coli*. *Emerg Infect Dis*, 14(8), 1287–1290. doi: 10.3201/eid1408.071082

Miñán, A., Bosch, A., Lasch, P., Stämmler, M., Serra, D. O., Degrossi, J.,…Naumann, D. (2009). Rapid identification of Burkholderia cepacia complex species including strains of the novel Taxon K, recovered from cystic fibrosis patients by intact cell MALDI-ToF mass spectrometry. *Analyst*, 134(6), 1138–1148. doi: 10.1039/b822669e

Moreno, E., Cloeckaert, A., & Moriyon, I. (2002). *Brucella* evolution and taxonomy. *Vet Microbiol*, 90(1–4), 209–227.

Muller, W., Hotzel, H., Otto, P., Karger, A., Bettin, B., Bocklisch, H.,…Tomaso, H. (2013). German *Francisella tularensis* isolates from European brown hares (*Lepus europaeus*) reveal genetic and phenotypic diversity. *BMC Microbiol*, 13, 61. doi: 10.1186/1471-2180-13-61

Nataro, J. P., & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev*, 11(1), 142–201.

Niyompanich, S., Jaresitthikunchai, J., Srisanga, K., Roytrakul, S., & Tungpradabkul, S. (2014). Source-identifying biomarker ions between environmental and clinical *Burkholderia pseudomallei* using whole-cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). *PLoS ONE*, 9(6), e99160. doi: 10.1371/journal.pone.0099160

Ojima-Kato, T., Yamamoto, N., Suzuki, M., Fukunaga, T., & Tamura, H. (2014). Discrimination of *Escherichia coli* O157, O26 and O111 from other serovars by MALDI-TOF MS based on the S10-GERMS method. *PLoS ONE*, 9(11), e113458. doi: 10.1371/journal.pone.0113458

Paauw, A., Trip, H., Niemcewicz, M., Sellek, R., Heng, J. M., Mars-Groenendijk, R. H.,… Tsivtsivadze, E. (2014). OmpU as a biomarker for rapid discrimination between toxigenic and epidemic *Vibrio cholerae* O1/O139 and non-epidemic Vibrio cholerae in a modified MALDI-TOF MS assay. *BMC Microbiol*, 14, 158. doi: 10.1186/1471-2180-14-158

Pappas, G., Papadimitriou, P., Akritidis, N., Christou, L., & Tsianos, E. V. (2006). The new global map of human brucellosis. *Lancet Infect Dis*, 6(2), 91–99. doi: 10.1016/S1473-3099(06)70382-6

Patel, R. (2013). Matrix-assisted laser desorption ionization-time of flight mass spectrometry in clinical microbiology. *Clin Infect Dis*, 57(4), 564–572. doi: 10.1093/cid/cit247

Preussel, K., Hohle, M., Stark, K., & Werber, D. (2013). Shiga toxin-producing Escherichia coli O157 is more likely to lead to hospitalization and death than non-O157 serogroups – except O104. *PLoS ONE*, 8(11), e78180. doi: 10.1371/journal.pone.0078180

Robert-Koch-Institut. (2011). Bericht: Abschließende Darstellung und Bewertung der epidemiologischen Erkenntnisse im EHEC O104:H4 Ausbruch, *Deutschland* 2011.

Sammon, J. (1969). A non-linear mapping for data structure analysis. *IEEE Trans Comp C*, 18, 401–409.

Sanchez, G. (2013). *DiscriMiner: Tools of the Trade for Discriminant Analysis*. R package version 0.1–29.

Sauer, S., Freiwald, A., Maier, T., Kube, M., Reinhardt, R., Kostrzewa, M., & Geider, K. (2008). Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS ONE*, 3(7). doi: 10.1371/journal.pone.0002843

Schäfer, M. O., Genersch, E., Funfhaus, A., Poppinga, L., Formella, N., Bettin, B., & Karger, A. (2014). Rapid identification of differentially virulent genotypes of *Paenibacillus larvae*, the causative organism of American foulbrood of honey bees, by whole cell MALDI-TOF mass spectrometry. *Vet Microbiol*, 170(3–4), 291–297. doi: 10.1016/j.vetmic.2014.02.006

Schmidt, F., Fiege, T., Hustoft, H. K., Kneist, S., & Thiede, B. (2009). Shotgun mass mapping of *Lactobacillus* species and subspecies from caries related isolates by MALDI-MS. *Proteomics*, 9(7), 1994–2003. doi: 10.1002/pmic.200701028

Schmitt, B. H., Cunningham, S. A., Dailey, A. L., Gustafson, D. R., & Patel, R. (2013). Identification of anaerobic bacteria by Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry with on-plate formic acid preparation. *J Clin Microbiol*, 51(3), 782–786. doi: 10.1128/jcm.02420-12

Scholz, H. C., Hubalek, Z., Sedláček, I., Vergnaud, G., Tomaso, H., Al Dahouk, S.,… Nöckler, K. (2008). *Brucella microti* sp. nov., isolated from the common vole *Microtus arvalis*. *Int J Syst Evol Microbiol*, 58(2), 375–382. doi: 10.1099/ijs.0.65356-0

Scholz, H. C., Nöckler, K., Llner, C. G., Bahn, P., Vergnaud, G., Tomaso, H.,…De, B. K. (2010). *Brucella inopinata* sp. nov., isolated from a breast implant infection. *International Journal of Systematic and Evolutionary Microbiology*, 60(4), 801–808. doi: 10.1099/ijs.0.011148-0

Schulthess, B., Brodner, K., Bloemberg, G. V., Zbinden, R., Bottger, E. C., & Hombach, M. (2013). Identification of Gram-positive cocci by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry: Comparison of different preparation methods and implementation of a practical algorithm for routine diagnostics. *J Clin Microbiol*, 51(6), 1834–1840. doi: 10.1128/jcm.02654-12

Sedo, O., Vavrova, A., Vad'urova, M., Tvrzova, L., & Zdrahal, Z. (2013). The influence of growth conditions on strain differentiation within the *Lactobacillus acidophilus* group using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry profiling. *Rapid Commun Mass Spectrom*, 27(24), 2729–2736. doi: 10.1002/rcm.6741

Seibold, E., Bogumil, R., Vorderwulbecke, S., Al Dahouk, S., Buckendahl, A., Tomaso, H., & Splettstoesser, W. (2007). Optimized application of surface-enhanced laser desorption/

ionization time-of-flight MS to differentiate *Francisella tularensis* at the level of subspecies and individual strains. *FEMS Immunol Med Microbiol*, 49(3), 364–373. doi: 10.1111/j.1574-695X.2007.00216.x

Seibold, E., Maier, T., Kostrzewa, M., Zeman, E., & Splettstoesser, W. (2010). Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: Fast, reliable, robust, and cost-effective differentiation on species and subspecies levels. *J Clin Microbiol*, 48(4), 1061–1069. doi: 10.1128/jcm.01953-09

Seng, P., Abat, C., Rolain, J. M., Colson, P., Lagier, J. C., Gouriet, F.,…Raoult, D. (2013). Identification of rare pathogenic bacteria in a clinical microbiology laboratory: Impact of matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*, 51(7), 2182–2194. doi: 10.1128/jcm.00492-13

Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P. E., Rolain, J. M., & Raoult, D. (2009). Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis*, 49(4), 543–551. doi: 10.1086/600885

Shah, H. N., Rajakaruna, L., Ball, G., Misra, R., Al-Shahib, A., Fang, M., & Gharbia, S. E. (2011). Tracing the transition of methicillin resistance in sub-populations of *Staphylococcus aureus*, using SELDI-TOF mass spectrometry and artificial neural network analysis. *Syst Appl Microbiol*, 34(1), 81–86. doi: 10.1016/j.syapm.2010.11.002

Tracz, D. M., McCorrister, S. J., Westmacott, G. R., & Corbett, C. R. (2013). Effect of gamma radiation on the identification of bacterial pathogens by MALDI-TOF MS. *J Microbiol Methods*, 92(2), 132–134. doi: 10.1016/j.mimet.2012.11.013

Vanlaere, E., Sergeant, K., Dawyndt, P., Kallow, W., Erhard, M., Sutton, H.,…Vandamme, P. (2008). Matrix-assisted laser desorption ionisation-time-of of-flight mass spectrometry of intact cells allows rapid identification of Burkholderia cepacia complex. *J Microbiol Methods*, 75(2), 279–286. doi: 10.1016/j.mimet.2008.06.016

Veloo, A. C., Elgersma, P. E., Friedrich, A. W., Nagy, E., & van Winkelhoff, A. J. (2014). The influence of incubation time, sample preparation and exposure to oxygen on the quality of the MALDI-TOF MS spectrum of anaerobic bacteria. *Clin Microbiol Infect*, 20(12), O1091–1097. doi: 10.1111/1469-0691.12644

Verger, J. M., Grimont, F., Grimont, P. A. D., & Grayon, M. (1985). *Brucella*, a monospecific genus as shown by deoxyribonucleic acid hybridization. *International Journal of Systematic Bacteriology*, 35(3), 292–295.

Vogler, A. J., Birdsell, D., Price, L. B., Bowers, J. R., Beckstrom-Sternberg, S. M., Auerbach, R. K.,…Keim, P. (2009). Phylogeography of *Francisella tularensis*: Global expansion of a highly fit clone. *J Bacteriol*, 191(8), 2474–2484. doi: 10.1128/jb.01786-08

Whatmore, A. M. (2009). Current understanding of the genetic diversity of Brucella, an expanding genus of zoonotic pathogens. *Infect Genet Evol*, 9(6), 1168–1184. doi: 10.1016/j.meegid.2009.07.001

Wilen, C. B., McMullen, A. R., & Burnham, C. A. (2015). Comparison of sample preparation methods, instrumentation platforms, and contemporary commercial databases for identification of clinically relevant mycobacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol*, 53(7), 2308–2315. doi: 10.1128/JCM.00567-15

Yagupsky, P., & Baron, E. J. (2005). Laboratory exposures to brucellae and implications for bioterrorism. *Emerg Infect Dis*, 11(8), 1180–1185.

12

MALDI-TOF-MS Based on Ribosomal Protein Coding in *S10-spc-alpha* Operons for Proteotyping

Hiroto Tamura

Laboratory of Environmental Microbiology, Department of Environmental Bioscience, Meijo University, Nagoya, Japan

12.1 Introduction

Over the last 20 years, 16S rRNA gene sequencing has been used widely for identification of bacteria. Due to the improvement in the accuracy of 16S rRNA gene sequencing techniques, an isolate, which shares less than 98.7%–99% similarity based on 16S rRNA gene sequencing, is assigned as a novel species [1]. However, in some cases, the isolate with greater than 99% similarity of 16S rRNA gene sequencing exhibits less than the DNA–DNA hybridization value of 70% [2,3]. Therefore, the usage of the 16S rRNA gene sequencing technique is still limited from family to species and is not applicable to discrimination at the strain level. On the other hand, unique signatures for bacterial characterization were observed by mass spectrometry (MS) obtained from bacterial extracts in 1975 [4] and low molecular biomarkers such as lipids were analyzed for bacterial profiling [5]. As a result of development of a robust MS approach for the rapid and cost-effective identification of microorganisms, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis of whole cells by soft ionization has been successfully applied to identify not only clinically important microorganisms in diagnostic laboratories [6–8] but also *Archaea* [9], fungi and yeast [10–13], and even viruses [14]. Furthermore, MALDI-TOF MS showed better potential to discriminate bacteria at the subspecies level than 16S rRNA gene sequencing, whose similarity is 99%–100% [15], by using the statistical coefficient of correlation and permitted the typing of microbial isolates at the strain or serovar level using discriminating peaks [16–18]. The technique most utilized in the bacterial identification by MALDI-TOF MS is the fingerprint method, which is the comparison of the mass spectra of target isolates with those of known reference strains in well-characterized commercially available databases, because the fingerprint method is easier, more rapid, and has a higher throughput and a lower cost than conventional techniques [19–24]. Although this fingerprint method has become increasingly popular in stable and robust bacterial identification at the species level independent of culture conditions and operator skill

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

[25,26], culture conditions and experimental factors still affect both the quality and reproducibility of the obtained mass spectra [27–29]. Therefore, the fingerprint method has one major drawback as a result of not specifying and identifying any peaks by MALDI-TOF MS analysis. Although the resulting mass spectra of closely related strains, in particular below the species level, gives us very similar mass spectra or mass spectra with subtle differences among them, the fact that the fingerprint method may lead to unreliable identification necessitates a considerable effort to address the optimization of reliable routine bacterial identification by using statistical algorithmic methods [30–33].

The most informative observed MALDI-TOF MS spectra consist of peaks derived mostly from ribosomal and other housekeeping proteins because of their following characteristics: significant amount of protein expression, appropriate molecular weight range (*m/z*: 4000–15000), and higher pI [34–40].

The bacterial ribosome consists of a 50S large subunit with 34 proteins (L1–L34), a 30S small subunit with 21 proteins (S1–S21), and 16S rRNA. Although high-throughput sequencing of the 16S rRNA genes using next-generation sequencer provides a powerful approach for the taxonomic classification of bacteria and characterization of bacterial community profiling [41], this genetic identification reflects the information only from a single gene with about 1600 bp, suggesting considerable concern regarding the phylogenetic relationship. Contrary to the 16S rRNA gene sequence, the advantage of using ribosomal proteins as biomarkers is the reflection of more information from 55 types of ribosomal subunit proteins, implying proteomic identification based on diverse and multiple genetic information similar to multilocus sequence typing (MLST) analysis, which is generally performed according to the seven-loci schemes for the target bacteria (Figure 12.1).

Because the masses of identified peaks of ribosomal proteins are deduced on the basis of their corresponding amino acids sequences associated with the target genes, the bioinformatics-based approach has been developed for a highly reliable advanced discrimination method at the strain level with a validation procedure [6,38–40,42–44].

Although the advancement in whole genome sequencing technology will allow the use of MALDI-TOF MS for the bioinformatics-based approach, the organization of operons coding many ribosomal proteins does not follow the promoter-structural gene-terminator paradigm [45]. Furthermore, even though the gene order is conserved in closely related taxa, it rapidly becomes less conserved with evolutionary distance [46,47]. These facts mean that researchers have to search a huge data bank of information for the target ribosomal proteins. In addition, the post-translational modifications and incorrect annotations in the sequenced genomes in the databases may cause differences between the calculated and observed masses of the ribosomal proteins as biomarkers [48–51]. Therefore, researchers may have no clue about how to search for the genes coding the target ribosomal proteins scattered around the bacterial genomes.

To overcome those problems, it is essential to establish and standardize a simple and reliable construction method for a ribosomal proteins database. Therefore, the *S10-spc-alpha* operon is selected for biomarker mines. The standardized MALDI-TOF MS method that combines genomics and proteomics is designated the *S10*-GERMS (*S10-spc-alpha* operon gene-encoded ribosomal protein mass spectrum) method. The *S10*-GERMS method offers an accurate method to construct a database by comparing the experimentally observed mass-to-ion ratio (*m/z*) values of the selected biomarkers with

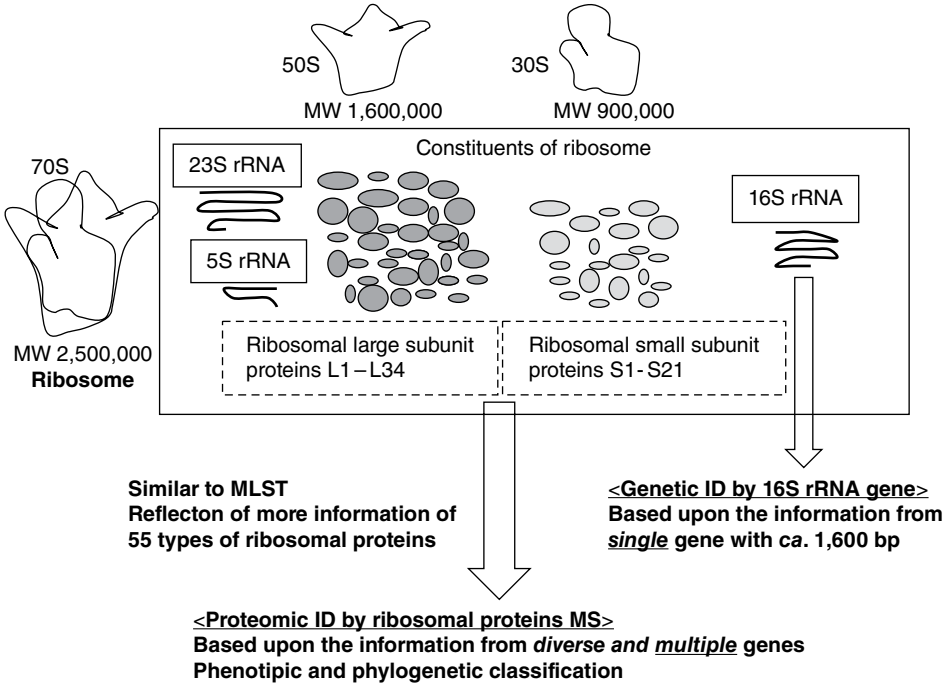


Figure 12.1 Advantage of using ribosomal proteins in comparison with 16S rRNA.

their theoretically calculated *m/z* values and has been employed as a typing method for various taxa.

In this chapter, the construction procedures of a working database for MALDI-TOF MS analysis is illustrated as follows: first, MALDI-TOF MS analysis of the genome-sequenced strains is performed to obtain the observed *m/z* values. Second, the theoretical *m/z* values of ribosomal proteins in this operon are calculated by sequence data from the NCBI databank for genome-sequenced strains or determination of the DNA sequence by using designed primers against the consensus DNA sequences, and then the candidate biomarkers are selected by comparison with the theoretical *m/z* values of each ribosomal protein in silico. Third, the reliable *m/z* values of candidate biomarkers are corrected by comparing the observed *m/z* values of the candidate biomarkers with their in silico–calculated *m/z* values (working database).

As the *S10*-GERMS method reflects different evolutionary lineages for ribosomal proteins backed by multi-gene sequence information, the *S10*-GERMS method, similar to MLST, provides an accurate means of discriminating the bacteria below the species level across the microbial kingdom with reproducibility.

Then, the following pages introduce the use of the standardized *S10*-GERMS method successfully to discriminate *Pseudomonas putida* at the strain level, *Pseudomonas syringae* at the pathovar level, the genera *Bacillus* and *Sphingopyxis* despite only two and one base differences in the 16S rRNA gene sequence, *Lactobacillus casei* at the (sub)species level, and the enterohemorrhagic *Escherichia coli* (EHEC) O157, O26, and O111 serovars with a high level of confidence.

Finally, development of computer-aided proteotyping of bacteria in combination with genomics based on the *S10*-GERMS method is discussed.

12.2 *S10*-GERMS Method

12.2.1 Background of Proteotyping

Although the usefulness of MALDI-TOF MS for bacterial identification is recognized in the fields of clinical and environmental microbiology, the food industry and safety, molecular epidemiology and counterterrorism, the technique most utilized in microbe identification by MALDI-TOF MS, fingerprinting [20,21,52] without specifying and identifying any observed peaks, does not reflect the microbial evolutionary relatedness, because those peaks in fingerprinting have no direct relation with genetic markers like the 16S rRNA gene sequence for taxonomic classification [36,37,53]. In particular, many isolated bacteria in clinical laboratories are very closely related, and their detail profiling is required at the strain and/or pathovar level. Therefore, extending the application to the bacterial discrimination at the strain level has been one of the expectations for MALDI-TOF MS analysis.

Because most of the ribosomal proteins could be observed by MALDI-TOF MS analysis, the bioinformatics-based approach has been proposed using ribosomal proteins as biomarkers for rapid identification of bacteria [6,38,42–44] because of their following characteristics: significant amount of protein expression, appropriate molecular weight range, higher pI, moderate sequence conservation, fewer post-translational modifications without *N*-terminal methionine loss, and database accessibility [38–40,49].

Although the complete microbial genome sequencing project has been advanced by next generation pyrosequencing, it is still essential for user-friendliness to establish simple and reliable construction procedures for a ribosomal proteins database by the conventional DNA sequencer.

To establish a standardized method, the technological requirements are as follows: (1) functionally related genes tend to be located at the same locus, (2) the organization of gene order is conserved, and (3) the genome size is readable by a conventional sequencer. Because the *S10-spc-alpha* operon encodes more than half of the ribosomal proteins and a number of housekeeping genes, is highly conserved among bacterial and archaeal genomes, and has a genome size of approximately 15–18 kb [54–58], this operon provides strong support for the concept. Moreover, the sequences of ribosomal proteins in this operon suggest that horizontal gene transfer may have played a significant role in the evolution of this operon [57]. Therefore, the *S10-spc-alpha* operon is selected for biomarker mines. Using this operon in silico, strain-specific candidate biomarkers can be predicted by in silico–calculated masses (*m/z*) calculated on the basis of the DNA sequence information coded in this operon prior to MALDI-TOF MS measurement. In the case of non-genome-sequenced strains, the DNA sequence of the ribosomal proteins encoded in the *S10-spc-alpha* operon can be analyzed by using primers designed against the consensus DNA sequences among the target genus and/or species of the bacteria.

12.2.2 Construction Procedures of the Working Database for MALDI-TOF MS Analysis

The theoretical masses (*m/z)* of the ribosomal proteins can be calculated by the sequence information in *S10-spc-alpha* operon of the genome-sequenced type strains. There is an important advantage of the *S10*-GERMS method, which means that some errors found in the in silico–calculated masses (*m/z*), including incorrect annotation and post-translational modification like *N*-terminal methionine loss, methylation, and acetylation in a few case [11,40,49,59], are easily corrected by a comparison with the corresponding experimentally observed masses by whole cell MALDI-TOF MS analysis (WC-MS). Thus, the selected biomarker proteins have a role in rapid bacterial discrimination at the strain level with reliability and reproducibility. A concrete workflow of *S10*-GERMS method is as follows (Figure 12.2).

Step 1: MALDI-TOF MS analysis of the genome-sequenced type strain or non- genome-sequenced type strain (observed data); mass calibration of whole cells of *P. putida* KT2440 was carried out by external calibration using two peaks of myoglobin ($[M+H]^+$, *m/z* 16952.6, and $[M+H]^{2+}$, *m/z* 8476.8), followed by self-calibration using four moderately strong peaks assigned to ribosomal subunit proteins L36 ($[M+H]^+$, *m/z* 4435.3), L29 ($[M+H]^+$, *m/z* 7173.3), S10 ($[M+H]^+$, *m/z* 10753.6), and L15 ($[M+H]^+$, *m/z* 15190.4) as internal references. Mass calibration of the other samples was performed by external calibration using the mass spectra observed for the whole cell of the KT2440 strain. Either bacterial colonies grown on agar plate or bacteria harvested from liquid culture by centrifugation are available for the conventional MALDI-TOF MS analysis. MALDI-TOF mass spectra in the range of *m/z* 4000–20000 are observed in the positive linear mode, and sinapinic acid is preferably used as a matrix reagent because it is found to be better for the detection of a higher range of *m/z* than α(alpha)-cyano-4-hydroxycinnamic acid.

Step 2: Design of primers to sequence the genes encoded in *S10-spc-alpha* operon: The sequencing primers are designed against the consensus nucleotide sequences in the *S10-spc-alpha* operon by using the sequence information of the genome read type strain and/or some genome-sequenced strains of the same genus obtained from the NCBInr database (http://www.ncbi.nlm.nih.gov/). If genome sequence information of the target strains is not available in the NCBInr database, the *S10-spc-alpha* operon of the target strains will be sequenced by the designed primers. The respective regions of ribosomal protein-encoding genes (≈5 kbp) are sequenced by conventional methods. The unregistered ribosomal proteins and missing annotations of the start codon could also be confirmed by sequencing the genes.

Step 3: Calculation of the theoretical ionized mass (*m/z*) of respective ribosomal protein and construction of in silico–calculated *m/z* database: The gene sequences obtained at step 2 are translated into the corresponding amino acids sequences. The theoretical ionized mass (*m/z*) of each protein is calculated on the basis of the translated amino acids sequence using a Compute pI/Mw tool on the ExPASy proteomics server (http://web.expasy.org/compute_pi/).

As a post-translational modification, *N*-terminal methionine is selectively cleaved by amino peptidase when the size of neighbor amino acid branches such as glycine, alanine, serine, proline, valine, threonine, and cysteine is less than 1.29 angstrom, i.e. " the

Step 1 : Measurement of samples by MALDI-TOF MS (WC or IC MS)

Step 2 : Design of primers to sequence the genes encoded in *S10-spc-alpha* operon

Point: Consensus DNA sequences encoded in *S10-spc-alpha* operon of genome read type strains. Designed primers in this chapter are listed in Table xx.

Step3 : Calculation of the theoretical ionized mass (*m/z*) of respective ribosomal protein and construction of *in silico*-calculated *m/z* database

S17 MAEAEKTVRT.....RAVEV
S19 MPRSLKKGPF.....KKAKR
L23 MNQERVFKVL.....SSSAE
S14 MAKKSMKNRE.....VKASW
L24 MQKIRRDDEI.....KAVDA
S10 MQNQQIRIRL.....QISLG
L22 MEVAAKLSGA.....KVADK
L18 MTDKKVIRLR.....GGLEF
S13 MARIAGVNIP.....KPIRK
S11 MAKPAARPRK.....KKRRV
S08 MSMQDPLADM.....LCTVF

Translation respective ribosomal protein

		P.putida 14164	*P.f luorescens* 14160	*P.alcaligenes* 14159	*P.aeruginosa* 12689	*P. asotoformans* 12693	*P.chlororaphis* 3904	*P.f. ulva* 16637	*P.mendocina* 14162	*P.straminea* 16665	*P.stutseri* 14165
S10	L22	**11912.0**	**11912.0**	11893.9	**11912.0**	**11912.0**	**11912.0**	**11912.0**	**11912.0**	**11912.0**	11897.9
	L23	*10900.7*	**10945.7**	10955.6	10950.7	**10945.7**	**10945.7**	*10900.7*	11015.7	11085.8	10920.6
	L29	**7173.3**	**7173.3**	7215.4	7202.4	**7173.3**	**7173.3**	**7173.3**	7205.4	7215.4	7274.4
	S10	**11753.6**	**11753.6**	*11783.6*	11767.6	**11753.6**	**11753.6**	**11753.6**	*11783.6*	11755.6	**11753.6**
	S17	**9902.5**	*9966.6*	9957.6	9955.6	*9966.6*	9984.6	**9902.5**	9974.6	10014.7	9973.6
	S19	**10218.1**	10246.1	10186.0	10227.1	10189.1	10204.0	**10218.1**	10176.0	10190.0	10163.0
spc	L14	**13410.9**	**13410.9**	*13396.8*	13412.9	**13410.9**	**13410.9**	**13410.9**	*13396.8*	**13410.9**	13436.9
	L18	12497.4	12556.4	12561.4	12531.4	**12512.4**	**12512.4**	12485.4	12413.3	12457.3	12477.3
	L24	**11330.2**	*11336.3*	11340.3	11471.5	*11336.3*	11345.3	**11330.2**	11344.3	11344.3	11413.4
	L30	6334.5	**6395.6**	6278.3	6347.4	**6395.6**	**6395.6**	6292.5	6363.5	6448.5	6463.6
	L36	**4435.4**	**4435.4**	4407.3	**4435.4**	**4435.4**	**4435.4**	**4435.4**	**4435.4**	**4435.4**	4421.4
	S08	13845.1	13962.3	13951.3	14040.4	13920.2	13973.3	13861.1	13928.2	13914.2	13869.1
	S14	11259.3	**11304.3**	11394.3	11435.3	**11304.3**	11274.2	11288.3	11359.2	11385.3	11326.2
alpha	S11	13529.5	13485.4	13517.5	**13513.5**	13499.4	**13513.5**	13543.5	13531.5	13493.4	13527.5
	S13	13126.3	13210.4	13177.4	13135.2	13164.5	13239.4	13140.3	13118.3	13058.3	13176.4

Figure 12.2 Construction procedures of the working database for MALDI-TOF MS analysis based on the *S10*-GERMS method.

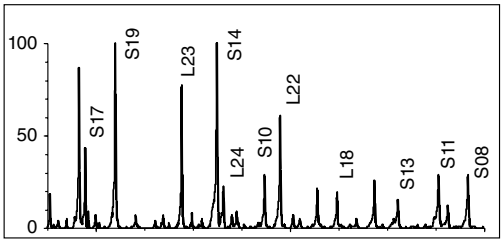
Step 4: Construction of accurate database by comparison of theoretical *vs.* observed data

In silico-calculated *m/z* data from step 3

		P.putida	*P.fluorescens*	*P.alcaligenes*	*P.aeruginosa*	*P.azotoformans*	*P.chlororaphis*	*P.fulva*	*P.mendocina*	*P.straminea*	*P.stutzeri*
		14164	14160	14159	12689	12693	3904	16637	14162	16665	14165
S10	L22	**11912.0**	**11912.0**	11893.9	**11912.0**	**11912.0**	**11912.0**	**11912.0**	**11912.0**	**11912.0**	11897.9
	L23	*10900.7*	**10945.7**	10955.6	10950.7	**10945.7**	**10945.7**	*10900.7*	11015.7	11085.8	10920.6
	L29	**7173.3**	**7173.3**	7215.4	7202.4	**7173.3**	**7173.3**	**7173.3**	7205.4	7215.4	7274.4
	S10	**11753.6**	**11753.6**	*11783.6*	11767.6	**11753.6**	**11753.6**	**11753.6**	*11783.6*	11755.6	**11753.6**
	S17	**9902.5**	*9966.6*	9957.6	9955.6	*9966.6*	9984.6	**9902.5**	9974.6	10014.7	9973.6
	S19	**10218.1**	10246.1	10186.0	10227.1	10189.1	10204.0	**10218.1**	10176.0	10190.0	10163.0
spc	L14	**13410.9**	**13410.9**	*13396.8*	13412.9	**13410.9**	**13410.9**	**13410.9**	*13396.8*	**13410.9**	13436.9
	L18	12497.4	12556.4	12561.4	12531.4	**12512.4**	**12512.4**	12485.4	12413.3	12457.3	12477.3
	L24	**11330.2**	*11336.3*	11340.3	11471.5	*11336.3*	11345.3	**11330.2**	11344.3	11344.3	11413.4
	L30	6334.5	**6395.6**	6278.3	6347.4	**6395.6**	**6395.6**	6292.5	6363.5	6448.5	6463.6
	L36	**4435.4**	**4435.4**	4407.3	**4435.4**	**4435.4**	**4435.4**	**4435.4**	**4435.4**	**4435.4**	4421.4
	S08	13845.1	13962.3	13951.3	14040.4	13920.2	13973.3	13861.1	13928.2	13914.2	13869.1
	S14	11259.3	**11304.3**	11394.3	11435.3	**11304.3**	11274.2	11288.3	11359.2	11385.3	11326.2
alpha	S11	13529.5	13485.4	13517.5	**13513.5**	13499.4	**13513.5**	13543.5	13531.5	13493.4	13527.5
	S13	13126.3	13210.4	13177.4	13135.2	13164.5	13239.4	13140.3	13118.3	13058.3	13176.4

Correction

Observed *m/z* data from step 1



Construction of biomarkers list with accurate *m/z* (working database)

	P.putida	*P.fulva*	*P.fluorescens*	*P.azotoformans*	*P.chlororaphis*	*P.aeruginosa*	*P.mendocina*	*P.straminea*	*P.stutzeri*	*P.alcaligenes*
	NBRC 14164	NBRC 16637	NBRC 14160	NBRC 12693	NBRC 3904	NBRC 12689	NBRC 14162	NBRC 16665	NBRC 14165	NBRC 14159
L22	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11897.92	11893.91
L23	10900.65	10900.65	10945.74	10945.74	10945.74	10950.71	11015.74	11085.83	10920.64	10955.60
L29	7173.31	7173.31	7173.31	7173.31	7173.31	7202.35	7205.35	7215.39	7274.42	7215.44
S10	11753.58	11753.58	11753.58	11753.58	11753.58	11767.61	11783.61	11755.55	11753.58	11783.61
S17	9902.53	9902.53	9966.58	9966.58	9984.61	9955.60	9974.55	10014.66	9973.57	9957.61
S19	10218.07	10218.07	10246.12	10189.07	10204.04	10227.08	10175.98	10190.01	10162.99	10186.01
L18	12497.41	12485.36	12556.43	12512.38	12512.38	12531.43	12413.29	12457.30	12477.34	12561.35
L24	11330.24	11330.24	11336.25	11336.25	11345.26	11471.46	11344.27	11344.27	11413.37	11340.32
L30	6334.54	6292.46	6395.60	6395.60	6395.60	6347.44	6363.48	6448.54	6463.62	6278.33
L36	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4421.36	4407.34
S08	13845.12	13861.12	13962.26	13920.23	13973.29	14040.42	13928.16	13914.18	13869.10	13951.25
S14	11259.27	11288.27	11304.32	11304.32	11274.24	11435.25	11359.15	11385.32	11326.23	11394.33
S11	13529.50	13529.50	13485.44	13485.44	13499.47	13499.47	13517.48	13479.39	13513.50	13503.46
S13	13126.31	13140.34	13210.43	13164.45	13239.43	13135.23	13118.33	13058.28	13176.42	13177.40

Figure 12.2 (Continued)

Step 5: Typing of isolates based on the selected biomarkers.

working data base

	P.putida NBRC 14164	*P.fulva* NBRC 16637	*P.fluorescens* NBRC 14160	*P.azotoformans* NBRC 12693	*P.chlororaphis* NBRC 3904	*P.aeruginosa* NBRC 12689	*P.mendocina* NBRC 14162	*P.straminea* NBRC 16665	*P.stutzeri* NBRC 14165	*P.alcaligenes* NBRC 14159
L22	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11911.95	11897.92	11893.91
L23	10900.65	10900.65	10945.74	10945.74	10945.74	10950.71	11015.74	11085.83	10920.64	10955.60
L29	7173.31	7173.31	7173.31	7173.31	7173.31	7202.35	7205.35	7215.39	7274.42	7215.44
S10	11753.58	11753.58	11753.58	11753.58	11753.58	11767.61	11783.61	11755.55	11753.58	11783.61
S17	9902.53	9902.53	9966.58	9966.58	9984.61	9955.60	9974.55	10014.66	9973.57	9957.61
S19	10218.07	10218.07	10246.12	10189.07	10204.04	10227.08	10175.98	10190.01	10162.99	10186.01
L18	12497.41	12485.36	12556.43	12512.38	12512.38	12531.43	12413.29	12457.30	12477.34	12561.35
L24	11330.24	11330.24	11336.25	11336.25	11345.26	11471.46	11344.27	11344.27	11413.37	11340.32
L30	6334.54	6292.46	6395.60	6395.60	6395.60	6347.44	6363.48	6448.54	6463.62	6278.33
L36	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4435.39	4421.36	4407.34
S08	13845.12	13861.12	13962.26	13920.23	13973.29	14040.42	13928.16	13914.18	13869.10	13951.25
S14	11259.27	11288.27	11304.32	11304.32	11274.24	11435.25	11359.15	11385.32	11326.23	11394.33
S11	13529.50	13529.50	13485.44	13485.44	13499.47	13499.47	13517.48	13479.39	13513.50	13503.46
S13	13126.31	13140.34	13210.43	13164.45	13239.43	13135.23	13118.33	13058.28	13176.42	13177.40

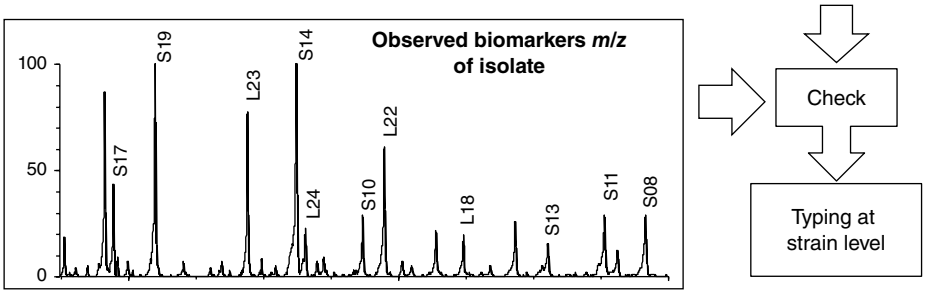


Figure 12.2 (Continued)

N-end rule" [60,61]. According to this rule, therefore, the theoretical mass is calculated by deducing the mass of methionine ($m/z = 131.19$) from the calculated ionized mass (m/z) of any relevant ribosomal protein (Figure 12.3). In some cases, methylation and/or acetylation of amino acid are considered as possible post-translational modifications. After each in silico–calculated m/z of the respective ribosomal protein is completed by the information of amino acid sequences, the candidate biomarkers are selected on the basis of the in silico–calculated m/z database.

Step 4: Construction of accurate database, working database, by a comparative analysis of in silico–calculated m/z and observed m/z and confirmation of reliable biomarker proteins: The in silico–calculated m/z of the candidate biomakers is confirmed and corrected by the comparison with the corresponding observed m/z by the MALDI-TOF MS analysis. The accurate m/z values of the confirmed biomarkers are tabulated as the working database.

Step 5: Typing of isolates based on the selected biomarkers: The working database consists of the confirmed biomarkers that are used as reliable and reproducible biomarkers for the discrimination of the target bacteria, regardless of the sample conditions. The proteotyping of isolates is performed by the results of mass matching profiles of the selected biomarkers using the working database.

In particular, the theoretically supported information of "absence of the selected biomarker m/z" plays a useful role in typing of bacteria because sequence analysis of the

1) Get amino acids sequence by translation of gene sequence encoded in the operon.

<example> Bacterium: *Pseudomonas putida* KT2440 strain
Protein: 50S ribosomal protein L35
Aminno acid sequnce:
MPKMKTKSGAAKRFLKTASGFKHKHAFKSHILTKMSTKRKRQLRGA
SLLHPSDVAKVERMLRVR

2) Calculate theoretical mass of target proteins based on the obtained sequence.

Free software is available to calculate m/z. i.e. Compute pI/Mw tool in ExPASy proteomics server = The obtained theoretical *m/z* of L35 is suspected to *m/z* **7344.88**.

3) Check modification after translation, like methionine loss.

N-terminal end rule: *N*-terminal methionine is selectively cleaved by amino peptidase when the size of branch of neighbor amino acid is less than 1.29 Å.

<example>

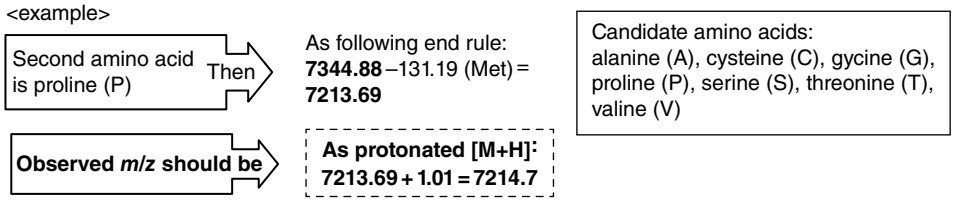


Figure 12.3 Calculation procedure of theoretical masses of ribosomal protein processed by *N*-terminal methionine loss.

biomarker gene can confirm a missing biomarker gene sequence or a point mutation of a putative start codon, such as from ATG to ATA [62].

Therefore, the *S10*-GERMS method clearly explains the rationale for the presence or absence of information on the selected biomarker masses that play a key role in the bacterial discrimination of isolated strains. Not only are their data imported into the available software like PAST (http://folk.uio.no/ohammer/past/, Natural History Museum, Oslo University, Norway) to calculate distance matrices using the neighbor-joining method with the Kimura algorithm, but a phylogenetic tree is also constructed using software like FigTree ver. 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) [63].

12.2.3 Application of Standardized *S10*-GERMS Method to Bacterial Typing

12.2.3.1 Classification of Genus *Pseudomonas* [64,65]

The bacteria of genus *Pseudomonas*, which are gram-negative rod-shaped saprotrophic soil bacteria, occupy an important ecological position in the life-cycle assessment of chemicals because they possess diverse metabolic abilities to assimilate a wide variety of organic compounds and are ubiquitous in nature. Our previous studies showed that some strain of *P. putida* degrade alkylphenol polyethoxylates to estrogenic and antiandrogenic metabolites [66,67]. Although MALDI-TOF MS has been applied especially in clinical laboratories, it was hard to discriminate *P. putida* at the strain level, and little information on the genome-sequenced strain of *P. putida* was available.

To evaluate the usefulness of the *S10*-GERMS method, it was applied for discrimination and classification of *P. putida* at the strain level. Because 17 genome-sequenced strains of genus *Pseudomonas* were available, specific primers were designed on the

basis of the consensus nucleotide sequences of *S10-spc-alpha* operon against 17 genome-sequenced strains. After sequencing the genes of ribosomal proteins coded in *S10-spc-alpha* operon of 10 type strains of genus *Pseudomonas* and following the procedure for the construction of a database in Section 12.2.2, the 14 ribosomal proteins with an *m/z* value lower than 15,000, L18, L22, L23, L24, L29, L30, L36, S08, S10, S11, S13, S14, S17, and S19 were selected as useful biomarkers for phylogenetic classification at the species level among 26 ribosomal proteins coded in the *S10-spc-alpha* operon, because the mass spectra of 14 ribosomal proteins had the reproducibility and the sensitivity of masses based on the *S/N* level by MALDI-TOF MS analysis (Figure 12.4).

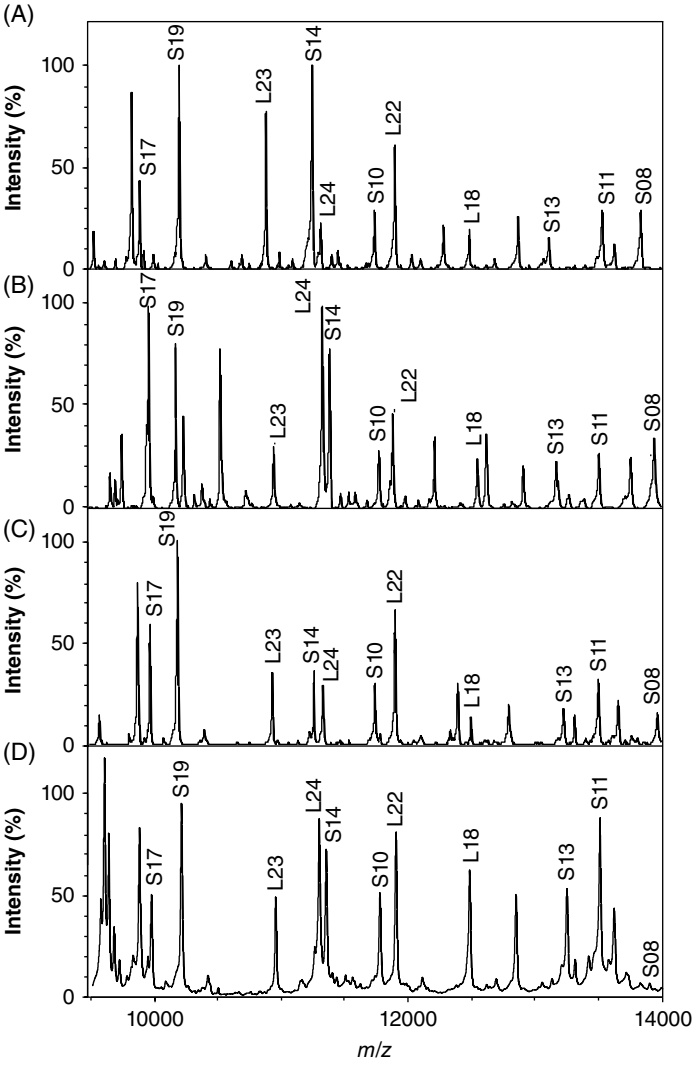


Figure 12.4 Representative MALDI-TOF mass spectra of ribosomal proteins of type strains of genus *Pseudomonas* encoded in *S10-spc-alpha* operon: (A) *P. putida* NBRC 14164^T, (B) *P. alcaligenes* NBRC 14159^T, (C) *P.chlororaphis* NBRC 3904^T, (D) *P. syringae* NBRC 14083.

Phylogenetic analysis based on the amino acid sequence of 14 biomarkers was compared with that based on *gyrB* sequences because the result of *gyrB* sequence analysis correlated very well with DNA–DNA reassociation values determined by DNA hybridization experiments [68,69]. The basic topologies of phylogenetic trees obtained, both of *gyrB* sequences and amino acid sequences of 14 biomarkers, were similar, except for the shift of the phylogenetic position of *P. mendocina* NBRC 14162^T, suggesting that the selected 14 biomarkers were sufficient for bacterial identification by MALDI-TOF MS analysis, and the reclassification of *P. mendocina* species might be required. Because those 14 reliable and reproducible biomarkers were useful for bacterial identification at species levels, the proteotyping and phylogenetic classification of *P. putida* strains whose values of 16S rRNA sequence similarity are more than 99% were carried out using the same 14 biomarkers. The numerator and denominator of the fractions show the number of identical biomarkers and the number of total biomarkers, respectively, indicating the possibility of proteotyping of those closely related strains of *P. putida* (Table 12.1). Because among 14 biomarkers the mass values of S14, S13, and L24 are different in the strains of *P. putida*, those three biomarkers (S14 with *m/z*: 11259.3, 11273.3, and 11289.3; L24 with *m/z*: 11315.2 and 11330.2; S13 with *m/z*: 13096.3, 13112.3, and 13126.3) played a key role in the discrimination of *P. putida* at the strain level (Figure 12.5). The phylogenetic analysis based on the 14 biomarkers gave almost the same results as the phylogenetic trees based on the *gyrB* sequences (Figure 12.6), showing that the *S10*-GERMS method combining genomic and proteomic approaches proved to be a highly reliable and reproducible classification method at the strain level with the capability of validating the obtained results based on the valid masses.

Some species or strains of genus *Pseudomonas* are also important pathogens of animals and plants. Phytopathogenic bacteria, *Pseudomonas syringae*, cause serious damage to crops and remain a crop protection burden because of their variety of pathovars that consist of closely related strains. Therefore, the proteotyping of *P. syringae* at the pathovar level represents a challenging case study to evaluate the usefulness of the *S10*-GERMS method. The difference in the in silico–calculated masses of 5 biomarkers (L24, L30, S10, S14, and S19) out of the 14 selected biomarkers demonstrated that 12 genome-sequenced strains could be discriminated at the pathover level (Table 12.2; JASMS) [65]. Each *m/z* of the candidate theoretical biomarkers was compared with the corresponding *m/z* obtained by MALDI-TOF MS analysis using seven commercially available strains (Table 12.2) [65]. In particular, the ribosomal protein S17 emerged as

Table 12.1 Comparison of the similarities of *Pseudomonas putida* at strain level.

***P. putida* strains**	**NBRC 14671**	**NBRC 15366**	**NBRC 100988**	**JCM 13061**
NBRC 14671	—	99.5[a]	99.6	99.5
NBRC 15366	9/14	—	99.7	99.0
NBRC 100988	11/14[b]	10/14	—	99.2
JCM 13061	11/14	8/14	10/14	—

[a] Values obtained by 16S rRNA sequencing similarities.
[b] The numerator and denominator of fractions show the number of identical biomarkers and the total number of biomarkers, respectively.

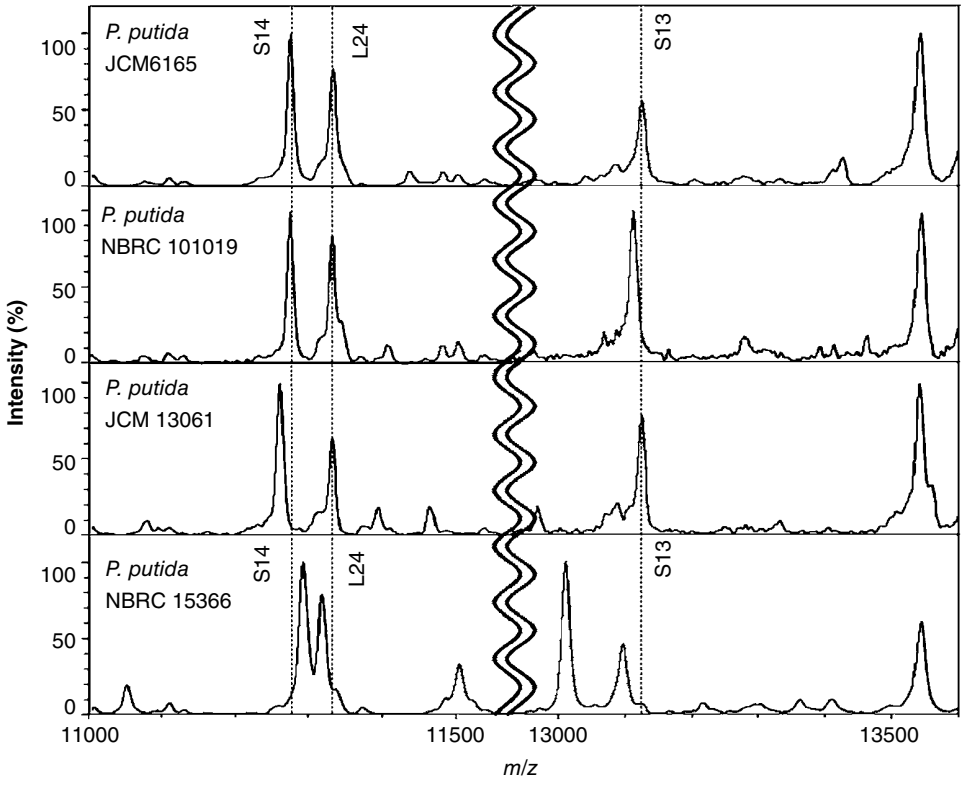


Figure 12.5 MALDI-TOF mass spectra of important biomarkers, S14, L24, and S13 for discrimination of *P. Putida* at the strain level.

a useful biomarker for discrimination of the strain of *P. syringae* though there was no mass difference in 12 genome-sequenced strains. Moreover, MALDI-TOF MS analysis found the additional two ribosomal proteins, S12 and S16, which were not encoded in the *S10-spc-alpha* operon, to be very effective for the discrimination of the strains of *P. syringae* (Table 12.2), which underscores the importance of the comparative verification of the in silico–calculated *m/z* by peak assignment using MALDI-TOF MS analysis.

To process a cluster analysis and/or generate a phylogenetic tree by using the biomarkers, the presence or absence of the mass ions (*m/z*) of the biomarkers was indicated by 1 and 0, respectively, as a binary peak matching profile, especially in fingerprinting analysis. In this case, however, identical ribosomal proteins like L24 had three different masses – for example, *m/z*: 11288.2, 11306.2, 11274.2 – making significantly useful biomarkers for the further typing of *P. syringae* at the pathovar level.

Therefore, when there was a mass difference among the biomarkers, they were categorized as types from I to III based on their corresponding *m/z*. As a rule, type I was first allocated to the ribosomal proteins of the type strain. If the mass of the biomarkers of other strains differed from type I, type II and III were then respectively allocated to the biomarkers. As for L26, type I was for *m/z* 11288.2, type II for *m/z* 11306.2, and type III for *m/z* 11274.2.

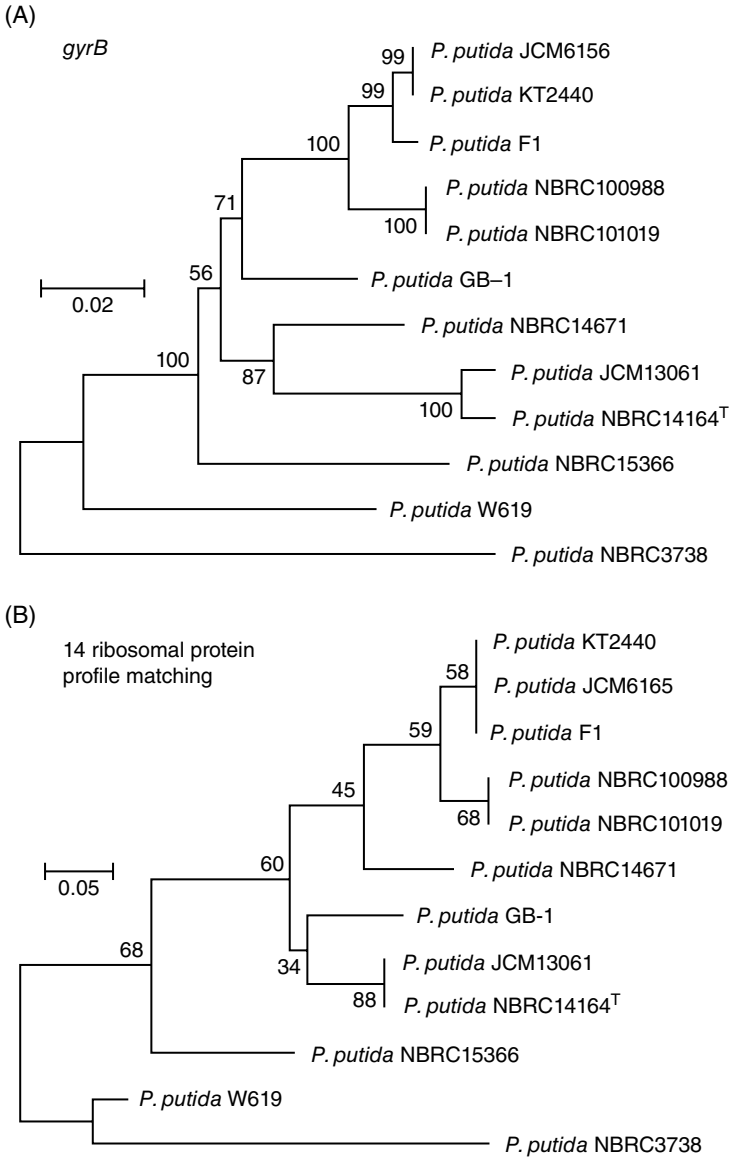


Figure 12.6 Phylogenetic analysis of 12 strains of *P. putida*: (A) phylogenetic analysis of *gyrB* sequences, (B) phylogenetic analysis based on the *S10*-GERMS method.

According to the selected eight biomarkers (L24, L30, S10, S14, S17, S19, S12, and S16), the seven commercially available strains were classified into five clusters as follows (Table 12.2): (1) NBRC 3310, (2) NBRC 3508 and NBRC 12656, (3) NBRC 12655, (4) NBRC 14053 and NBRC 14083, and (5) NBRC 14084.

Further, identical pathovars like *P. syringae* pv. Mori may be divided into subgroups using pathovar-specific biomarkers like S17, suggesting the strong possibility of future application of the *S10*-GERMS method to this problem.

Table 12.2 Ribosomal protein profiling table of *P. syringae* strains.

		Ribosomal protein types															
Pathover	**Strain**	**L36**	**L30**	**L29**	**S17**	**S19**	**L23**	**L24**	**S14**	**S10**	**L22**	**L18**	**S13**	**S11**	**S08**	**S16**	**S12**
Genome-sequenced strains																	
Aesculi	2250	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
Aesculi	NCPPB3681	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
Phaseolicola	1448A	I	I	I	I	I	I	I	II	III	I	I	I	I	I	I	I
Tabaci	ATCC 1 1528	I	I	I	I	I	I	I	II	II	I	I	I	I	I	I	I
Syringae	B728a	I	I	I	I	I	I	II	III	II	I	I	I	I	I	I	I
Syringae	642	I	II	I	I	I	I	II	II	II	I	I	I	I	I	I	I
Oryzae	1_6	I	I	I	I	I	I	III	II	II	I	I	I	I	I	-	I
Tomato	K 40	I	I	I	I	II	I	I	II	II	I	I	I	I	I	I	II
Tomato	Max 13	I	I	I	I	II	I	I	II	II	I	I	I	I	I	II	I
Tomato	NCPPB 1 108	I	I	I	I	II	I	I	II	II	I	I	I	I	I	II	I
Tomato	DC3000	I	I	I	I	II	I	I	II	II	I	I	I	I	I	II	I
Tomato	T1	I	III	I	I	II	I	I	II	II	I	-	I	I	I	II	I
Sample strains																	
Coronafaciens	NBRC3310	I	I	I	I	I	I	III	II	I	I	I	I	I	I	II	I
Tabaci	NBRC3508	I	I	I	I	I	I	I	II	I	I	I	I	I	I	I	I
Aptata	NBRC12655	I	I	I	I	I	I	II	II	I	I	I	I	I	I	I	I
Phaseolicola	NBRC12656	I	I	I	I	I	I	I	II	I	I	I	I	I	I	I	I
Mori	NBRC14053	I	I	I	I	I	I	I	II	I	I	I	I	I	I	II	III
Mori	NBRC14083	I	I	I	I	I	I	I	II	I	I	I	I	I	I	II	III
Mori	NBRC14084	I	I	I	II	I	I	I	II	I	I	I	I	I	I	II	I

Those biomarkers become relevant to the proteotyping of bacteria only if genomic data are combined with proteomic data by MALDI-TOF MA analysis.

Because the *S10*-GERMS method illustrates the rapid discrimination of *P. syringae* at the pathovar level, it may be a powerful tool for plant disease diagnosis caused by not only *P. syringae* but also by other plant pathogenic bacteria to overcome the burden of crop protection.

12.2.3.2 Classification of Genus *Sphingomonaceae* [70]

The biodegraded products of organic compounds may have a direct deleterious effect on wildlife when they have significant toxicity. Therefore, identification of bacteria is particularly important in understanding the dynamic relationship between microbial diversity and the microbial capacity for the biodegradation of organic compounds in the environment. The genus *Sphingomonas*, which consists of strictly aerobic, chemoheterotrophic, yellow-pigmented, gram-negative, rod-shaped bacteria [71], was reclassified into four genera: *Sphingomonas*, *Sphingobium*, *Novosphingobium*, and *Sphingopyxis* [72].

Although it has been reported that sphingomonad species can degrade a wide range of aromatic hydrocarbons [73], only one strain of the genus *Sphingopyxis* had been completely sequenced by December 1, 2011. Therefore, the *S10*-GERMS method was applied for the discrimination of strains of *Sphingomonadaceae* by using the consensus nucleotide sequences of *S10* and *spc* operons from seven genome-sequenced strains of *Sphingomonaceae* [70], which provide useful information for the design of specific primers. The ribosomal protein database was constructed by both sequencing the operons and MALDI-TOF MS analysis of 16 available type strains of *Sphingomonadaceae* following the procedures in Section 12.2.2.

The nine selected ribosomal proteins as biomarkers coded in the *S10-spc* operon, L18, L22, L24, L29, L30, S08, S14, S17, and S19, were significantly useful for bacterial classification to identify $APEO_n$-degrading bacteria (Table 12.3). Although there was no significant difference in the relatedness of the phylogenetic trees based on the *S10*-GERMS method and the 16S rRNA gene sequences, which formed four genera clusters of the *Sphingomonadaceae*, respectively, minor difference between two phylogenetic trees could be observed in the phylogenetic position of *Sphingomonas jaspsi* NBRC 102120^T and *Sphingomonas wittichii* NBRC 105917^T (Figure 12.7A, B). In the case of 16S rRNA gene sequences, *Sphingomonas wittichii* NBRC 105917^T was closely clustered into the genus *Sphingobium* and not the genus *Sphingomonas*, suggesting that more research into the *Sphingomonadaceae* may be required.

Nonionic surfactant alkylphenol polyethoxylates ($APEO_n$) were degraded to endocrine disruptors in the environment [74–76]. In the life-cycle impact assessment of synthetic chemicals, bacteria play a key role in their biodegradation. Seven strains of the genus *Sphingopyxis* and one strain of the genus *Sphingobium* identified on the basis of the 16S rRNA gene sequence were isolated as $APEO_n$-degrading bacteria in our laboratory. Because the *S10*-GERMS method was proved to be efficacious for the discrimination of the strains of *Sphingomonadaceae*, nine selected biomarkers were applied for the typing of the $APEO_n$ -degrading isolates.

The 16S rRNA sequence identity between the $APEO_n$-degrading bacteria strain BSN20 and *Sphingopyxis terrae* NBRC 15098^T was 99.9%, meaning that the difference in the 16S rRNA gene sequence was only one base among about 1500 bp; however, the

Table 12.3 Theoretical masses of nine selected ribosomal subunit proteins of *Sphingomonacecae*.

	Subunit No.								
Strains	**L30**	**L29**	**S19**	**S17**	**L24**	**S14**	**L18**	**L22**	**SO 8**
Genus *Sphingopyxis*									
BSN11	6789.2	7536.5	9978.6	10434.2	11085.8	11545.6	12393.2	13701.9	14577.7
NBRC 15033	6789.2	7536.5	9978.6	10434.2	11085.8	11545.6	12393.2	13701.9	14577.7
BSN51	6770.1	7536.5	9978.6	10505.2	11058.8	11545.6	12377.2	13701.9	14566.7
JCM 11457	6789.2	7461.5	9978.6	10205.9	10990.7	11540.6	12377.2	13700.8	14607.8
BSN54	6743.0	7445.5	9978.6	10405.2	11072.8	11545.6	12333.2	13701.9	14607.8
JCM 14844	6741.1	7514.6	9978.6	10834.6	10976.7	11527.6	12329.2	13715.9	14591.8
BSN22	6784.1	7523.5	9978.6	10565.2	11072.8	11513.6	12393.2	13715.9	14577.7
BSN48	6784.1	7523.5	9978.6	10565.2	11072.8	11513.6	12393.2	13715.9	14577.7
BSN53	6869.2	7487.5	9948.5	10269.9	11056.8	11499.6	12422.2	13701.9	14593.7
JCM 15910	6869.2	7459.5	9948.5	10269.9	11056.8	11499.6	12309.0	13701.9	14579.7
BSN20	6846.2	7471.5	9948.5	9833.5	10962.7	11527.6	12403.2	13779.9	14621.8
NBRC 15098	6846.2	7471.5	9948.5	9833.5	10962.7	11513.6	12403.2	13779.9	14621.8
JCM 14161	6565.8	8726.9	10054.5	11326.0	11330.2	11645.7	12320.2	13697.8	14231.2
Genus Novosphingobium									
NBRC 16086	6213.5	7665.6	10013.5	9893.6	11179.9	11408.3	12270.1	13440.7	14062.2
Genus Sphingobium									
NBRC 16172	6224.5	7629.6	10048.6	9838.4	11030.8	11502.5	12510.4	13793.0	14266.4
NBRC 16415	6395.7	7587.6	10064.6	9838.4	10972.8	11502.5	12219.1	13774.9	14192.3
NBRC 102517	6841.2	7601.6	9963.5	9865.4	11004.8	11532.5	12440.3	13778.9	14206.3

NBRC 15102	6425.8	7601.6	10161.8	9824.3	10984.8	11529.6	12460.3	13744.9	14194.3
NBRC 105917	6751.9	7381.5	9997.5	9837.3	11025.8	11571.5	12490.2	13605.9	14264.2
Genus Sphingomonas									
NBRC 102120	6354.5	7698.8	9993.7	9777.3	11185.0	11653.7	12113.0	13590.8	14439.5
NBRC 15742	6495.7	8140.3	10095.7	10145.7	10887.7	11447.4	12702.5	13281.5	14233.3
NBRC 16722	6411.5	7590.5	9970.6	10201.7	11022.8	11650.8	12316.1	13228.4	14225.3
NBRC 15099	6508.7	7317.2	9927.5	10056.6	11002.8	11401.4	12268.0	13297.4	14244.3

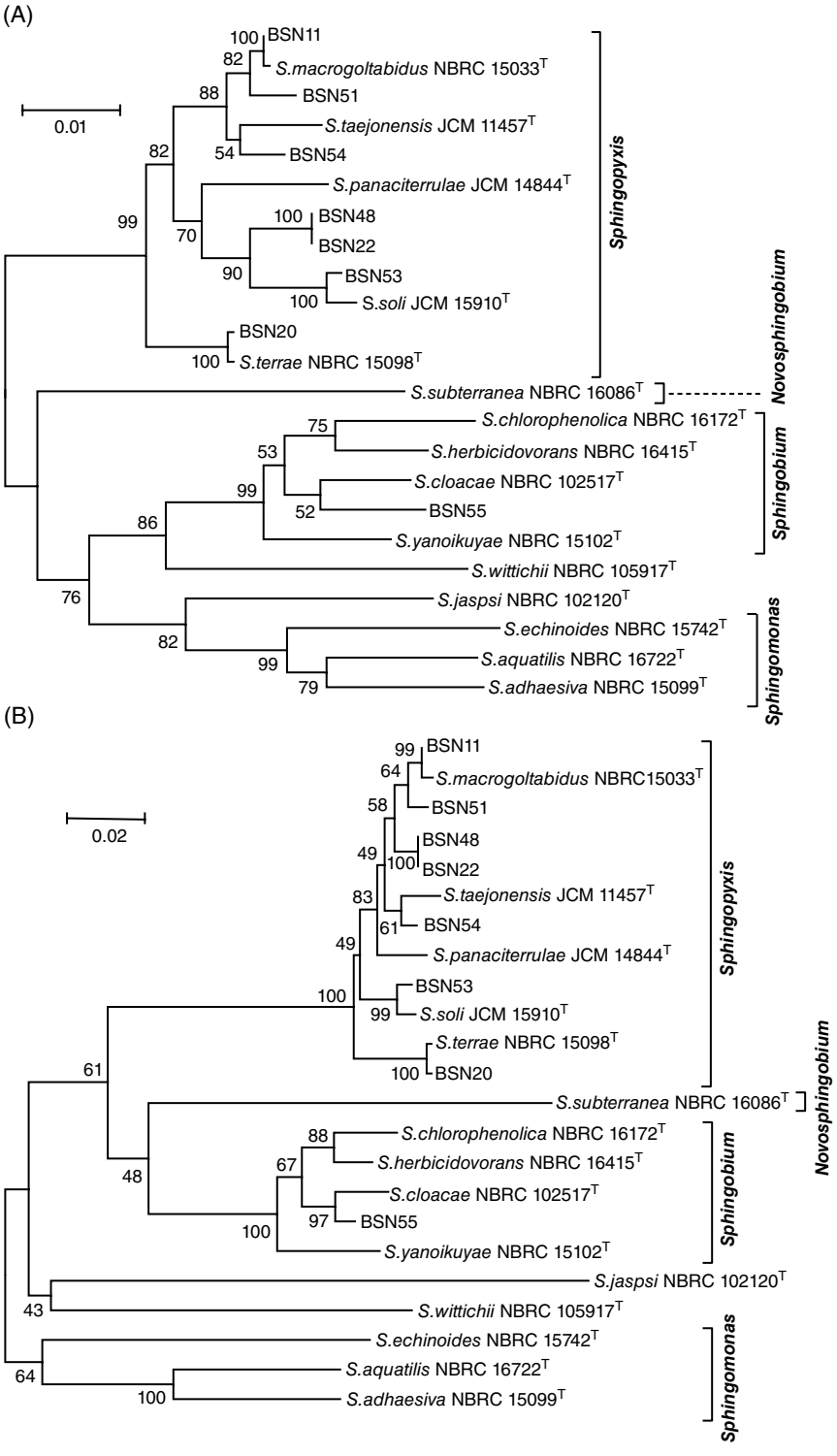


Figure 12.7 Phylogenetic trees of $APEO_n$-degrading bacteria and type strain of genus *Sphingomonas* based on (A) 16S rRNA gene sequences and on (B) the *S10*-GERMS method, and (C) MALDI mass spectra of ribosomal protein S14 of *Sphingopyxis terrae*: NBRC 15098^T, $APEO_n$-degrading bacterium strain BSN20, NBRC 15593, NBRC 15598, and NBRC 15599, respectively.

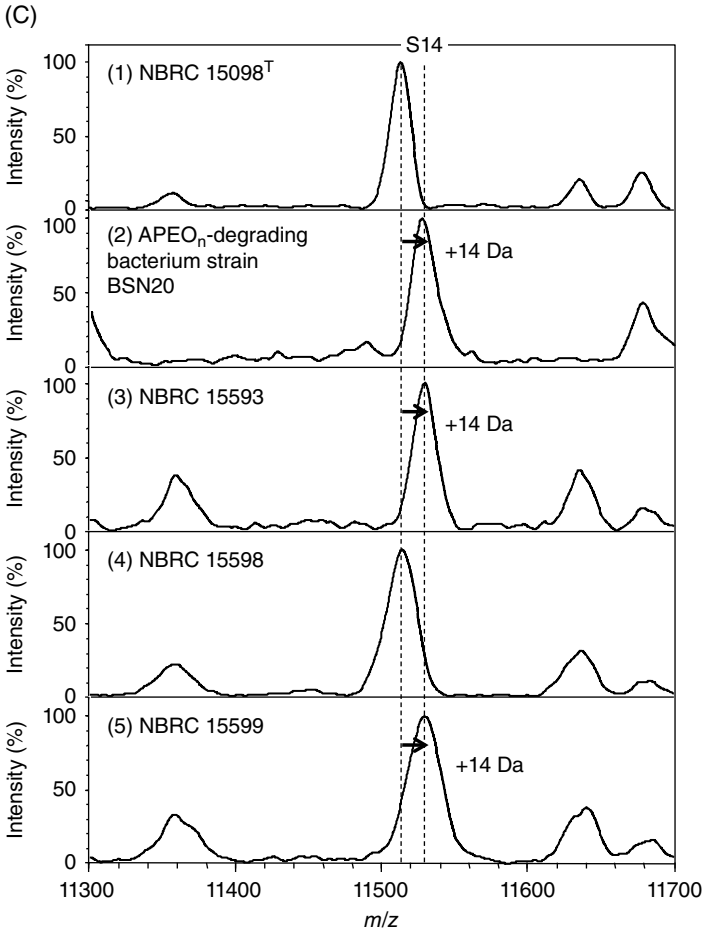


Figure 12.7 (Continued)

MALDI mass spectra revealed that there was a mass difference, especially in the selected biomarker S14, whose *m/z* was 11513.6 or 11527.6, respectively (Figure 12.7C). By using this *m/z* difference, *Sphingopyxis terrae* NBRC 15098^T and the $APEO_n$-degrading bacteria strain BSN20 could be successfully distinguished despite only one base difference in the 16S rRNA gene sequence between them.

Moreover, although the strain NBRC 15593 isolated as polyethylene glycol (PEG)-degrading bacteria is registered as *Sphingopyxis macrogoltabidus* in NBRC, the MALDI mass spectra of the strain NBRC 15593 was not that of the type strain, *Sphingopyxis macrogoltabidus* NBRC 15033^T, but was that of the strain of BSN20. This result was also supported by the 16S rRNA gene sequence analysis, suggesting that the strain NBRC 15593 might be registered as *Sphingopyxis terrae* and not *Sphingopyxis macrogoltabidus* (Figure 12.7C).

The *S10*-GERMS method may become a potent way to type isolates with insufficient genomic information and managing bacterial strains in culture collections.

12.2.3.3 Classification of Genus *Bacillus* [77]

In the gram-positive bacteria of genus *Bacillus, B. subtilis, B. amyloliquefaciens, B. pumilus,* and *B. licheniformis* are used for food processing of fermented soybean foods because of their capability to produce polyglutamic acid (PGA) [78]. Meanwhile, *B. anthracis, B. subtilis, B. mycoides*, *B. licheniformis, B. megaterium,* and *B. cereus* raise conflicting issues for human pathogens and food spoilage [79–81]. The whole genome sequences of 20 strains belonging to *Bacillus anthracis*, *B. atrophaeus*, *B. cereus*, *B. licheniformis*, *B. macerans*, *B. megaterium*, *B. mycoides*, and *B. subtilis* are now available [82]. However, because the genus *Bacillus* is genetically and physiologically diverse, the proteotyping of the whole cell analysis using the *S10*-GERMS method might be of significant value in the management of the genus *Bacillus*. Because the ribosomal proteins of gram-positive bacteria with a rigid cell wall are hard to analyze by MALDI-TOF MS [39], the acid extraction treatment was applied to the genus *Bacillus* for the disruption of the cell walls as usual. In this study, the best signal quality with MALDI-TOF MS analysis was observed for the 30% (v/v) formic acid (FA) treatment rather than for trifluoroacetic acid treatment (Figure 12.8). Due to the FA treatment, in particular, high-molecular-weight ribosomal proteins with masses exceeding an *m/z* of 11,000, such as L18, L22, and S10, coded in *S10* and *spc* operons, were detected with high sensitivity. However, because both S11 and S13 ribosomal proteins coded in *alpha* operon were not found, the PCR primers and sequencing primers were designed on the basis of consensus nucleotide sequences of *S10* and *spc* operons from 13 genome-sequenced strains. The unregistered ribosomal proteins and registration errors of genome-sequenced type strains were corrected by a comparison with in silico–calculated *m/z*

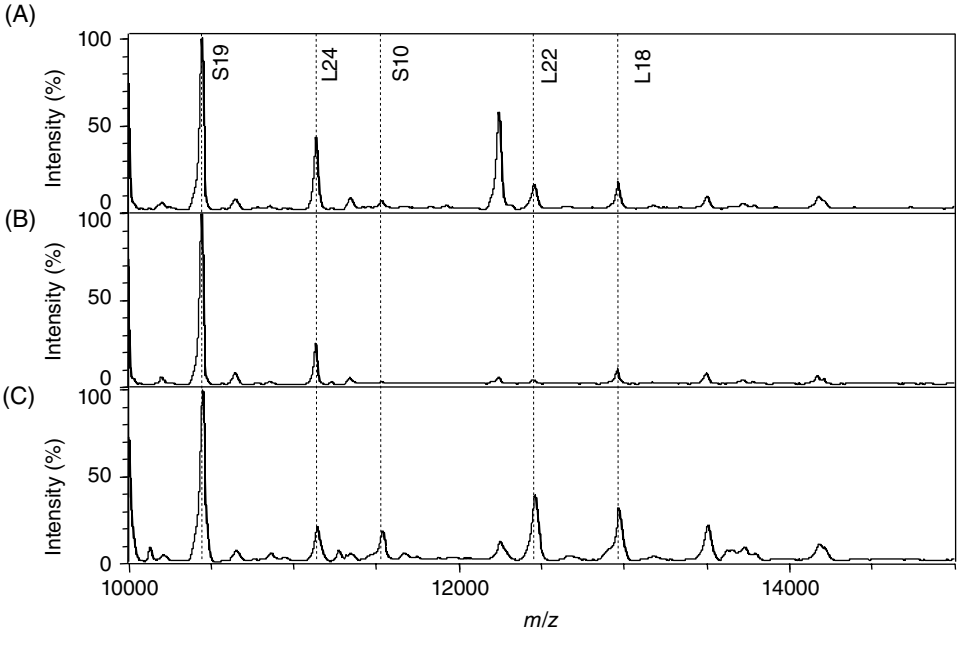


Figure 12.8 MALDI mass spectra of *B. subtilis* subsp. *subtilis* NBRC 13719^T using acid extraction: (A) TFA1.0%, (B) TFA2.5, (C) FA30%.

and the observed *m/z* by MALDI-TOF MS, for example, misannotation of the start codon of L22 and L32 and sequence error of L29 proteins in *B. cereus* NBRC 15305^T; the presence of unregistered L22, L32, L34, S14, S15, and S18 proteins in *B. pumilus* NBRC 12092^T; misannotation of the start codons of L29 and S20 proteins and the presence of an unregistered L24 protein in *B. thuringiensis* NBRC 101235^T; and misannotation of the start codons of L29, S14, and S18 and sequence errors of L22 and S20 proteins, and the presence of an unregistered L24 protein in *B. mycoides* NBRC 101228^T. Although Lauber *et al.* reported that the error of ribosomal protein masses of *B. subtilis* subsp. subtilis NBRC 13719^T was not easily accounted for by post-translational modifications [83], ribosomal proteins of *B. subtilis* subsp. subtilis NBRC 13719^T considering *N*-terminal methionine loss only as a post-translational modification could be assigned by MALDI-TOF MS analysis.

The eight ribosomal subunit proteins, L18, L22, L24, L29, L30, S10, S14, and S19 in *S10-spc* operon, were selected as biomarkers because their mass spectra had mass reproducibility and sensitivity based on the *S/N* level. Although there was no significant difference in the relatedness of the phylogenetic trees based on the *S10*-GERMS method and 16S rRNA gene sequences which formed two clusters, cluster I for the *B. subtilis* group and cluster II for the *B. cereus* group, respectively (Figure 12.9), minor difference between two phylogenetic trees could be observed in the phylogenetic position of *B. vietnamensis* NBRC 101237^T.

In the *S10*-GERMS method, *B. vietnamensis* NBRC 101237^T formed a cluster with *B. badius* NBRC 15713^T and *B. megaterium* NBRC 15308^T, whose position was linked to the *B. subtilis* group, whereas the phylogenetic position of *B. vietnamensis* NBRC 101237^T in the 16S rRNA gene sequences was included in cluster I.

Noguchi *et al.* also reported that *B. vietnamensis* formed a cluster with *B. aquimaris* and *B. marisflavi*, whose cluster was linked to the *B. subtilis* group [84]. However, the phylogenetic tree constructed by Cerritos *et al.* clearly separated them [85]. Therefore, the discussion of the phylogenetic position of *B. vietnamensis* NBRC 101237^T may require more information on the species of genus *Bacillus*.

Although three subspecies are known in *B. subtilis*: *B. subtilis* subsp. *subtilis*, *B. subtilis* subsp. *spizizenii*, and *B. subtilis* subsp. *inaquosorum* [86,87], the 16S rRNA gene sequence identity between *B. subtilis* subsp. *subtilis* NBRC 13719^T and *B. subtilis* subsp. *spizizenii* NBRC 101239^T is 99.8% (1473/1475 bases), which shows 63% and 67% of DNA–DNA relatedness value [86]. Although this two-base difference in the 16S rRNA gene sequence makes it difficult to discriminate *B. subtilis* at the subspecies level, in the *S10*-GERMS method, the difference in the masses of eight biomarkers indicated the possibility of discriminating *B. subtilis* at the subspecies level, according to the binary peak matching profile (Table 12.4).

In particular, biomarker L29 played a pivotal role in typing between *B. subtilis* subsp. *subtilis* NBRC 13719^T and *B. subtilis* subsp. *spizizenii* NBRC 101239^T. Moreover, both ribosomal proteins of L18 and L22 were useful biomarkers for discrimination of *B. subtilis* at the strain level. The results of the phylogenetic trees of *B. subtilis* based on eight ribosomal proteins by the NJ method using *B. subtilis* subsp. *subtilis* NBRC 13719^T as a reference strain indicated the clear discrimination of *B. subtilis* at the subspecies and strain levels rather than 16S rRNA gene analysis (Figure 12.10). As a consequence, despite the fact that identification of the isolated bacteria on the basis of the fingerprint is only possible inside the frame of the available reference database, only eight selected

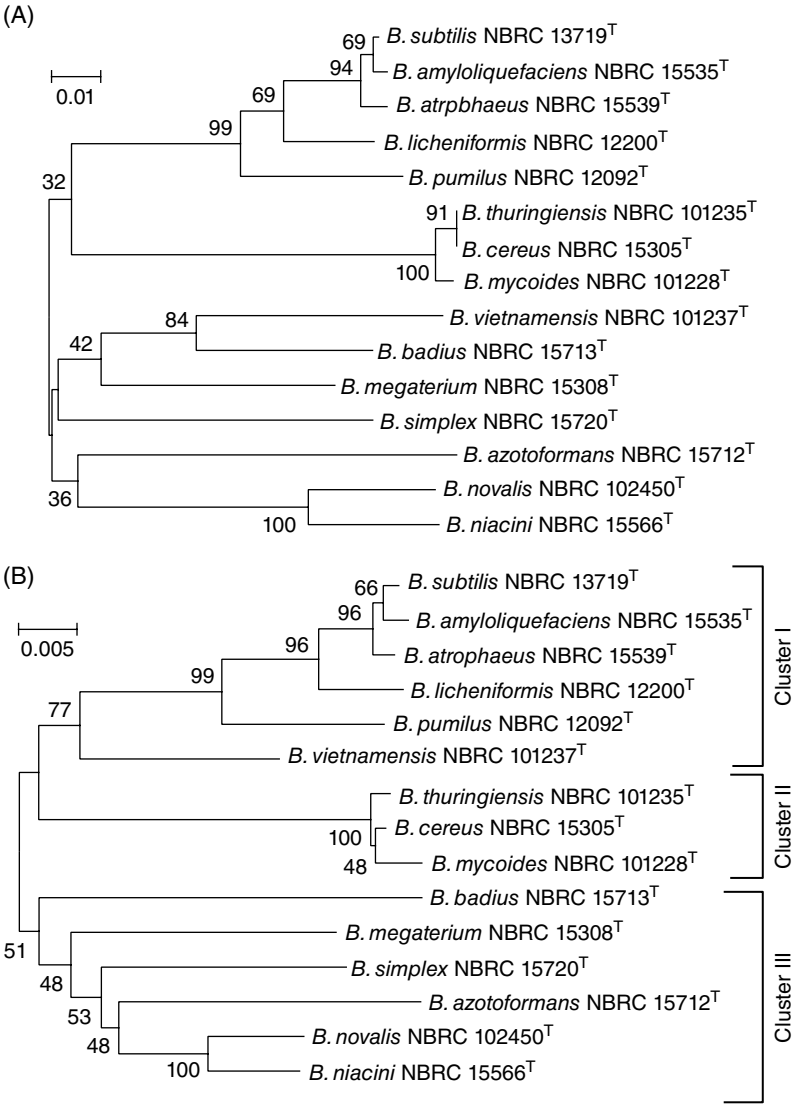


Figure 12.9 Phylogenetic trees of type strains of genus *Bacillus* strains: phylogenetic tree based on amino acid sequences of eight selected ribosomal proteins in (A) *S10-spc* operon, (B) phylogenetic tree based on 16S rRNA gene sequence.

reliable and reproducible biomarkers were significantly useful for rapid bacterial classification of genus *Bacillus* strains at the species and strain levels with the *S10*-GERMS method.

12.2.3.4 Characterization of the *Lactobacillus casei* Group [88]

Although the accumulating scientific knowledge on lactic acid bacteria (LAB) has been accelerating the development of probiotic functional foods even in pharmaceuticals, the phylogenetic similarity and controversial nomenclatural status of lactobacillus

Table 12.4 Binary peak matching profile of *B. subtilis* strains.

Strains	Reference peaks																			
NBRC No.	**L36**	**L34**	**L30**	**L32**	**L28**	**S14**	**L35**	**L29**	**S18**	**S20**	**S16**	**S15**	**S19**	**L24**	**S10**	**L22**	**L18**	**L20**	**S09**	**S12**
13719 T	1[a]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14415	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14474	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
101588	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
101246	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	0	1	1	0
13722	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1
101239	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	0	1	1
104440	1	0	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	1	0

[a] Present peaks are denoted by 1 and absent peaks by 0. Bold: Eight ribosomal subunit proteins selected for non-genome-sequenced genus *Bacillus*.

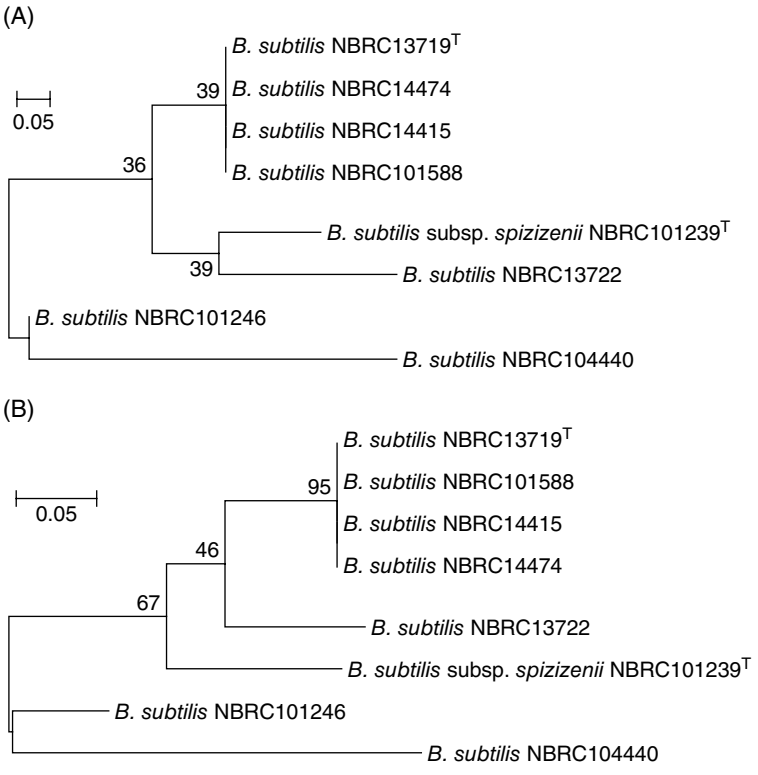


Figure 12.10 Phylogenetic trees of *B. subtilis* strains based on ribosomal protein profile matching using ribosomal protein of *B. subtilis* subsp. *subtilis* NBRC 13719^T as a reference strain: (A) phylogenetic tree based on eight ribosomal protein biomarkers, (B) phylogenetic tree based on 20 ribosomal protein biomarkers.

probiotic are responsible for the complicated taxonomy of *Lactobacillus casei* and related organisms (the *L. casei* group) [89,90]. This is because the current taxonomy of the *L. casei* group is still based on the proposal by Collins *et al.* in 1989 that *L. casei* comprises three species: *L. casei* (containing ATCC 393 as the type strain), *L. paracasei* with two subspecies (subsp. paracasei and subsp. tolerans), and *L. rhamnosus* [91].

Because there were no genome-sequenced type strains of the *L. casei* group with a finalized status, ribosomal protein genes were sequenced by using specific primers designed on the basis of consensus nucleotide sequences of *S10-spc-alpha* operons from 28 genome-sequenced strains, and a set of theoretical masses was calculated and corrected for the ribosomal protein database by using the construction procedure (Section 12.2.2). For MALDI-TOF MS analysis to characterize 33 strains of the *L. casei* group, each bacterial culture was mashed with a Mini Bead-Beater 8 (Biospec, Bartleville, OK, USA), and the resulting lysate was mixed with the matrix solution at a concentration of 10 mg/ml sinapinic acid in 50% acetonitrile with 1% trifluoroacetic acid solution.

Only four ribosomal proteins (L22, L36, S11, and S19) were common to the *L. casei* group, and these were identical only within the *L. casei* group strains by the results of a BLASTp search, making them potentially useful biomarkers for categorizing the *L. casei* group.

Fourteen ribosomal proteins, L14, L18, L22, L23, L24, L29, L30, L36, S8, S11, S13, S14, S17, and S19, were used as biomarkers to construct a phylogenetic tree of the *L. casei* group. When identical ribosomal protein had a different *m/z*, it was categorized into five types (types I to V) based on their own *m/z* to characterize ribosomal protein as the biomarker; for example, in the case of L23: type I for 11564.4, type II for 11561.4, type III for 11589.5, and type IV for 11533.4 (Table 12.5). The result of the UPGMA cluster analysis based on ribosomal protein profiling by the *S10*-GERMS method could classify the strains of the *L.casei* group into four major clusters that had each type strain in the clusters (Figure 12.11). Cluster 1 contained the type strain of *L.casei* JCM 1134^T. Due to a significant contribution of L23-IV, L24-IV, and S14-IV to the discrimination of the species *L. rhamnosus*, all strains of *L. rhamnosus* formed a single group as cluster 2, which is also supported by the result of ribotyping [92]. Cluster 3 reflected the history of the *L. casei* group because it contained almost all the strains of *L. paracasei* subsp. *paracasei* (except for JCM 1172), six strains of *L. casei*, and one strain of *L. acidophilus* JCM 20315. The results of the *S10*-GERMS method suggested that strains of JCM 20024, JCM 20304, and JCM 20315 should be correctly reidentified as *L. paracasei* subsp. *paracasei* as well as ATCC 334 and four strains of *L. casei*, BD-II [93], BL23 [94], LC2W [95], and Zhang [96], which have sequences that are nearly identical with ATCC 334. Cluster 4, defined by the two biomarkers L23-III and L24-III, had two strains: *L. paracasei* subsp. *tolerans* JCM 1171^T and *L. paracasei* subsp. *paracasei* JCM 1172. Because the phylogenetic position of *L. paracasei* subsp. *paracasei* JCM 1172 was also supported by the result of ribotyping and PCR fingerprint [92,97], *L. paracasei* subsp. *paracasei* JCM 1172 should be revised to *L. paracasei* subsp. *tolerans* on the basis of the cumulative scientific evidence.

The *S10*-GERMS method may shed new light on and help resolve the controversy arising from the nomenclatural status of the *L. casei* group that has persisted since 1989.

12.2.3.5 Characterization of Enterohemorrhagic *Escherichia coli* [98]

Shiga toxin-producing *Escherichia coli*, known as enterohemorrhagic *E. coli* (EHEC), is an important pathogen that causes life-threatening infection in humans worldwide. In particular, *E. coli* serotype O157:H7 is the most commonly reported EHEC serotype responsible for a large number of outbreaks.

Because MALDI-TOF MS is a rapid, cost-effective, and robust approach for the identification of microorganisms, this method has been rapidly developed and successfully applied for the diagnosis of infectious bacteria, especially in clinical microbiology.

Although MALDI-TOF MS has been used to classify EHEC serotypes, the results obtained by the fingerprinting method are not very reliable because of the high similarity of their spectra from isolates and less reproducible spectra depending on experimental factors like sample preparation and/or culture conditions [99]. To respond to the growing expectations for a specific diagnostic demand, the parameters need to be customized in an optimal way when using commercially available software like BioTyper software (Burker Daltonics, Germany) [30,100].

Although the identification of specific biomarkers for EHEC serotypes has emerged as a formidable challenge to overcome the limitation of application of MALDI-TOF MS, *E. coli* O157:H7-specific biomarkers HdeA, HdeB, CspC, YbgS, YjbJ, and YbgO were identified using MALDI-TOF/TOF-MS/MS [101], providing a precious clue to the typing of EHEC serotypes. Therefore, an effort was made to apply the *S10*-GERMS method for the proteotyping of EHEC serotypes.

Table 12.5 Peak pattern of selected biomarkers for the *L. casei* group.

		Biomarker types [a] and theoretical masses as $[M+H]^+$				
Protein name	**Coded operon**	***L. casei* JCM 1134^T**	***L. paracasei* subsp. *paracasei* JCM 8130^T**	***L. paracasei* subsp. *tolerans* JCM 1171^T**	***L. rhamnosus* JCM 1136^T**	***'L. zeae'* JCM11302**
L22	S10	I (12599.5)	I (12599.5)	I (12599.5)	I (12599.5)	I (12599.5)
L23	S10	I (11564.4)	II (11561.4)	III (11589.5)	IV (11533.4)	I (11564.4)
L29	S10	I (7578.9)	II (7479.7)	II (7479.7)	I (7578.9)	I (7578.9)
S17	S10	I (9985.5)	II (10071.7)	II (10071.7)	I (9985.5)	I (9985.5)
S19	S10	I (10447.0)	I (10447.0)	I (10447.0)	I (10447.0)	I (10447.0)
L14	spc	I (13031.2)	II (13036.2)	II (13036.2)	II (13036.2)	I (13031.2)
L18	spc	I (13015.8)	II (13001.8)	II (13001.8)	I (13015.8)	I (13015.8)
L24	spc	I (11231.1)	I (11231.1)	III (11203.0)	IV (11247.1)	I (11231.1)
L30	spc	I (6661.8)	II (6618.8)	II (6618.8)	I (6661.8)	I (6661.8)
S08	spc	I (14676.1)	II (14644.0)	II (14644.0)	II (14644.0)	I (14676.1)
S14	spc	I (6992.3)	I (6992.3)	I (6992.3)	IV (7006.4)	I (6992.3)
L36	alpha	I (4449.5)	I (4449.5)	I (4449.5)	I (4449.5)	I (4449.5)
S11	alpha	I (13641.6)	I (13641.6)	I (13641.6)	I (13641.6)	I (13641.6)
S13	alpha	I (13426.6) [b]	II (13360.5)	II (13360.5)	I (13426.6) [b]	V (13412.6)

[a] Type I was first allocated to the ribosomal proteins of *L. casei* JCM 1134^T. If the mass of the biomarkers of *L. paracasei* subsp. *paracasei* JCM 8130^T differed from type I, type II was then allocated to the biomarker. In this way, types III to V were respectively allocated to the characteristic ribosomal proteins of *L. paracasei* subsp. *tolerans* JCM 1171^T, *L. rhamnosus* JCM 1136^T, and *L. casei* ('*L. zeae*') JCM 11302.

[b] Amino acid sequences of S13 of *L. casei* JCM 1134^T and *L. rhamnosus* JCM 1136^T are different, but have the same mass.

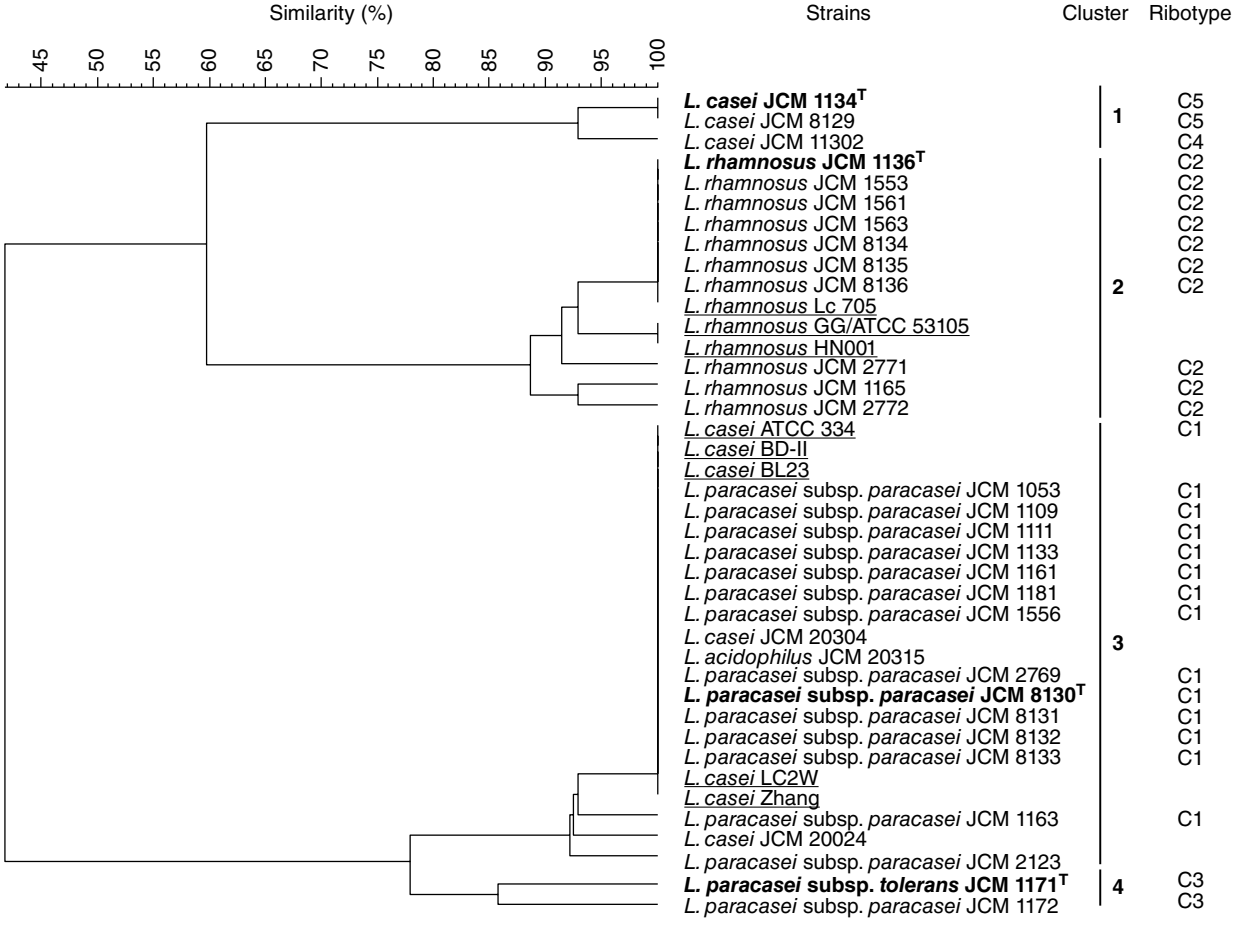


Figure 12.11 The UPGMA cluster analysis of the *L. casei* group based on ribosomal protein biomarkers. Underlined strains are genome-sequenced. Clusters were divided at a similarity of 80%. The result of ribotyping reported by Ryu *et al.* [92] is used as a reference.

The in silico–calculated masses of ribosomal proteins encoded by the *S10-spc-alpha* operon were calculated on the basis of sequence analysis and were corrected by the actual analytical results of MALDI-TOF MS (Table 12.6).

The masses of the *S10-spc-alpha* operon-encoded ribosomal proteins, namely, S10, L3, L4, L23, L2, S19, L22, S3, L16, L29, S17, L14, L5, S14, S8, L6, L18, S5, L30, L36, S13, S11, S4, and L17, were all identical respectively in all of the *E. coli* strains used for database construction, making them potentially useful biomarkers for categorizing *E. coli*.

Because the candidate biomarkers like L24, S5, and S13 encoded in the *S10-spc-alpha* operon gave unclear peaks because of small differences in masses or high molecular weights, in the case of *E. coli*, strain typing or serotyping using ribosomal proteins encoded in the *S10-spc-alpha* operon appears to be unsuitable for biomarker mines because of a lower diversity of masses. However, the ribosomal proteins S15 and L25 gave unique and clear mass shifts specific in *E. coli* O157 compared with the other *E. coli* serotypes (Figure 12.12 and Table 12.6).

Sequence analysis revealed that the O157-specific point mutation A239G on ribosomal protein S15 and G150A on ribosomal protein L25 was responsible for an amino acid substitution Q80R and M50I residue, which led to a mass shift of *m/z* 10138.6 to 10166.6 and *m/z* 10694.4 to 10676.4, respectively. Moreover, the peak intensity and sharpness for proteins S15 and L25 in O157 serovars were sufficient to distinguish them from other *E. coli* serovars (Figure 12.12). Out of more than thousands of *E. coli* strains available in the NCBI database, the theoretical masses of ribosomal protein S15 and L25 in all non-O157 *E. coli* strains were calculated as *m/z* 10138.6 and 10694.4, respectively.

Although the mass spectrum at *m/z* 9066.2 $[M+H]^+$ was identified as the acid stress chaperone HdeB in non-EHEC strains [101], no peak was observed in all EHEC serotype O157 used in this study with complete reproducibility, as supported by the evidence of Carter *et al.* [62] (Figure 12.12 and Table 12.6). Sequence analysis of the *hdeB* gene also confirmed that the start codon, ATG, had a point mutation (ATA) in all EHEC serotype O157 strains, whereas in all other *E. coli* strains, including other EHEC serotypes, the start codon, ATG, was held in a normal position in the sequence of the *hdeB* gene. This strongly supported the suggestion that this mutation correlates to the lack of the HdeB peak only in the *E. coli* serotype O157 [62].

EHEC serotypes O26 and O111, which are also responsible for a large number of EHEC outbreaks, could be distinguished from other *E. coli* strains by the peak at *m/z* 15425.4 $[M+H]^+$ (Table 12.6 and Figure 12.12). Because a previous study reported that the protein corresponding to *m/z* 15409.4 $[M+H]^+$ in *E. coli* K-12 strain (accession number P0ACF8) may be the DNA-binding protein H-NS [102], sequence analysis confirmed that an amino acid substitution (A81S) in the DNA-binding protein H-NS in strains O26 and O111 led to a mass shift of *m/z* 15409.4 to 15425.4, implying conclusive biomarker specificity for EHEC serotypes O26 and O111 in MALDI-TOF MS analysis.

Although the peak at *m/z* 6040 has been reported as a biomarker specifically present in EHEC serotype O157 [99], in our study, the intensity of the peak at *m/z* 6040 was too low to detect or the peak was absent in some cases, suggesting that the presence or absence of suspicious biomarker proteins with low intensity is insufficient for use as a biomarker for discrimination at the strain or serotype level.

Although the expression levels of ribosomal protein S15 and L25 were not affected by any growth medium, in accordance with the previous report that the impact of growth

Table 12.6 Peak pattern of selected biomarkers for *E. coli* discrimination.

	Group of mass pattern															
Biomarkers	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**
L23	1 (11200.1)	1	1	1	1	1	2 (11147.1)	1	1	1	1	1	1	1	1	1
L24	1 (11186.0)	1	1	1	1	1	1	1	2 (11216.0)	1	1	1	1	1	1	2
S14	1 (11450.3)	1	1	1	1	1	1	1	1	1	1	1	1	2 (11464.3)	1	1
L15	1 (14967.4)	1	1	1	1	1	2 (14981.4)	3 (14945.0)	1	2	2	1	2	1	1	1
S11 + Me	1 (13728.8)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2 (13756.8)
YdaQ	1 (8325.6)	0	1	1	0	1	1	1	1	1	1	0	0	1	0	0
S15	2 (10166.6)	2	1 (10138.6)	1	1	1	1	1	1	1	1	1	1	1	1	1
HdeB	0	0	0	1 (9066.2)	1	1	1	1	1	1	1	1	1	1	1	1
H-NS	1 (15409.4)	1	1	2 (15425.4)	2	1	1	1	1	1	3 (15882.0)	1	1	1	1	1

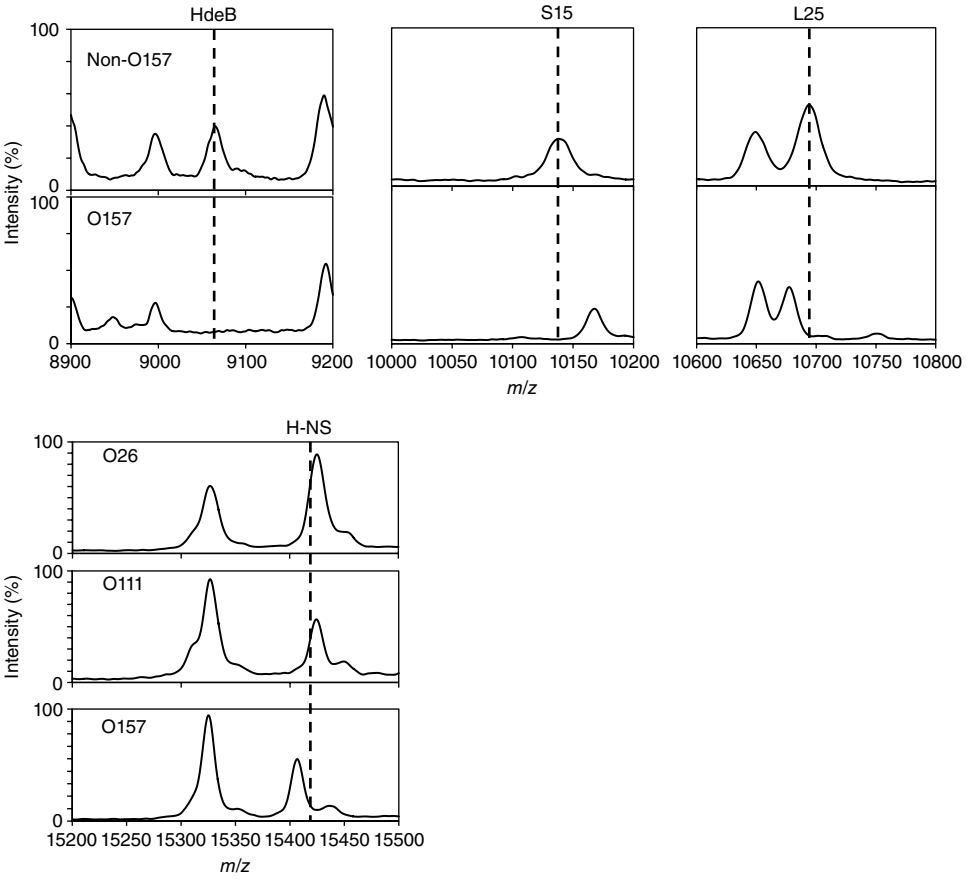


Figure 12.12 Typical MALDI mass spectra of four biomarker proteins in *E. coli*: HdeB (*m/z* 9066.2 $[M+H]^+$), S15 (*m/z* 10138.6/10166.6 $[M+H]^+$), L25 (*m/z* 10676.4/10694.4 $[M+H]^+$), and H-NS (*m/z* 15409.4/15425.4 $[M+H]^+$) for strains O 157, O26, O111, and other *E. coli* (non-O157).

conditions on ribosomal proteins is minimal [103], HdeB in some strains of serotypes O111 and O26 were less detectable when grown on chromagar X-gal or VRBL. Therefore, to ensure the discrimination of O157, O26, and O111 from the others in colony-directed MALDI-TOF MS analysis, in this case, the normal growth media, desoxycholate agar or CT-SMAC, will be recommended for the pre-selection of *E. coli*. Either sinapic acid or CHCA as the matrix reagent is available whether the sample is a colony or a liquid extracted with FA.

Our discrimination approach was verified by performing blind tests using 57 *E. coli* strains including O157, O26, O111, and O121 isolated as wild types. The mass shifts of S15 (*m/z* 10166.6) and L25 (*m/z*10676.4), combined with the absence of an HdeB peak at *m/z* 9066.2, could be universally applied for the discrimination of *E. coli* serotype O157. Similarly, *E. coli* serotypes O26 and O111 were correctly classified on the basis of the specific masses of H-NS (Table 12.6 and Figure 12.13).

The *S10*-GERMS-based discrimination method established a possible strategy for the effective discrimination of *E. coli* serotypes O157, O26, and O111 using four specific

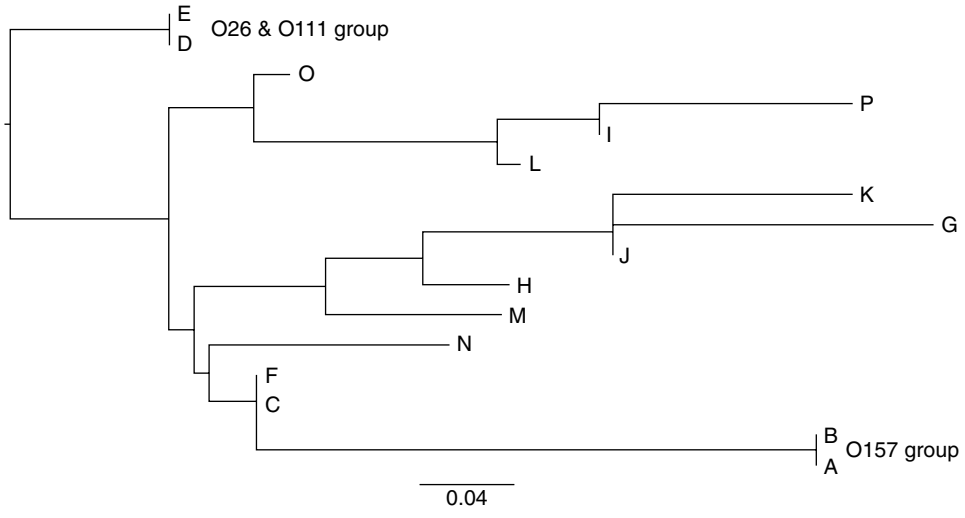


Figure 12.13 Cluster analysis for *E. coli* strains with selected biomarkers. Phylogenetic tree is made based on Table 12.6. A to P indicate the *E. coli* groups classified in Table 12.6.

biomarkers by MALDI-TOF MS as shown in Figure 12.14. In an *E. coli* study, the *S10*-GERMS-based discrimination method stretched the concept of the original *S10*-GERMS method, which uses the information encoded in the *S10-spc-alpha* operon.

In fact, the theoretically selected masses can play a key role as biomarkers in typing of bacteria at the strain level when their masses are confirmed by both genomic and proteomic approaches.

Therefore, to overcome the limitation of the MALDI-TOF MS fingerprint method for distinguishing mixtures containing the same strains of bacteria, the functionality of the *S10*-GERMS method can be expanded from isolated bacterial discrimination at the strain level to discrimination of mixtures consisting of two different types of *E. coli* strains, that is, the O157 strain and the K12 strain as a non-O157 representative, because they are usually identified as "*E. coli*" with the conventional fingerprinting method.

In this case, the theoretically supported *m/z* of biomarkers S15 (10138) and L25 (10694) in the *E. coli* K 12 strain is sufficiently different from the corresponding *m/z* of S15 (10166) and L25 (10676) in the *E. coli* O157 strain. The experimentally mixed sample was successfully discriminated by the *S10*-GERMS method using those biomarkers, which were detected as double peaks at the corresponding values of *m/z* (Figure 12.15).

This scientific rationale of the *S10*-GERMS method will open a new window in monitoring contaminations with similar bacteria in the fields of diagnostics, food safety, and public culture collections that deal with multiple related bacterial samples.

Because the *S10*-GERMS method reflects different evolutionary lineages for biomarkers backed by multi-gene sequence information (Figure 12.1), the concept of the *S10*-GERMS method is a significantly useful tool for bacterial typing below the species level across the microbial kingdom, including clinical, environmental, and food-related fields. Thus, the higher bacterial discrimination performance of MALDI-TOF MS analysis will be achieved in combination with conventional MALDI-TOF MS fingerprinting and the *S10*-GERMS method as an integrated system in future.

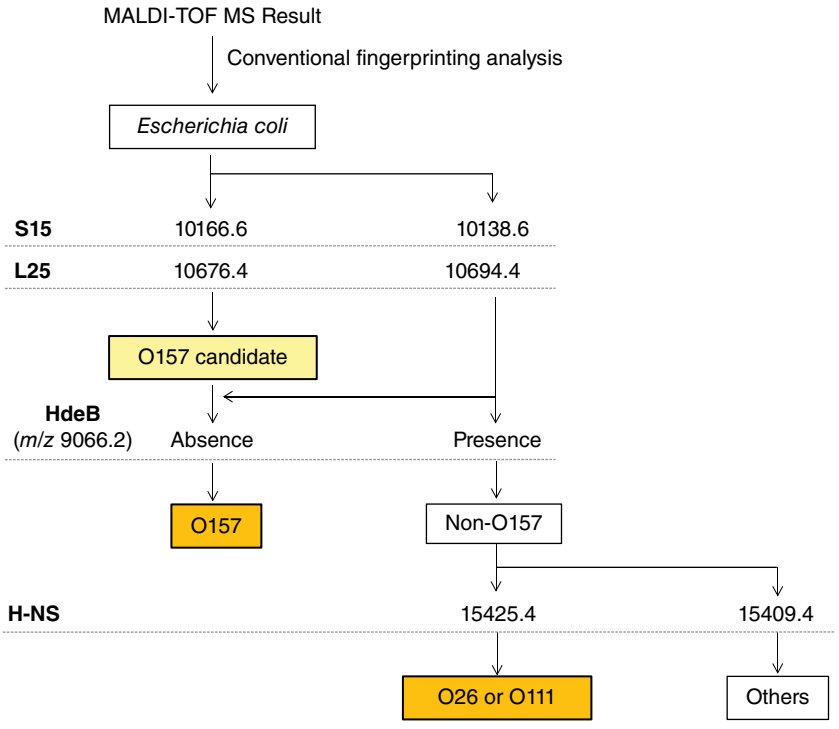


Figure 12.14 Strategy for MALDI-TOF MS proteotyping between *E. coli* strains O157, O26, O111, and the others using four biomarkers.

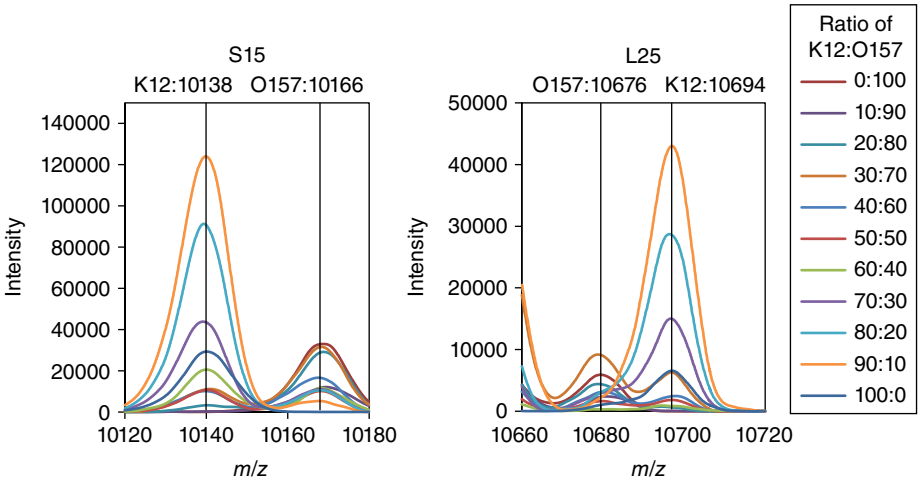


Figure 12.15 Discriminating mixtures of two different types of *E. coli* strains. Strains of K12 and O157 were used as non-O157 representative and O157, respectively. The *m/z* values of the biomarkers are S15 (10138) and L25 (10694) in *E. coli* K 12, and S15 (10166) and L25 (10676) in *E. coli* O157 strain, respectively. The mixed ratio of K12:O157 is referred to at the right. This figure was modified based on Reference 105.

12.3 Conclusion: Computer-Aided Proteotyping of Bacteria Based on the *S10*-GERMS Method

It is well known that the experimental procedure of MALDI-TOF MS analysis is significantly simpler and faster than that of the genome-based method, which requires time-consuming procedures such as DNA extraction and amplification, gel electrophoresis, and DNA sequencing. In particular, making the ribosomal proteins encoded in the *S10-spc-alpha* operon a target, the *S10*-GERMS method was developed as a more highly reliable, advanced method for phylogenetic analysis at the strain level with a validation procedure than conventional MALDI-TOF MS fingerprinting and achieved the desired object (standing goal) of reducing the influence of culture condition. Because the *S10*-GERMS method, based on a combination of genomics and proteomics, reflects different evolutionary lineages for ribosomal proteins backed by multi-gene sequence information (Figure 12.1), the proteopyping using the *S10*-GERMS method may take bacterial identification to the next generation. The main difference between the *S10*-GERMS method and every other conventional method is shown in Figure 12.16. The cutting edge of the *S10*-GERMS method is that it focuses on "the differences" of characteristic biomarkers based on the biological rationale for their *m/z*, whereas other conventional methods, including 16S rRNA sequencing and MS fingerprint methods, search their database systematically for "the similarities."

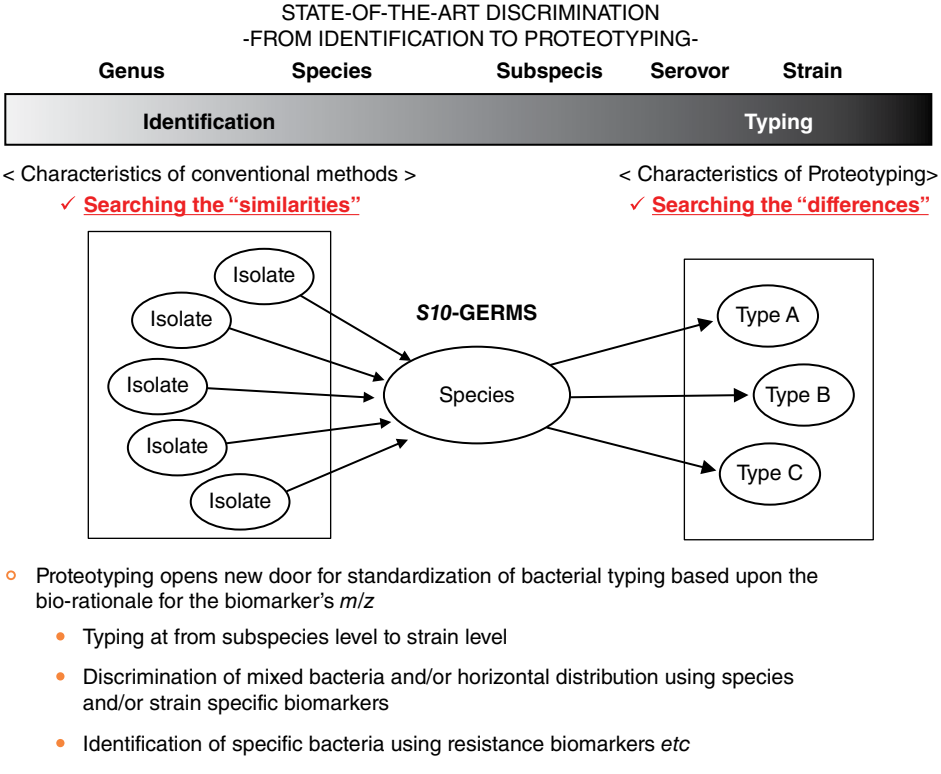
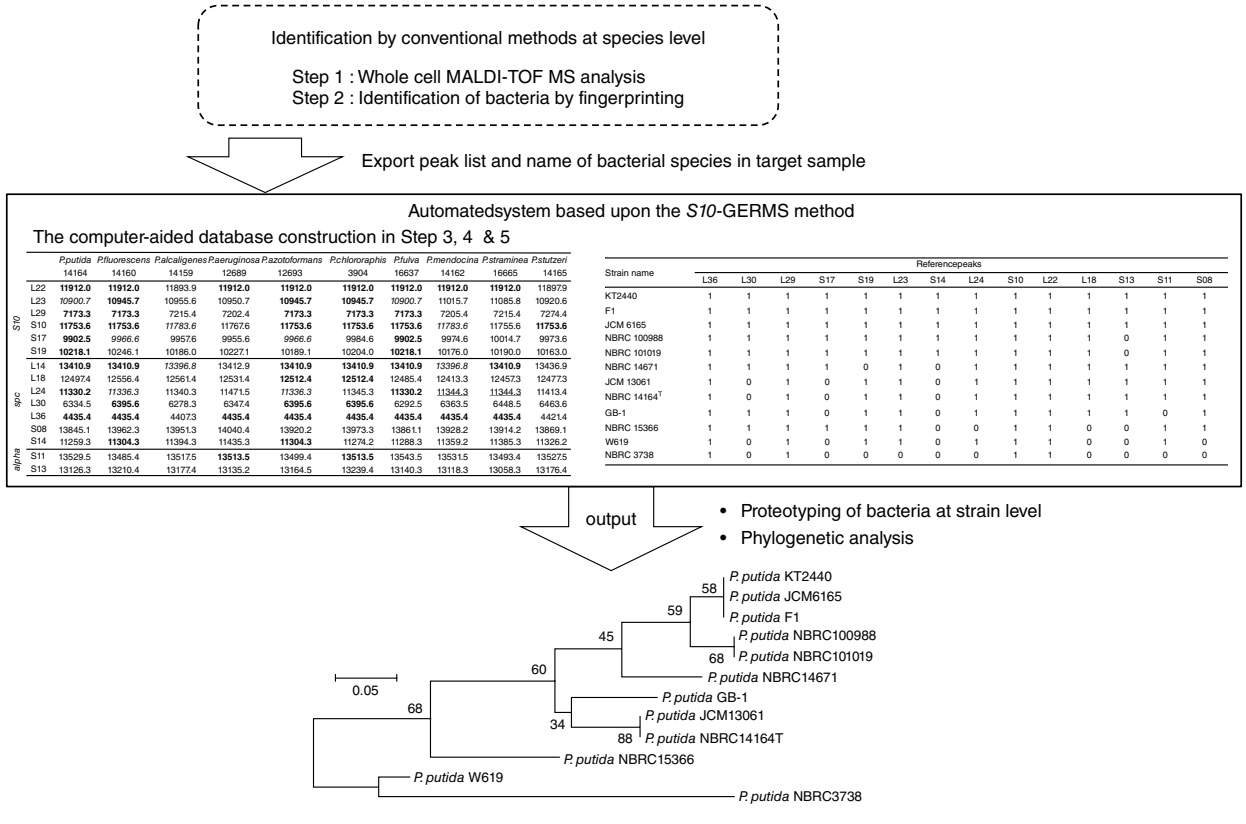


Figure 12.16 The main differences between the *S10*-GERMS method and other conventional methods.



		P.putida 14164	P.fluorescens 14160	P.alcaligenes 14159	P.aeruginosa 12689	P.azotoformans 12693	P.chlororaphis 3904	P.fulva 16637	P.mendocina 14162	P.straminea 16665	P.stutzeri 14165
S10	L22	**11912.0**	**11912.0**	11893.9	**11912.0**	**11912.0**	**11912.0**	**11912.0**	**11912.0**	**11912.0**	11897.9
	L23	10900.7	**10945.7**	10955.6	10950.7	**10945.7**	**10945.7**	10900.7	11015.7	11085.8	10920.6
	L29	**7173.3**	**7173.3**	7215.4	7202.4	**7173.3**	**7173.3**	**7173.3**	7205.4	7215.4	7274.4
	S10	**11753.6**	**11753.6**	*11783.6*	11767.6	**11753.6**	**11753.6**	**11753.6**	*11783.6*	11755.6	**11753.6**
	S17	**9902.5**	*9966.6*	9957.6	9955.6	*9966.6*	9984.6	**9902.5**	9974.6	10014.7	9973.6
	S19	**10218.1**	10246.1	10186.0	10227.1	10189.1	10204.0	**10218.1**	10176.0	10190.0	10163.0
spc	L14	**13410.9**	**13410.9**	*13396.8*	13412.9	**13410.9**	**13410.9**	**13410.9**	*13396.8*	**13410.9**	13436.9
	L18	12497.4	12556.4	12561.4	12531.4	**12512.4**	**12512.4**	12485.4	12413.3	12457.3	12477.3
	L24	**11330.2**	*11336.3*	11340.3	11471.5	*11336.3*	11345.3	**11330.2**	11344.3	11344.3	11413.4
	L30	6334.5	**6395.6**	6278.3	6347.4	**6395.6**	**6395.6**	6292.5	6363.5	6448.5	6463.6
	L36	**4435.4**	**4435.4**	4407.3	**4435.4**	**4435.4**	**4435.4**	**4435.4**	**4435.4**	**4435.4**	4421.4
alpha	S08	13845.1	13962.3	13951.3	14040.4	13920.2	13973.3	13861.1	13928.2	13914.2	13869.1
	S14	11259.3	**11304.3**	11394.3	11435.3	**11304.3**	11274.2	11288.3	11359.2	11385.3	11326.2
	S11	13529.5	13485.4	13517.5	**13513.5**	13499.4	**13513.5**	13543.5	13531.5	13493.4	13527.5
	S13	13126.3	13210.4	13177.4	13135.2	13164.5	13239.4	13140.3	13118.3	13058.3	13176.4

Strain name	Referencepeaks													
	L36	L30	L29	S17	S19	L23	S14	L24	S10	L22	L18	S13	S11	S08
KT2440	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JCM 6165	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NBRC 100988	1	1	1	1	1	1	1	1	1	1	1	0	1	1
NBRC 101019	1	1	1	1	1	1	1	1	1	1	1	0	1	1
NBRC 14671	1	1	1	1	0	1	0	1	1	1	1	1	1	1
JCM 13061	1	0	1	0	1	1	0	1	1	1	1	1	1	1
NBRC 14164^{T}	1	0	1	0	1	1	0	1	1	1	1	1	1	1
GB-1	1	1	1	0	1	1	0	1	1	1	1	1	0	1
NBRC 15366	1	1	1	1	1	1	0	0	1	1	0	0	1	1
W619	1	0	1	0	1	1	0	1	1	1	0	0	1	0
NBRC 3738	1	0	1	0	0	0	0	0	1	1	0	0	0	0

Figure 12.17 Concept of the automatic proteotyping system combining conventional MALDI-TOF MS fingerprinting with the *S10*-GERMS method.

Therefore, automated processing and clustering of the data generated by the *S10*-GERMS method is desired. To implement a solution, a combination of the conventional MALDI-TOF MS fingerprinting with the *S10*-GERMS method as an integrated system will provide higher performance of bacterial discrimination by MALDI-TOF MS analysis, that is, ease of use, cost performance, accuracy, reproducibility, high throughput, and data portability.

In addition, the standardized *S10*-GERMS method also has the potential to develop into a new method for the characterization of a bacterial mixture based on the information of biomarkers with their own molecular weight information, because the *m/z* values of the observed ions are specific for the corresponding species or strains, despite the fact that the new approach of automatic identification of mixed bacterial species has been developed on the basis of fingerprints [104].

The concept of automated processing and clustering of the data generated by the *S10*-GERMS method is shown in Figure 12.17. The automated system will result in the following features [105]:

- Discrimination of genetically similar bacteria that are difficult to differentiate by16S rRNA gene sequencing
- Phylogenetic analysis at strain and serotype level
- Identification of mixed bacterial species

Moreover, when anyone deposits isolated bacteria in the culture collection with their MALDI-TOF MS spectra, for example, Spectra Bank [106], in the future, establishment of a useful search engine for computer-aided proteotyping based on the standardized *S10*-GERMS method will open new avenues in the field of bacterial discrimination.

References

1 Stackebrandt, E. and Ebers, J. Taxonomic parameters revisited: Tarnished gold standards. *Microbiol. Today*, 4, 152–155, 2006.

2 Stackebrandt, E. and Goebel, B. M. Taxonomic Note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.*, 44, 846–849, 1994.

3 Viulu, S., Nakamura, K., Kojima, A., Yoshiyasu, Y., Saitou, S. and Takamizawa, K. *Geobacter sulfurreducens* subsp. *ethanolicus*, subsp.nov., an ethanol-utilizing dissimilatory Fe (III)-reducing bacterium from a lotus field. *J. Gen. Appl. Microbiol.*, 59, 325–334, 2013.

4 Anhalt, J. P. and Fenselau, C. Identification of bacteria using mass spectrometry. *Anal. Chem.*, 47, 219–225, 1975.

5 Shah, H. N. and Collins, M. D. Fatty acid and isoprenoid quinone composition in the classification of *Bacteroides melaninogenicus* and related taxa. *J. Appl. Bacteriol.*, 48, 75–87, 1980.

6 Fenselau, C. and Demirev, P. A. Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.*, 20, 157–171, 2001.

7 Croxatto, A., Prod'hom, G. and Greub, G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS Microbiol. Rev.*, 36, 380–407, 2012.

8 Wieser, A., Schneider, L., Jung, J. and S. Schubert, MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (nimi review). *Appl. Microbiol. Biotechnol.*, 93, 965–974, 2012.

9 Dridi, B., Raoult, D. and Drancourt, M. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of *Archaea*: Towards the universal identification of living organisms (review). *APMIS*, 120, 85–91, 2011.

10 Welham, K. J., Domin, M., A. Johnson, K., Jones, L. and Ashton, D. S. Characterization of fungal spores by laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.*, 14, 307–310, 2000.

11 Amiri-Eliasi, B. and Fenselau, C. Characterization of protein biomarkers desorbed by MALDI from whole fungal cells. *Anal. Chem.*, 73, 5228–5231, 2001.

12 Marinach-Patrice, C., Lethuillier, A., Marly, A., Brossas, J. Y., Gené, J. Symoens, F., Datry, A., Guarro, J., Mazier, D. and Hennequin, C. Use of mass spectrometry to identify clinical *Fusarium* isolates. *Clin. Microbiol. Infect.*, 15, 634–642, 2009.

13 Usbeck J. C., Kern, C. C., Vogel, R. F. and Behr, J. Optimization of experimental and modelling parameters for the differentiation of beverage spoiling yeasts by matrix-assisted-laser-desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS) in response to varying growth conditions. *Food Microbiol.*, 36, 379–387, 2013.

14 Yao, Z. P., Demirev, P. A. and Fenselau, C. Mass spectrometry-based proteolytic mapping for rapid virus identification. *Anal. Chem.*, 74, 2529–2534, 2002.

15 Vargha, M., Takats, Z., Konopka, A. and Nakatsu, C. H. Optimization of MALDI-TOF MS for strain level differentiation of *Arthrobacter* isolates. *J. Microbil. Methods*, 66, 399–409, 2006.

16 Siegrist, T. J., Anderson, P. D., Huen, W. H., Kleinheinz, G. T., McDermott, C. M. and Sandrin, T. R. Discrimination and characterization of environmental strains of *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). *J. Microbiol. Methods*, 68, 554–562, 2007.

17 Barbuddhe, S. B., Maier, T. and Schwarz, G. Rapid identification and typing of *Listeria* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Proteomics*, 2, 747–753, 2008.

18 Dieckmann, R. and Malorny, B. Rapid screening of epidemiologically important *Salmonella enterica* subsp. *enterica* serovars using Whole-Cell MALDI-TOF mass spectrometry. *Appl. Environ. Microbiol.*, 77, 4136–4146, 2011.

19 Grosse-Herrenthey, A., Maier, T., Gessler, F., Schaumann, R., Böhnel, H., Kostrzewa, M. and Krüger, M. Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Anaerobe*, 14, 242–249, 2008.

20 Nagy, E., Maier, T., Urban, E., Terhes, G. and Kostrzewa, M. Species identification of clinical isolates of *Bacteroides* by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. *Clin. Microbiol. Infect.*, 15, 796–802, 2009.

21 Cherkaoui, A., Hibbs, J., Emonet, S., Tangomo, M., Girard, M., Francois, P. and Schrenzel, J. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *J. Clin. Microbiol.*, 48, 1169–1175, 2010.

22 Welker, M. and Moore, E. R. Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology (mini review). *Syst. Appl. Microbiol.*, 34, 2–11, 2011.

23 Idelevich, E. A., Schüle, I., Grünastel, B., Wüllenweber, J., Peters, G. and Becker, K. Rapid identification of microorganisms from positive blood cultures by MALDI-TOF mass spectrometry subsequent to very short-term incubation on solid medium. *Clin. Microbiol. Infect.*, Apr 3. 1–6, 2014.

24 Machen, A., Drake, T. and Wang, Y.F. Same day identification and full panel antimicrobial susceptibility testing of bacteria from positive blood culture bottles made possible by a combined lysis-filtration method with MALDI-TOF VITEK mass spectrometry and the VITEK2 system. *PLoS ONE*, 9, e87870, 2014.

25 Carbonnelle, E., Grohs, P., Jacquier, H., Day, N., Tenza, S., Dewailly, A., Vissouarn, O., Rottman, M., Herrmann, J. L., Podglajen, I. and Raskine L. Robustness of two MALDI-TOF mass spectrometry systems for bacterial identification. *J. Microbiol. Methods*, 89, 133–136, 2012.

26 Ford, B. A. and Burnham, C. A. Optimization of routine identification of clinically relevant Gram-negative bacteria by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry and the Bruker Biotyper. *J. Clin. Microbiol.*, 51, 1412–1420, 2013.

27 Goldstein, J. E., Zhang, L., Borror, C. M., Rago, J. V. and Sandrin, T. R. Culture conditions and sample preparation methods affect spectrum quality and reproducibility during profiling of *Staphylococcus aureus* with matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry. *Lett. Appl. Microbiol.*, 57, 144–150, 2013.

28 Wieme, A. D., Spitaels, F., Aerts, M., De Bruyne, K., Van Landschoot, A. and Vandamme, P. Effects of growth medium on matrix-assisted laser desorption-ionization time of flight mass spectra: A case study of acetic acid bacteria. *Appl. Environ. Microbiol.*, 80, 1528–1538, 2014.

29 Balážová, T., Makovcová, J., Šedo, O., Slaný, M., Faldyna, M. and Zdráhal, Z. The influence of culture conditions on the identification of *Mycobacterium* species by MALDI-TOF MS profiling. *FEMS Microbiol. Lett.*, 353, 77–84, 2014.

30 Khot, P. D. and Fisher, M. A. Novel approach for differentiating *Shigella* species and Escherichia coli by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.*, 51, 3711–3716, 2013.

31 Zhang, L., Borror, C. M. and Sandrin, T. R. A designed experiments approach to optimization of automated data acquisition during characterization of bacteria with MALDI-TOF mass spectrometry. *PLoS ONE*, 9, e92720, 2014.

32 Branquinho, R., Sousa, C, Lopes, J., Pintado, M. E., Peixe, L. V. and Osório, H. Differentiation of *Bacillus pumilus* and *Bacillus safensis* using MALDI-TOF-MS. *PLoS ONE*, 9, e110127, 2014.

33 Mahé, P., Arsac, M., Chatellier, S., Monnin, V., Perrot, N., Mailler, S., Girard, V., Ramjeet, M., Surre, J., Lacroix, B., van Belkum, A. and Veyrieras, J. B. Automatic identification of mixed bacterial species fingerprints in a MALDI-TOF mass-spectrum. *Bioinformatics*, 30, 1280–1286, 2014.

34 Wittmann, H. G. Components of bacterial ribosomes. *Annu. Rev. Biochem.*, 51, 155, 1982.

35 Arnold, R. J. and Reilly, J. P. Observation of Escherichia coli ribosomal proteins and their posttranslational modifications by mass spectrometry. *Anal. Biochem.*, 269, 105–112, 1999.

36 Ochi, K. Heterogeneity of ribosomal proteins among *Streptomyces* species and its application to identification. *J. Gen. Microbiol.*, 135, 2635–2642, 1989.

37 Ochiai, K., Uchida, K. and Kawamoto, I. Similarity of ribosomal proteins studied by two-dimensional coelectrophoresis for identification of Gram-positive bacteria. *Int. J. Syst. Bacteriol.*, 43, 69–76, 1993.

38 Pineda, F. J., Antoine, M. D., Demirev, P. A., Feldman, A. B., Jackman, J., Longenecker, M. and Lin, J. S. Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. *Anal. Chem.*, 75, 3817–3822, 2003.

39 Sun, L., Teramoto, K., Sato, H., Torimura, M., Tao, H. and Shintani, T. Characterization of ribosomal proteins as biomarkers for matrix-assisted laser desorption/ionization mass spectral identification of *Lactobacillus plantarum*. *Rapid Commun. Mass Spectrom.*, 20, 3789–3798, 2006.

40 Teramoto, K., Sato, H., Sun, L., Torimura, M., Tao, H., Yoshikawa, H., Hotta, Y., Hosoda, A. and Tamura, H. Phylogenetic classification of *Pseudomonas putida* strains by MALDI-MS using ribosomal subunit proteins as biomarkers. *Anal. Chem.*, 79, 8712–8719, 2007.

41 Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogestraat, D. R., Cummings, L. A., Sengupta, D. J., Harkins, T. T., Cookson, B. T. and Hoffman, N. G. Performance comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-based Bacterial Community Profiling. *Appl. Environ. Microbiol.*, 80, 7583–7591, 2014.

42 Demirev, P. A., Ho, Y. P., Ryzhov, V. and Fenselau, C. Microorganism identification by mass spectrometry and protein database searches. *Anal. Chem.*, 71, 2732–2738, 1999.

43 Pineda, F. J., Lin, J. S., Fenselau, C. and Demirev, P. A. Testing the significance of microorganism identification by mass spectrometry and proteome database search. *Anal. Chem.*, 72, 3739–3744, 2000.

44 Demirev, P. A., Lin, J. S., Peneda, F. J. and Fenselau, C. Bioinformatics and mass spectrometry for microorganism identification: Proteome-wide post-translational modifications and database search algorithms for characterization of intact *H. pylori*. *Anal. Chem.*, 73, 4566–4573, 2001.

45 Lindahl, L. and Zengel, J. M. Ribosomal genes in *Escherichia coli*. *Annu. Rev. Genet.*, 20, 297–326, 1986.

46 Suyama, M. and Bork, P. Evolution of prokaryotic gene order: Genome rearrangements inclosely related species. *Trends Gen.*, 17, 10–13, 2001.

47 Tamames, J. Evolution of gene order conservation in prokaryotes. *Genome Biol.*, 2, 0020.1–0020.11, 2001.

48 Suh, M. J., Hamburg, D. M., Gregory, S.T., Dahlberg, A. E. and Limbach, P. A. Extending ribosomal protein identifications to unsequenced bacterial strains using matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics*, 5, 4818–4831, 2005.

49 Teramoto, K., Sato, H., Sun, L., Torimura, M. and Tao, H. A simple intact protein analysis by MALDI-MS for characterization of ribosomal proteins of two genome-sequenced lactic acid bacteria and verification of their amino acid sequences. *J. Proteome Res.*, 6, 3899–3907, 2007.

50 Running, W. E., Ravipaty, S., Karty, J. A. and Reilly, J. P. A top-down/bottom-up study of the ribosomal proteins of *Caulobacter crescentus*. *J. Proteome Res.*, 6, 337–347, 2007.

51 Sato, H., Teramoto, K., Ishii, Y., Watanabe, K. and Benno, Y. Ribosomal protein profiling by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for phylogenety-based subspecies resolution of *Bifidobacterium longum*. *Syst. Appl. Microbiol.*, 34, 76–80, 2011.

52 Erhard, M., Hipler, U. C., Burmester, A., Brakhage, A. A. and Wostemeyer, J. Identification of dermatophyte species causing onychomycosis and tinea pedis by MALDI-TOF mass spectrometry. *Exp. Dermatol.*, 17, 356–361, 2008.
53 Matte-Tailliez, O., Brochier, C., Forterre, P. and Philippe, H. Archaeal phylogeny based on ribosomal proteins. *Mol. Biol. Evol.*, 19, 631–639, 2002.
54 Watanabe, H., Mori, H., Itoh, T. and Gojobori, T. Genome plasticity as a paradigm of eubacteria evolution. *J. Mol. Evol.*, 44, S57–S64, 1997.
55 Itoh, T., Takemoto, K., Mori, H. and Gojobori, T. Evolutionary instability of Operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol. Biol. Evol.*, 16, 332–346, 1999.
56 Barloy-Hubler, F., Lelaure, V. and Galibert, F. Ribosomal protein gene cluster analysis in eubacterium genomics: homology between *Sinorhizobium meliloti* strain 1021 and *Bacillus subtilis*. *Nucleic Acids Res.*, 29, 2747–2756, 2001.
57 Coenye, T. and Vandamme, P. Organisation of the *S10, spc* and *alpha* ribosomal protein gene clusters in prokaryotic genomes. *FEMS Microbiol. Lett.*, 242, 117–126, 2005.
58 Victoria, B., Ahmed, A., Zuerner, R. L., Ahmed, N., Bulach, D. M., Quinteiro, J. and Hartskeerl, R. A. Conservation of the *S10–spc–alpha* locus within otherwise highly plastic genomes provides phylogenetic insight into the genus Leptospira. *PLoS ONE*, 3, e2752, 2008.
59 Valentine, N. B., Wahl, J. H., Kingsley, M. T. and Wahl, K. L. Direct surface analysis of fungal species by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.*, 16, 1352–1357, 2002.
60 Sherman, F., Stewart, J. W. and Tsunasawa, S. Methionine or not methionine at the beginning of a protein. *Bioessays*, 3, 27–31, 1985.
61 Moerschell, R. P., Hosokawa, Y., Tsunasawa, S. and Sherman, F. The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine *in vivo*. Processing of altered *iso*-1-cytochromes *c* created by oligonucleotide transformation. *J. Biol. Chem.*, 265, 19638–19643, 1990.
62 Carter, M. Q., Louie, J. W., Fagerquist, C. K., Sultan, O., Miller, W. G. and Mandrell, R. E. Evolutionary silence of the acid chaperone protein HdeB in enterohemorrhagic *Escherichia coli* O157:H7. *Appl. Environ. Microbiol.*, 78, 1004–1014, 2012.
63 Ziegler, D., Mariotti, A., Pflüger, V., Saad, M., Vogel, G., Tonolla, M. and Perret, X. *In situ* identification of plant-invasive bacteria with MALDI-TOF mass spectrometry. *PLoS ONE*, 7, e37189, 2012.
64 Hotta, Y., Teramoto, K., Sato, H., Yoshikawa, H., Hosoda, A. and Tamura, H. Classification of genus *Pseudomonas* by MALDI-TOF MS based on ribosomal protein coding in *S10-spc-alpha* operon at strain level. *J. Proteome Res.*, 9, 6722–6728, 2010.
65 Tamura, H., Hotta, Y. and Sato, H. Novel accurate bacterial discrimination by MALDI-Time-of-Flight MS based on ribosomal proteins coding in *S10-spc-alpha* Operon at strain level *S10*-GERMS. *J. Am. Soc. Mass Spectrom.*, 24, 1185–1193, 2013.
66 Sato, H., Shibata, A., Wang, Y., Yoshikawa, H. and Tamura, H. *Polym. Degrad. Stab.*, 74, 69–75, 2001.
67 Nishino, E., Ichiki, Y., Tamura, H., Morita, S., Watanabe, K. and Yoshikawa, H. *Biosci. Biotechnol. Biochem.*, 66, 1792–1798, 2002.
68 Yamamoto, S. and Harayama, S. Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int. J. Syst. Bacteriol.*, 48, 813–819, 1998.

69 Yamamoto, S., Bouvet, P. J. M. and Harayama, S. Phylogenetic structures of the genus *Acinetobacter* based on *gyrB* sequences: Comparison with the grouping by DNA-DNA hybridization. *Int. J. Syst. Bacteriol.*, 49, 87–95, 1999.

70 Hotta, Y., Sato, H., Hosoda, A. and Tamura, H. MALDI-TOF MS analysis of ribosomal proteins coded in *S10* and *spc* operons rapidly classified *Sphingomonaceae* as alkylphenol polyethoxylate-degrading bacteria from the environment. *FEMS Microbiol Lett.*, 330, 23–29, 2012.

71 Yabuuchi, E., Yano, I., Oyaizu, H., Hashimoto, Y., Ezaki, T. and Yamamoto, H. Proposals of *Sphingomonas paucimobilis* gen. nov. and comb. nov., *Sphingomonas parapaucimobilis* sp. nov., *Sphingomonas yanoikuyae* sp. nov., *Sphingomonas adhaesiva* sp. nov., *Sphingomonas capsulata* comb. nov., and two genospecies of the genus *Sphingomonas*. *Microbiol. Immunol.*, 34, 99–119, 1990.

72 Takeuchi, M., Hamana, K. and Hiraishi, A. Proposal of the genus *Sphingomonas sensu stricto* and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. *Int. J. Syst. Evol. Microbiol.*, 51, 1405–1417, 2001.

73 Lyu, Y., Zheng, W., Zheng, T. and Tian, Y. Biodegradation of polycyclic aromatic hydrocarbons by *Novosphingobium pentaromativorans* US6-1. *PLoS ONE*, 9, e101438, 2014.

74 White, G. F. Bacterial biodegradation of ethoxylated surfactants. 1993. *Pestic. Sci.*, 37, 159–166.

75 Laws, S. C., Carey, S. A., Ferrell, J. M., Bodman, G. J. and Cooper, R. L. Estrogenic activity of octylphenol, nonylphenol, bisphenol A and methoxychlor in rats. *Toxicol. Sci.*, 54, 154–167, 2000.

76 Shibata, A., Ishimoto, Y., Nishizaki, Y., Hosoda, A., Yoshikawa, H. and Tamura, H. The effect of calcium ion on the biodegradation of octylphenol polyethoxylates, and the antiandrogenic activity of their biodegradates. *Appl. Microbiol. Biotechnol.*, 77, 195–201, 2007.

77 Hotta, Y., Sato, J., Sato, H., Hosoda, A. and Tamura, H. Classification of Genus *Bacillus* based on MALDI-TOF MS analysis of ribosomal protein coded in *S10* and *spc* operons. *J. Agric. Food. Chem.*, 59, 5222–5230, 2011.

78 Meerak, J., Yukphan, P., Miyashita, M., Sato, H., Nakagawa, Y. and Tahara, Y. Phylogeny of γ-polyglutamic acid-producing *Bacillus* strains isolated from a fermented locust bean product manufactured in West Africa. *J. Gen. Appl. Microbiol.*, 54, 159–166, 2008.

79 Yamada, S., Ohashi, E., Agata, N. and Venkateswaran, K. Cloning and nucleotide sequence analysis of *gyrB* of *Bacillus cereus*, *B. thuringiensis*, *B. mycoides*, and *B. anthracis* and their application to the detection of *B. cereus* in rice. *Appl. Environ. Microbiol.*, 65, 1483–1490, 1999.

80 Granum, P. E. and Lund, T. *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol. Lett.*, 157, 223–228, 1997.

81 Priest, F. G., Barker, M., Baillie, L. W., Holmes, E. C. and Maiden, M. C. Population structure and evolution of the *Bacillus cereus* group. *J. Bacteriol.*, 186, 7959–7970, 2004.

82 Daligault, H. E., Davenport, K. W., Minogue, T. D., Bishop-Lilly, K. A., Broomall, S. M., Bruce, D. C., Chain, P. S., Coyne, S. R., Frey, K. G., Gibbons, H. S,, Jaissle, J., Koroleva, G. I., Ladner, J. T., Lo, C. C., Munk, C., Palacios, G. F., Redden, C. L., Rosenzweig, C. N., Scholz, M. B. and Johnson, S. L. Twenty Whole-Genome Bacillus sp. assemblies. *Genome Announc.*, 2, e00958–14, 2014.

83 Lauber, M. A., Running, W. E. and Reilly, J. P. *B. subtilis* ribosomal proteins: structural homology and post-translational modifications. *J. Proteome Res.*, 8, 4193–4206, 2009.

84 Noguchi, H., Uchino, M., Shida, O., Takano, K., Nakamura, L. K. and Komagata, K. *Bacillus vietnamensis* sp. nov., a moderately halotolerant, aerobic, endospore-forming bacterium isolated from Vietnamese fish sauce. *Int. J. Syst. Evol. Microbiol.*, 54, 2117–2120, 2004.

85 Cerritos, R., Vinuesa, P., Eguiarte, L. E., Herrera-Estrella, L., Alcaraz-Peraza, L. D., Arvizu-Gómez, J. L., Olmedo, G.; Ramirez, E., Siefert, J. L. and Souza, V. *Bacillus coahuilensis* sp. nov., a moderately halophilic species from a desiccation lagoon in the Cuatro Ciénegas Valley in Coahuila, Mexico. *Int. J. Syst. Evol. Microbiol.*, 58, 919–923, 2008.

86 Nakamura, L. K., Roberts, M. S. and Cohan, F. M. Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: A proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov, and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. *Int. J. Syst. Bacteriol.*, 49, 1211–1215, 1999.

87 Rooney, A. P., Price, N. P., Ehrhardt, C., Swezey, J. L. and Bannan, J. D. Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *inaquosorum* subsp. nov. *Int. J. Syst. Evol. Microbiol.*, 59, 2429–2436, 2009.

88 Sato, H., Torimura, M., Kitahara, M., Ohkuma, M., Hotta, Y. and Tamura, H. Characterization of the *Lactobacillus casei* group based on the profiling of ribosomal proteins coded in *S10-spc-alpha* operons as observed by MALDI-TOF MS. *System. Appl. Microbiol.*, 35, 447–454, 2012.

89 Klein, G., Pack, A., Bonaparte, C. and Reuter, G. Taxonomy and physiology of probiotic lactic acid bacteria. *Int. J. Food Microbiol.*, 41, 103–125, 1998.

90 Singh, S., Goswami, P., Singh, R. and Heller, K. J. Application of molecular identification tools for *Lactobacillus*, with a focus on discrimination between closely related species (review). *LWT-Food Sci. Technol.*, 42, 448–457, 2009.

91 Collins, M. D., Phillips, B. A. and Zanoni, P. Deoxyribonucleic acid homology studies of *Lactobacillus casei*, *Lactobacillus paracasei* sp. nov., subsp. *paracasei* and subsp. *tolerans*, and *Lactobacillus rhamnosus* sp. nov., comb. nov. *Int. J. Syst. Bacteriol.*, 39, 105–108, 1989.

92 Ryu, C. S., Czajka, J. W., Sakamoto, M. and Benno, Y. Characterization of the *Lactobacillus casei* group and the *Lactobacillus acidophilus* group by automated ribotyping. *Microbiol. Immunol.*, 45, 271–275, 2001.

93 Ai, L. Z., Chen, C., Zhou, F. F., Wang, L., Zhang, H., Chen, W. and Guo, B. H. Complete genome sequence of the probiotic strain *Lactobacillus casei* BD-II. *J. Bacteriol.*, 193, 3160–3161, 2011.

94 Maze, A., Boel, G., Zuniga, M., Bourand, A., Loux, V., Yebra, M. J., Monedero, V., Correia, K., Jacques, N., Beaufils, S., Poncet, S., Joyet, P., Milohanic, E., Casaregola, S., Auffray, Y., Perez-Martinez, G., Gibrat, J. F., Zagorec, M., Francke, C., Hartke, A. and Deutscher, J. Complete genome sequence of the probiotic *Lactobacillus casei* strain BL23. *J. Bacteriol.*, 192, 2647–2648, 2010.

95 Chen, C. L., Ai, Z., Zhou, F. F., Wang, L., Zhang, H., Chen, W. and Guo, B. H. Complete genome sequence of the probiotic bacterium *Lactobacillus casei* LC2W. *J. Bacteriol.*, 193, 3419–3420, 2011.

96 Zhang, W. Y., Yu, D. L., Sun, Z. H., Wu, R. N., Chen, X., Chen, W., Meng, H., Hu, S. N. and Zhang, H. P. Complete genome sequence of *Lactobacillus casei* Zhang, a new probiotic strain isolated from traditional homemade koumiss in Inner Mongolia, *China. J. Bacteriol.*, 192, 5268–5269, 2010.

97 Berthier, F., Beuvier, E., Dasen, A. and Grappin, R. Origin and diversity of mesophilic lactobacilli in Comte cheese, as revealed by PCR with repetitive and species-specific primers. *Int. Dairy J.*, 11, 293–305, 2001.

98 Ojima-Kato, T., Yamamoto, N., Suzuki, M., Fukunaga, T. and Tamura, H. Discrimination of *Escherichia coli* O157, O26 and O111 from other serovars by MALDI-TOF MS based on the *S10*-GERMS method, *PLoS ONE*, 9, e113458, 2014.

99 Clark, C. G., Kruczkiewicz, P., Guan, C., McCorrister, S. J., Chong, P., Wylie, J., van Caeseele, P., Tabor, H. A., Snarr, P., Gilmour, M. W., Taboada, E. N. and Westmacott, G. R. Evaluation of MALDI-TOF mass spectroscopy methods for determination of *Escherichia coli* pathotypes. *J. Microbiol. Methods.*, 94, 180–191, 2013.

100 Karger, A., Ziller, M., Bettin, B., Mintel, B., Schares. S. and Geue, L. Determination of serotypes of Shiga toxin-producing Escherichia coli isolates by intact cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.*, 77, 896–905, 2011.

101 Fagerquist, C. K., Garbus, B. R., Miller, W. G., Williams, K. E., Yee, E., Bates, A. H., Boyle, S., Harden, L. A., Cooley, M. B. and Mandrell, R. E. Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight-time-of-flight mass spectrometry and top-down proteomics. *Anal. Chem.*, 82, 2717–2725, 2010.

102 Momo, R. A., Povey, J. F., Smales, C. M., O'Malley, C. J., Montague, G. A. and Martin, E. B. MALDI-ToF mass spectrometry coupled with multivariate pattern recognition analysis for the rapid biomarker profiling of *Escherichia coli* in different growth phases. *Anal. Bioanal. Chem.*, 405, 8251–8265, 2013.

103 Valentine, N., Wunschel, S., Wunschel, D., Petersen, C. and Wahl, K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. *Appl. Environ. Microbiol.*, 71, 58–64, 2005.

104 Mahé, P., Arsac, M., Chatellier, S., Monnin, V., Perrot, N., Mailler, S., Girard, V., Ramjeet, M., Surre, J., Lacroix, B., van Belkum, A. and Veyrieras, J. B. Automatic identification of mixed bacterial species fingerprints in a MALDI-TOF mass-spectrum. *Bioinformatics*, 30, 1280–1286, 2014.

105 Ojima-Kato, T., Yamamoto, N., Iijima, Y. and Tamura, H., Assessing the performance of novel software Strain Solution on automated discrimination of *Escherichia coli* serotypes and their mixtures using matrix-assisted laser desorption ionization-time of flight mass spectrometry, *J. Microbiol. Methods*, doi:10.1016/j.mimet.2015.11.005.

106 Böhme, K., Fernández-No, I. C., Barros-Velázquez, J., Gallardo, J. M., Cañas, B. and Calo-Mata P. SpectraBank: An open access tool for rapid microbial identification by MALDI-TOF MS fingerprinting. *Electrophoresis*, 33, 2138–2142, 2012.

Part II

Tandem MS/MS-Based Approaches to Microbial Characterization

13

Tandem Mass Spectrometry Analysis as an Approach to Delineate Genetically Related Taxa

Raju V. Misra,[1] Tom Gaulton,[2] Nadia Ahmod,[2] Min Fang,[2] Martin Hornshaw,[3] Jenny Ho,[3] Saheer E. Gharbia[1] and Haroun N. Shah[2]

[1] *Genomics Research Unit, Public Health England, London, UK*
[2] *Proteomics Research, Public Health England, London, UK*
[3] *ThermoFisher Scientific, Hemel Hempstead, Hertfordshire, UK*

This chapter will demonstrate through worked examples the clinical utility of tandem mass spectrometry. Part A uses a model of the challenging family Enterobacteriaceae, ultimately showing how proteomics, combined with genomics, can be used to identify and characterize a recent outbreak strain of *Escherichia coli*, strain type 0104:H4. Part B demonstrates the utility of proteomics as a biomarker discovery and validation platform for the identification of protein and genomic features, for the delineation of high-profile, highly pathogenic biothreat agents.

Part A

13.1 Introduction

The family Enterobacteriaceae is a large heterogeneous group of bacteria comprising a diverse range of organisms inhabiting a wide range of ecological environments, including the intestinal flora of man and animals. Phylogenetic studies indicate that many of the genera within the Enterobacteriaceae are monophyletic, sharing a common ancestor, from which a conserved core of gene sequences can be used to infer phylogeny, for example, 16S rRNA, gyrB or rpoB (Dauga, 2002). However, unlike DNA hybridization and phenotypic studies, molecular approaches based on single genes fail to resolve some species or even genera into monophyletic groups. Consequently, both at the generic and species levels, some taxa merge, even between strains from different species, resulting in ill-defined boundaries (Delmas *et al.*, 2006; Paradis *et al.*, 2005; Pham *et al.*, 2007; Stecher *et al.*, 2012). Among these, *Escherichia coli* is the most studied; there are strains that are pathogenic, causing a wide range of diseases affecting the respiratory and urinary tracts, the bloodstream, and intestinal infections that can be fatal if not treated (Kaper *et al.*, 2004). On the basis of the virulence factors present and the host clinical

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

symptoms, pathogenic *E. coli* strains can be categorized into one or more pathogenic groups, 'pathotypes' such as Enteroaggregrative *E. coli* (EAEC), Enterohemorrhagic *E. coli* (EHEC), Uropathogenic *E. coli* (UPEC) and Enteropathogenic *E. coli* (EPEC) (Kaper *et al.*, 2004).

Virulence factors in *E. coli* comprise a broad range of protein products, which allow pathogens to successfully inhabit and persist in a host. They include, but are not limited to, proteins associated with adhesion, antimicrobial resistance, toxicity and motility (Chen *et al.*, 2012). Ideally, to characterize enteric pathogens, the ability to identify virulence-associated features, coupled with the correct species/strain identification, offers a more comprehensive and clinically relevant description of a pathogen. Recently, molecular approaches such as multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), variable number tandem repeats (VNTRs), optical mapping and whole genome sequencing (WGS) have been used to resolve many species within the family Enterobacteriaceae (Sabat *et al.*, 2013). However, molecular typing methods currently in use are unable to constantly mirror the pathovars or highlight key virulence factors which can be used to characterize *E. coli* strains (Rasko *et al.*, 2008).

Species identification methods vary and have evolved greatly over the decades. Traditionally, morphological and biochemical phenotypic tests characterized a bacterium into distinct taxonomic groups, which are still required for formal species descriptions today. Subsequently, chemotaxonomic analyses (e.g. lipid components, peptidoglycan composition), and later molecular approaches such as moles % G + C content and DNA-DNA hybridization were used to help define a species and consolidate taxonomic divisions (Schleifer, 2009). Comparative 16S rRNA provided the first molecular phylogenetic basis for microbial classification, changing the taxonomic and clinical diagnostic landscape, and has been referred to as the benchmark for species classification (Schleifer, 2009). Despite the huge impact 16S rRNA sequencing has had, the correct identification of many closely related species, such as those belonging to the Enterobacteriaceae, remains challenging and some – for example, *E. coli* and the genus *Shigella* – are routinely mistaken.

Following more than a decade of concerted research, mass spectrometry (MS)-based identification has been successfully introduced into the clinical laboratory. MS techniques have been previously used to analyze components and products of organisms, such as fatty acids and respiratory molecules (Corina & Sesardic, 1980; Mayberry, 1981; Shah & Collins, 1983). The development of 'soft' ionization techniques such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) permitted the analysis of larger, intact biological molecules (Karas & Hillenkamp, 1988). The primary MS method currently used for microbial identification utilizes 'soft' ionization techniques, in particular MALDI sources coupled with a mass analyzer such as time of flight (TOF) (van Belkum *et al.*, 2013; Dingle & Butler-Wu, 2013). The current implementation of MALDI-TOF MS for bacterial identification primarily utilizes a single mass TOF analyzer, which generates mass spectral profiles which are statistically matched to a database. We established a robust method (Shah *et al.,* 2000; Shah *et al.*, 2002) and reported the first MALDI-TOF MS microbial database using type and reference strains held in the National Collection of Type Collections (Keys *et al.*, 2004, see Chapter 1). In its current guise, the ionized masses of proteins are collated from a range of bacterial species and used to construct a database, against which genus- or species-specific

masses can be identified (Dingle & Butler-Wu, 2013). Experimentally determined masses are statistically compared against the database to obtain identifications.

MALDI-TOF MS has ushered in a new era in clinical diagnostics, resulting in greatly reduced costs and rapid species identification (Carbonnelle *et al.*, 2011; Fournier *et al.*, 2013). It is now routinely used for the identification of species, profiling mainly ribosomal proteins, but generally excludes expressed virulence factors. However, there are some species which consistently remain unresolved by both 16S rRNA and MALDI-TOF, for example, species from the genera *Escherichia* and *Shigella*. Although through manual peak classification and using the software package ClinProTools, it may be possible to resolve some strains of *E. coli* below the species (Everley *et al.*, 2008), the method is highly cumbersome and remains outside the realm of routine use in clinical laboratories. Even more challenging is accurate subspecies identification, which requires the identification of a greater number of markers with greater discriminatory power than those currently being used.

The identification of more proteins/peptides, apart from capturing a more complete profile of an organism's proteome, increases the likelihood of discovering high-resolution markers for bacterial identification. MADLI-TOF has made MS accessible to clinical scientists, paving the way for alternative MS approaches to be used, such as nano-liquid chromatography coupled with high-resolution MS/MS instruments (nano-LC-MS/MS). Nano-LC-MS/MS can be used to profile a greater number of proteins/peptides. Although not as simple to use as current MALDI-TOF methods, they have greater resolution and can identify proteins over a greater mass range (as well as other chemicals in a sample). This enables the identification of more proteins and allows for a more complete characterization of an organism's proteome, from which features used for identification, as well as those associated with virulence factors, can be resolved and used for high-resolution bacterial identification (Al-Shahib *et al.*, 2010; Misra *et al.*, 2012).

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) has been used for decades to proteotype bacterial species (Costas, 1990). The development of protein digestion, such as trypsin, which generates peptide fragments amenable for MS, enabled SDS-PAGE-separated proteins to be identified. This technique is now referred to as GeLC-MS/MS, that is, SDS-PAGE coupled with LC-MS/MS methods, for example, nano-LC-MS/MS. Previous studies have demonstrated the power of GeLC-MS/MS for the characterization of proteins as well as for the identification of bacteria (Al-Shahib *et al.*, 2010; Ho & Reddy, 2010; Misra *et al.*, 2012). To profile a bacterial proteome using nano-LC-MS/MS, the most common method is to lyse the cells from a bacterial culture, releasing all of the cell contents, including the majority of soluble proteins localized in the cytoplasm and periplasm (gram-negative bacterium). Some membrane proteins are also released, but those embedded in the membrane tend to be more hydrophobic and thus rely on more in-depth analysis to capture them. Assuming the lysis methods were successful, the proteins ideally should be enriched, removing any potential contaminants which may conflict with analysis via nano-LC-MS/MS, for example, cell wall debris, lipids, nucleic acids and detergents (if used during lysis). Following protein enrichment, analysis can still be challenging due to the huge variability of protein properties, both chemical and morphological, which can impact their 'suitability' for nano-LC-MS/MS analysis. Therefore, to better profile proteins, somewhat counter-intuitively, the sample is made more complex in terms of the number of detectable features, whereby the thousands of enriched proteins are enzymatically fragmented into tens of thousands

of peptides. It is these peptides that are analyzed via nano-LC-MS/MS and identified using database matching algorithms, which map the peptides to their parent proteins, thus enabling for the proteins to be identified. Using this approach, the proteome of any bacterium can be profiled, and far more proteins and peptides are identified, providing the pool from which higher-resolution markers can be identified

To better describe the resolution and detection limits of 16S rRNA, MALDI-TOF and nano-LC-MS/MS, a selection of challenging bacteria belonging to the family Enterobacteriaceae were analyzed and the results compared.

13.2 Methods

For 16S rRNA analysis, each of the annotated strains was grown for 24 h, on CBA plates at 37 °C, under aerobic conditions. DNA was extracted using the PrepMan Ultra reagent (Applied Biosystems, UK) by suspending 1 µl loop of cells in 60 µl of the reagent. The extracts were then placed on a heating block at 99 °C for 10 min and centrifuged at 10,000 g for 3 min. The supernatant was recovered and used for DNA amplification. The PCR mix contained 1 µl of template, 10 nmol forward primer (ANT1F: 5′ AGA GTT TGA TCC TGG CTC AG3′), 10 nmol reverse primer (1392R: 5′ ACG GGC GGT GTG TAC AAG 3′), 25 µl PCR ready mix (Promega) and 22 µl water. PCR cycling conditions were as follows: initial activation at 95 °C for 2 min, followed by 35 cycles of 95 °C for 45 s, 56 °C for 45 s and 72 °C for 1 min. The final extension step was at 72 °C for 5 min. The size and intensity of the PCR products were confirmed by agarose gel electrophoresis, using 2% E-gels (Invitrogen, UK). The PCR products were cleaned using AMPure XP® PCR Purification Magnetic Bead Kit, and samples were eluted in 50 µl sterile water. Sequencing was performed using forward primer (357 F: 5′ CTC CTA CGG GAG GCA GCA G 3′) and reverse primer (3R: 5′ GTT GCG CTC GTT GCG GGA CT 3′) with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, UK), using the ABI 3730 sequencer.

The sequences generated were taken and searched against the publicly available, curated 16S rRNA resource, 'The Ribosomal Database Project'. It is important to note, as with all identification methods where a database is used for matching, that the database is of high quality. With regard to 16S rRNA, there are 10,000 s of sequences which have been deposited in various databases such as the NCBI and EBI; however, curation and quality control are performed by the depositor. This can lead to spurious matches, which can greatly impact any inferences made. To resolve this, while ensuring access to 16S rRNA data, databases such as RDP and SILVA were created with the aim of providing a curated repository of 16S rRNA sequences including quality metrics and well-annotated taxonomies. To determine the species for the samples in this study, the 16S rRNA sequence data generated in this study were taken and searched against the RDP 16S rRNA database using RDP SeqMatch. Matches were given a 'similarity score' and 'seqmatch score' and used to assign a genus and/or species to the sample, and the highest ranked scores were taken.

As with the 16S rRNA comparisons, the same annotated sample set was taken and analyzed using the Bruker Biotyper MALDI-TOF identification system. Rather than trying to extract, amplify, sequence and annotate the samples, as with the 16S rRNA approach above, the samples were grown aerobically overnight at 37 °C on CBA plates, from which single colonies were recovered with a toothpick and directly spotted in duplicate onto an

Table 13.1 The strains used in this example are listed below, where * = type strain, D = strains used to construct the database and T = strains used to test blindly whether they can be identified using the workflow presented in this study.

Species	Strain accession
Citrobacter freundii	NCTC09750^{*,D}
Enterobacter aerogenes	NCTC10006^{*,D};NCTC10336^{D}
Enterobacter cloacae	NCTC10005^{*,D,T};NCTC11570^{D};NCTC11571^{D}
Escherichia coli	NCTC10538^{D};NCTC09001^{*D,T};NCTC10964^{D};NCTC11151^{D} NCTC13128^{D};NCTC8621^{D};H112160280^{T};H112160540^{T} H112160541^{T};E99518^{T};EDL933^{T};H10302^{T};NCTC12900^{T}
Escherichia fergusonii	DSM13698^{*,D}
Escherichia vulneris	NCTC12130^{D}
Klebsiella pneumoniae	NCTC05050^{*,D};NCTC12463^{D};NCTC5056^{D};NCTC9633^{*,D}
Morganella morganii	NCTC235^{*,D,T};NCTC12286^{D};NCTC12287^{D}
Proteus mirabilis	NCTC11938^{*,D};NCTC10374^{D};NCTC9559^{D}
Proteus vulgaris	NCTC10020^{D,T};NCTC10740^{D};NCTC4175^{D}
Providencia alcalifaciens	NCTC10286^{*,D}
Providencia rettgeri	NCTC11801^{*,D,T};NCTC7475^{D};NCTC7477^{D};NCTC7480^{D}
Providencia stuartii	NCTC11800^{*,D};NCTC12254^{D};NCTC12256^{D}
Serratia marcescens	NCTC10211^{D};NCTC1377^{D};NCTC2446^{D}
Shigella boydii	NCTC12985^{D,T};NCTC9327^{D}
Shigella flexneri	NCTC08192^{D}
Shigella sonnei	DSM5570^{*,D}
Yersinia pseudotuberculosis	NCTC10275^{*,D,T};NCTC10278^{D};NCTC12718^{D}
Campylobacter coli	NCTC11366^{*,D}
Campylobacter fetus	NCTC10842^{*,D}
Campylobacter hyointestinalis	NCTC11608^{*,D}
Campylobacter jejuni	NCTC11351^{*,D}
Campylobacter upsaliensis	NCTC11541^{D}
Enterobacter gergoviae	NCTC11434^{D}
Cronobacter sakazakii	NCTC11467^{*,D};NCTC8155^{D};NCTC9238^{D}
Escherichia hermannii	NCTC12129^{D}
Hafnia alvei	NCTC8105^{*,D};NCTC6578^{D};NCTC8535^{D}
Klebsiella aerogenes	NCTC8172^{D};NCTC9527^{D};NCTC9499^{D}
Proteus penneri	NCTC11972^{D};NCTC12737^{D}
Salmonella enterica subsp. Enterica	NCTC04444^{*,D,T};NCTC00074^{D}
Yersinia enterolitica	NCTC10460^{D};NCTC10938^{D};NCTC11174^{D}

MSP 96 target plate (Bruker Daltonics, UK). Each spot was overlaid with 1 µl of matrix (saturated solution of α-cyano-4-hydroxycinnaminic acid CHCA in 50% acetonitrile and 2.5% trifluoroacetic acid (Bruker Daltonics, UK)). The samples were air-dried to co-crystallize the sample and matrix, and the target plate was then processed using a Microflex LT mass spectrometer (Bruker Daltonics, UK). Data collection was done in 'automatic mode' by collecting 240 laser shots from six different positions within the spot, and MALDI measurements were recorded in a positive linear mode with a mass range of 2–20 kDa. Sample identification was far more automated; using the Bruker Biotyper identification algorithms, for every match a Biotyper score was given which describes how good a match was. Strains were identified with a score ≥2, where scores of n ≥ 2 were considered confident species identifications; scores of 1.7–2 were used to confidently assign a strain to the genus identification; and scores of ≤1.7 gave no confident identification.

Unlike the 16S rRNA and MALDI-TOF identification methods, nano-LC-MS/MS-based identification is still very much in the research and development phase. As such, the methods are comparatively laborious and cumbersome. Despite this, the potentially higher-resolution identification and clinical characterization of strains warrants the investigation of nano-LC-MS/MS as a possible future alternative to MALDI-TOF.

A typical nano-LC-MS/MS method relies, as mentioned earlier, on extracting proteins from a culture, possibly separating the proteins into smaller batches and enzymatically digesting them into peptide fragments. It is these peptides which are separated via LC, ionized and analyzed using MS. Taking the strains in this study as an example, a more detailed method is as follows. Using the same aerobic cultures used for the MALDI workflow, that is, bacterial cells grown aerobically on CBA plates, overnight at 37 °C, rather than taking a single colony, a plate full of cells was taken and suspended in 1 ml of a lysis buffer. There are many buffers to choose from, but when choosing lysis reagents, not all are compatible with LC-MS systems, such as some salts, detergents, lipids and polyethylene glycols. The lysis buffer used in this study was nano-LC-MS/MS compatible and consisted of 7 M urea and 2 M thiourea (GE Healthcare, UK), 4% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS – Melford, Ipswich, UK) and 40 mM dithiothreitol (DTT). The culture–lysis suspension was mixed thoroughly, to which 300 µl of glass beads were added, and the cells were mechanically disrupted using a FastPrep homogenizer (MP Biomedicals, USA). The combination of the lysis buffer and mechanical disruption greatly increases the likelihood of breaking open bacterial cells and releasing their contents, including proteins. The FastPrep, as the name suggests, is a quick method to break cells; the suspension was pulsed for 3 cycles of 20 s at a speed of 4 m/s, followed by a 20 s cooling-down period. The resulting mixture is referred to as the crude extract, and to clarify the extract – that is, remove the glass beads and large cell debris – the crude extract was centrifuged for 30 min at 21,000 g at 4 °C. The supernatant, which contains the protein extract, was removed, quantified using the Bradford assay and stored at –20 °C until required.

To simplify the protein extracts and increase the proteome coverage, the extracts were separated by one-dimensional SDS-PAGE (1D-SDS-PAGE). 10 µg of protein extract was loaded and separated using pre-made MES running buffer (Invitrogen, UK), in accordance with the manufacturer's instructions. The MES buffer is made up of 50 mM MES, 50 mM Tris Base, 0.1% SDS and 1 mM EDTA, pH 7.3. It is the SDS, an anionic detergent which disrupts the non-covalent bonds in a protein, resulting in denatured proteins, which have lost their native shape and applies a negative charge to the denatured proteins, proportional to its mass. This when used in conjunction with

polyacrylamide gels such as the NuPAGE® Novex 4–12% Bis-Tris gels (1.0 mm, 12 well, Invitrogen, UK), enables proteins to be separated according to their mass, into distinct bands, whereby each band will contain one or more proteins. The proteins bands were visualized by staining with InstantBlue (Expedeon Ltd., UK) and each sample lane was cut into 24 pieces and subjected to in-gel tryptic digestion as described below.

Protein digestion is the enzymatic process by which proteins are fragmented into peptides, and it is these peptides which are ultimately used as the 'sample' for nano-LC-MS/MS analysis. Protein extracts (30 µg in 100 mM NH_4HCO_3) were reduced with DTT (5 µl, 200 mM in 100 mM NH_4HCO_3) at 60 °C for 45 min, followed by sulfhydryl alkylation with iodoacetamide (4 µl, 1 M iodoacetamide in 100 mM NH_4HCO_3) at room temperature for 45 min in the dark. Alkylation was stopped by adding DTT (20.0 µl, 200 mM in 100 mM NH_4HCO_3) to neutralize the remaining iodoacetamide and incubated for 45 min at room temperature. Overnight digestion was carried out using trypsin with an enzyme:protein ratio of 1:30 at 37 °C. The digestion was then stopped by adding concentrated glacial acetic acid (4 µl).

The peptide mixture, that is, the 'sample', was separated and analyzed using online nano-LC-MS/MS. In this study, tryptic peptide mixtures from in-gel trypsin digestion were separated using an Ultimate 3000 Dionex nano/capillary HPLC system and analyzed on an LTQ Orbitrap Classic mass spectrometer. HPLC systems, such as the Dionex nano-LC, rely on a stationary phase and an aqueous, mobile phase. The stationary phase captures/attracts the peptides of interest based on one or more properties, for example, charge, size and so on. To release the captured peptides in a controlled manner, that is, to slowly releasing the peptides, the mobile phase is washed over the stationary phase, which in turn carries the peptides for downstream analysis. When a single mobile phase is added, that is, the composition remains static over a period of time, it is referred to as isocratic flow, which is useful for 'simple' samples when separating a few compounds. However, for more complex samples, such as peptide mixtures, two or more mobile phases are used and mixed in different ratios over time, and the changing composition over time is referred to as gradient flow. All of the 'samples' in this study were very complex, comprising 10,000 s of peptides; therefore, a gradient flow was used. The mobile phase in this study comprised Buffer A (2% CH_3CN/0.1% formic acid in water) and Buffer B (10% water/0.1% formic acid in CH_3CN), mixed in different ratios, where the ratio influences the rate at which peptides are released from the stationary phase. Peptide mixtures (10 µl) were initially trapped and desalted in a reversed-phase trap column (Acclaim PepMap C18, 300 µm i.d. × 3 mm, Dionex Ltd., UK) using Buffer A at a flow rate of 25 µl/min and further separated, on the stationary phase, using an analytical C18 reversed-phase (RP) nano-column (75 µm i.d. × 15 cm, 3 µm particles, Dionex Ltd., UK) at 35 °C. The peptides were separated, that is, released over a 60-min gradient: from 10% to 45% solvent B over 45 min, followed by 45% to 90% solvent B over 0.5 min, maintained at 90% B for an additional 5 min and then returned to 10% B over 0.5 min at a flow rate of 300 nl/min. The total run time was 60 min.

When coupled with MS, the separated peptides are passed directly into the MS system for analysis. MS/MS refers to the way in which peptides, that is, the precursor ions, are selected and fragmented for further analysis in the MS. Using the samples in this study as the example, the peptides eluted from the LC gradient are injected into the MS, where a first round of ion selection is performed (MS1), referred to as 'precursor ion selection'. The precursor ions are then fragmented with high-pressure gas, resulting in 'daughter

ions', which are analyzed in MS2. For the LTQ Orbitrap Classic, the mass spectrometer was operated in positive ion mode, and a top six method was used. The precursor ion scan (*m/z* 440–2000) was acquired in the Orbitrap with a resolution $R = 60{,}000$ at *m/z* 400. The six most abundant peptide precursor ions detected in the survey scan were dynamically selected and subjected to collision-induced dissociation (CID) in the linear ion trap to generate MS/MS spectra. General mass spectrometric conditions were set as follows: spray voltage at 1.6 kV, capillary voltage at 38 V, capillary temperature at 200 °C and tube lens at 125 V. Helium was used as collision gas but no sheath or auxiliary gas was applied. Tandem MS (MS/MS) data was acquired in data-dependent mode with automatic switching between MS and MS/MS modes. A normalized collision energy of 35%, an activation of $q = 0.25$ and activation time of 30 ms were applied for ion trap CID in MS/MS acquisition. The lock mass option was enabled using the polydimethylcyclosiloxane ion generated in the electrospray process from ambient air, and protonated $(Si(CH_3)_2O)_6$ at *m/z* 445.120025 was used for internal recalibration in real time to enable accurate mass measurement. Samples were analyzed as technical triplicates.

A huge amount of data is generated by MS instruments, presented in the form of MS profiles, 'spectra'. To interpret the spectra and use them to characterize a protein would be impossible manually; instead, there are many software algorithms (MS search engines) that can be used. In this study, the MS search engine Mascot (version 2.2 – Matrix Sciences) was used, although there are many alternative free programs that can be used, for example, MS Amanda, Tide, Morpheus, Comet, X!Tandem, OMMSSA, to name a few. The algorithms employed in each of these programs differ, although they all follow the same basic procedure: identifying good-quality spectra, creating a reference database comprised of in silico–generated spectra, based on a database of sequences against which the experimental spectra are compared. Although the closest match is considered the most likely, these approaches are greatly dependent on the database, and a variety of quality control methods are applied to determine how good a match is and the chances of getting a false match. One commonly used quality metric is the false discovery rate (FDR), which tells us the likelihood that a spectral match was by chance. This does not explicitly tell you which one is wrong, just the likelihood and therefore lowering the FDR cut-off increases the accuracy. There are other metrics, many of which are algorithm dependent, which can be used to increase the accuracy of a match. In parallel, there are algorithms whose sole purpose is to validate MS data matches, such as Peptide/Protein Prophet. Mascot searches were set up using the following parameters: fragment ion (daughter ion) tolerance of 0.50 Da and a parent ion (precursor ion) tolerance of 10.0 ppm. During enzymatic digestion and preparation for nano-LC-MS/MS analysis, the chemicals used could modify the peptides; therefore, to account for this, the following modifications were included: oxidation of methionine and the iodoacetamide derivative of cysteine, as well as deamidation of asparagine. Protein identifications were accepted where >1 peptide was matched, the charge state was 2+ and was present in all replicates. To further validate the peptides and parent proteins identified, a commercial program, Scaffold (version 3.2, Proteome Software), was used which normalizes the mass spectra and using the search engines match statistics, for example, Mascot ion and identity scores, generates a probability score, referred to as a Discriminant score. It is this Discriminant score which is used to determine cut-offs between good- and bad-quality spectra and subsequent matches. As with the search algorithms, there are many other validation tools that can be used, the most popular being Peptide/Protein Prophet, Percolator and PepDistiller. Using Scaffold

and its proprietary algorithms, probability scores were automatically adjusted to be equivalent to the more conventional FDR, whereby an FDR of 2% was used to filter out the weakest matches.

As stated earlier, one of the primary variables influencing spectral matching to a protein sequence is the search database used. For this study, the NCBInr database was used, which comprises a non-redundant list of all annotated proteins made available through the web resource NCBI (alternatives include EBI/UniProt and DDJB).

13.3 Results

13.3.1 16S rRNA Identification

The current gold standard for microbial identification against which identification methods are benchmarked against is the molecular technique, 16S rRNA sequencing. The strains used in this study were characterized using partial 16S rRNA sequencing, whereby 800 bases were sequenced, ensuring a base quality of > Q20. This dataset was used as the 'benchmark' against which other methods such as MALDI-TOF MS and LC-MS/MS can be compared.

Seventy isolates, representative of clinically relevant Enterobacteriaceae that are encountered in most clinical laboratories, were tested. The RDP taxonomic identifications, which are listed in Table 13.2, demonstrated great variability between what was expected and what was observed, as well as limits of taxonomic resolution. As shown in Table 13.2, the RDP identifications to the genus and species level were approximately 78% and 51%, respectively. The resolution varied greatly, and species identification was particularly challenging among the genera *Proteus*, *Yersinia*, *Enterobacter*, *Providencia*, *Klebsiella*, *Escherichia* and *Shigella*. For example, *P. rettgeri* (NCTC7477) was misidentified as *P. sp/P. vermicola*, *E. aerogenes* (NCTC9735) could only be resolved to the genera *Enterobacter* or *Kluyvera*, *K. aerogenes* (NCTC8172) was identified as *K. pneumonia*, *K. aerogenes* (NCTC9527) was identified as *Raotella sp.*, *E. coli* and *S. flexnerii* was identified as Escherichia/Shigella. Strains NCTC11608 (*C. hyointestinalis*) and NCTC8155 (*C. sakazakii*) could not be identified owing to low-quality sequence products.

13.3.2 MALDI-TOF MS Identification

MALDI-TOF MS identification was performed using the Biotyper method (Bruker) as described in Section 13.2. Strains were identified using the scores as follows: ≥2 (where $n \geq 2$) was considered to give a confident species identification; scores in the range 1.7–2, genus identification; whereas scores ≤1.7 were considered poor and could not be identified. On the basis of these criteria, 91.5% of the isolates (64/70) were identified to the genus level, whereas 88.6% (62/70) matched the expected species identification. The genus *Salmonella* could not be resolved to the species level and was identified as *Salmonella* sp. only. Similar to the 16S rRNA result, *Escherichia coli* could not be resolved to the correct genus or species and was consistently identified as *Escherichia*/*Shigella.* The resulting spectrum of each isolate was compiled to generate a dendrogram (see Figure 13.1). The bacterial strains used to construct the dendrogram are listed in Table 13.2 with their MALDI-TOF-MS-assigned identifications. The dendrogram is based on MALDI-TOF MS spectra and can be used to assign taxonomic boundaries, for example, distinct species clusters. Strains that fall outside of

Table 13.2 Comparison of identifications generated by 16S rRNA (RDP) and MALDI-TOF (Biotyper). No ID = unable to determine genus or species. Bold highlighted = incorrectly identified species when compared to the NCTC designation.

Accession	Strain name (*type strain)	16S rRNA id (RDP)	Biotyper id
NCTC09750	*Citrobacter freundii**	*C. freundii*	*C. freundii*
NCTC10006	*Enterobacter aerogenes**	*E. aerogenes*	*E. aerogenes*
NCTC10336	*Enterobacter aerogenes*	*E. aerogenes*	*E. aerogenes*
NCTC10005	*Enterobacter cloacae**	*E. cloacae*	*E. cloacae*
NCTC11570	*Enterobacter cloacae*	***Enterobacter***	*E. cloacae*
NCTC11571	*Enterobacter cloacae*	***Enterobacter***	*E. cloacae*
NCTC10538	*Escherichia coli*	***Escherichia/Shigella***	***Escherichia/Shigella***
NCTC09001	*Escherichia coli**	***Escherichia/Shigella***	***Escherichia/Shigella***
NCTC10964	*Escherichia coli*	***Escherichia/Shigella***	***Escherichia/Shigella***
NCTC11151	*Escherichia coli*	***Escherichia/Shigella***	***Escherichia/Shigella***
NCTC13128	*Escherichia coli*	***Escherichia/Shigella***	***Escherichia/Shigella***
NCTC8621	*Escherichia coli*	***Escherichia/Shigella***	***Escherichia/Shigella***
DSM13698	*Escherichia fergusonii**	***Escherichia/Shigella***	*E. fergusonii*
NCTC12130	*Escherichia vulneris*	*E. vulneris*	*E. vulneris*
NCTC05050	*Klebsiella pneumonia subsp. ozaenae**	*K. pneumoniae*	*K. pneumoniae*
NCTC1246	*Klebsiella pneumonia*	***K. granulomatis/ K. pneumonia***	*K. pneumoniae*
NCTC5056	*Klebsiella pneumonia*	***K. granulomatis /K. pneumonia***	*K. pneumoniae*
NCTC12286	*Morganella morganii*	*M. morganii*	*M. morganii*
NCTC10374	*Proteus mirabilis*	*P. mirabilis*	*P. mirabilis*
NCTC11938	*Proteus mirabilis**	*P. mirabilis*	*P. mirabilis*
NCTC9559	*Proteus mirabilis*	*P. mirabilis*	*P. mirabilis*
NCTC10020	*Proteus vulgaris*	***P. penneri/P. vulgaris***	*P. vulgaris*
NCTC10740	*Proteus vulgaris*	*P. vulgaris*	*P. vulgaris*
NCTC4175	*Proteus vulgaris*	***P. vulgaris/P. hauseri***	*P. vulgaris*
NCTC10286	*Providencia alcalifaciens**	***P. alcalifaciens/ P. rustigianii***	*P. alcalifaciens*
NCTC11801	*Providencia rettgeri**	***P. rettgeri/P. vermicola***	*P. rettgeri*
NCTC7475	*Providencia rettgeri*	*P. rettgeri*	*P. rettgeri*
NCTC7477	*Providencia rettgeri*	***P. sp./P. vermicola***	*P. rettgeri*
NCTC7480	*Providencia rettgeri*	***P. rettgeri/P. vermicola***	*P. rettgeri*
NCTC11800	*Providencia stuartii**	*P. stuartii*	*P. stuartii*
NCTC12254	*Providencia stuartii*	*P. stuartii*	*P. stuartii*
NCTC12256	*Providencia stuartii*	*P. stuartii*	*P. stuartii*
NCTC10211	*Serratia marcescens*	*S. marcescens*	*S. marcescens*

Table 13.2 (Continued)

Accession	Strain name (*type strain)	16S rRNA id (RDP)	Biotyper id
NCTC1377	*Serratia marcescens*	***Serratia***	*S. marcescens*
NCTC2446	*Serratia marcescens*	***Serratia***	*S. marcescens*
NCTC12985	*Shigella boydii*	***Escherichia/Shigella***	*S. boydii*
NCTC9327	*Shigella boydii*	*S. boydii*	*S. boydii*
NCTC08192	*Shigella flexneri*	***Escherichia/Shigella***	*S. flexneri*
DSM5570	*Shigella sonnei**	***Escherichia/Shigella***	*S. sonnei*
NCTC10275	*Yersinia pseudotuberculosis**	***Yersinia***	*Y. pseudotuberculosis*
NCTC10278	*Yersinia pseudotuberculosis*	***Yersinia***	*Y. pseudotuberculosis*
NCTC12718	*Yersinia pseudotuberculosis*	***Yersinia***	*Y. pseudotuberculosis*
NCTC11366	*Campylobacter coli**	*C. coli*	*C. coli*
NCTC10842	*Campylobacter fetus**	*C. fetus*	*C. fetus*
NCTC11608	*Campylobacter hyointestinalis**	**No ID**	*C. hyointestinalis*
NCTC11351	*Campylobacter jejuni**	*C. jejuni*	*C. jejuni*
NCTC11541	*Campylobacter upsaliensis*	*C. upsaliensis*	*C. upsaliensis*
NCTC9735	*Enterobacte raerogenes*	***Enterobacter/Kluyvera***	*E. aerogenes*
NCTC11434	*Enterobacter gergoviae**	***D. acidovorans***	*E. gergoviae*
NCTC11467	*Cronobacter sakazakii**	*C. sakazakii*	*C. sakazakii*
NCTC8155	*Cronobacter sakazakii*	**No ID**	*C. sakazakii*
NCTC9238	*Cronobacter sakazakii*	*C. sakazakii*	*C. sakazakii*
NCTC12129	*Escherichia hermannii*	*E. hermannii*	*E. hermannii*
NCTC6578	*Hafnia alvei*	*H. alvei*	*H. alvei*
NCTC8105	*Hafnia alvei**	*H. alvei*	*H. alvei*
NCTC8535	*Hafnia alvei*	*H. alvei*	*H. alvei*
NCTC8172	*Klebsiella aerogenes*	*K. pneumoniae*	*K. aerogenes*
NCTC9499	*Klebsiella aerogenes*	***Klebsiella. sp.***	*K. aerogenes*
NCTC9527	*Klebsiella aerogenes*	***Raotella sp.***	*K. aerogenes*
NCTC5050	*Klebsiella pneumonia**	*K. pneumonia*	*K. pneumonia*
NCTC9633	*Klebsiella pneumonia**	*K. pneumonia*	*K. pneumonia*
NCTC12287	*Morganella morganii*	*M. morganii*	*M. morganii*
NCTC235	*Morganella morganii**	*M. morganii*	*M. morganii*
NCTC11972	*Proteus penneri*	***P. penneri/P. vulgaris***	*P. penneri*
NCTC12737	*Proteus penneri*	***P. penneri/P. vulgaris***	*P. penneri*
NCTC04444	*Salmonella enterica subsp. enterica**	*S. enterica*	***Salmonella sp.***
NCTC00074	*Salmonella enterica subsp. Enterica*	*S. enterica*	***Salmonella sp.***
NCTC10460	*Yersinia enterolitica*	*Y. enterolitica*	*Y. enterolitica*
NCTC10938	*Yersinia enterolitica*	*Y. enterolitica*	*Y. enterolitica*
NCTC11174	*Yersinia enterolitica*	*Y. enterolitica*	*Y. enterolitica*

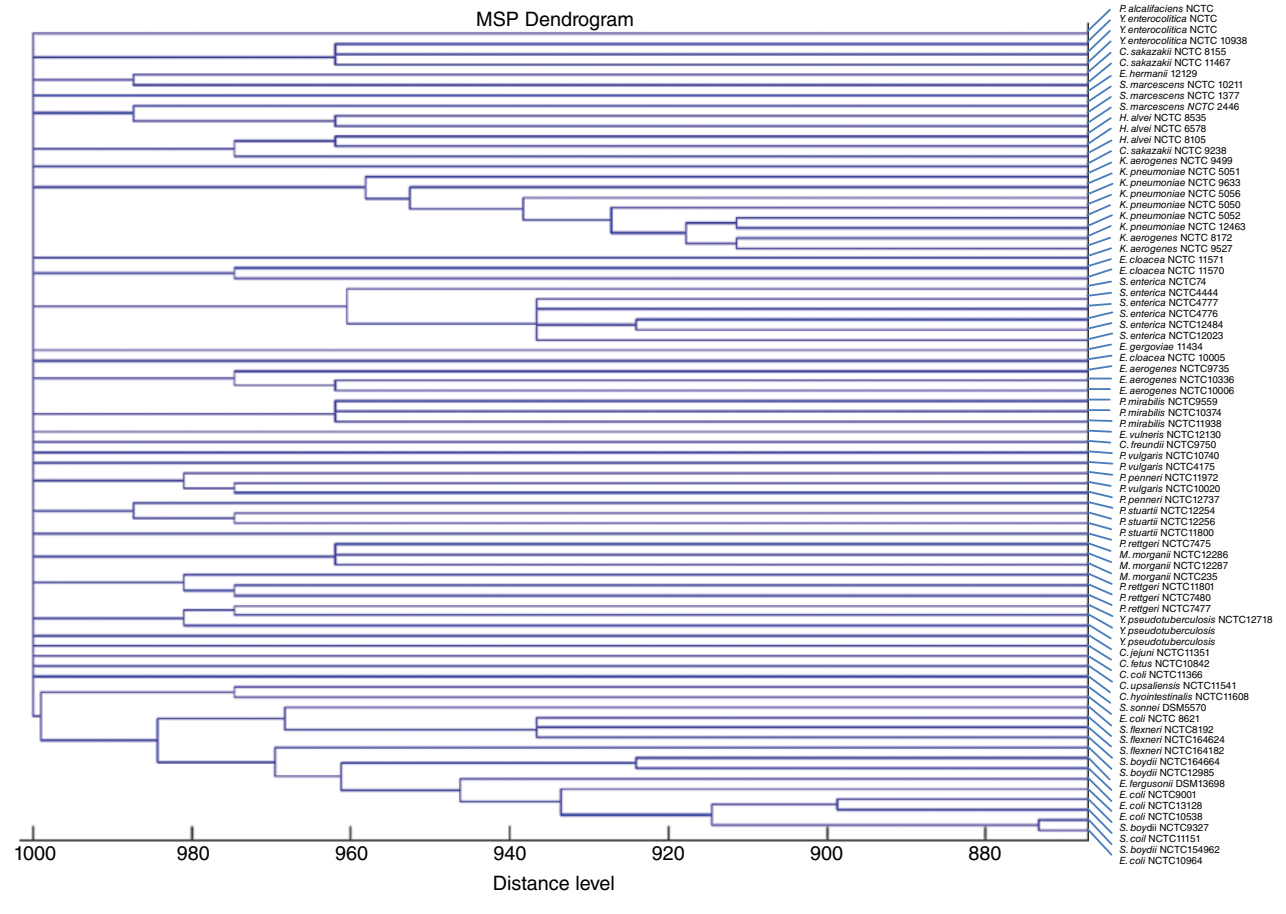


Figure 13.1 Dendrogram describing MALDI spectra similarity between isolates in the Enterobacteriaceae database.

these distinct species clusters may be indicative of species outliers, highlighting possible misidentifications or ill-defined species clusters. The dendrogram (Figure 13.1) shows that whereas many species were grouped together, for example, *Klebsiella sp.*, *Hafnia sp.* and *Salmonella sp.*, species of *Shigella*, *Escherichia* and *Enterobacter* were present in different clusters across the dendrogram.

13.4 Candidate Biomarker Discovery: Shotgun Sampling of Enterobacteriaceae Proteomes by GeLC-MS/MS

To generate the GeLC-MS/MS database, a total of 32 species (70 strains) designated 'D' in Table 13.2 were analyzed. In all, $\sim 1 \times 10^6$ non-redundant peptides were identified from the 70 strains investigated in this study, with an average of approximately 600 proteins (groups) per strain. Following the established marker selection methods (Al-Shahib *et al.*, 2010; Misra *et al.*, 2012), this list was filtered to remove false positives, that is, peptide sequences that were not specific for a designated species. The resultant database comprised peptides from a non-redundant list of 20,678 protein sequences, from all the strains in this study. When investigating candidate genus/species markers, it is important to be able to characterize as much of the proteome as possible, irrespective of subcellular location or function. Analysis of parent protein functions and subcellular locations demonstrate that the identified proteins varied greatly in relation to function and subcellular location (Figures 13.2 and 13.3). The majority of proteins were from the cytoplasm (61%), but many proteins were characterized as from the cytoplasmic membrane (19.1%), outer membrane (1.8%), periplasm (2.8%) or extracellular (1.6%) subcellular locations. The remainder could not be localized to any one subcellular location and were designated unknown.

13.4.1 Database Optimization and Testing

A test panel of strains was created, whereby following the GeLC-MS/MS methodology, eight isolates (Table 13.1, labelled 'T') were analyzed. These were strains from *Escherichia coli*, *Citrobacter freundii*, *Enterobacter cloacae*, *Salmonella enterica*, *Morganella morganii*, *Yersinia pseudotuberculosis* and *Proteus vulgaris*. All of the 'T' strains were reliably identified; however, some of the database matches appeared to be false positives, that is, database markers matching incorrect species. On the basis of the 'T' strains, the number of known false positives ranged from 0.29% (7/2424 markers) to 7.7% (955/12447 markers) per strain. Removal of known false positive markers and re-searching the 'T' strains against the new optimized database (DB-FP) resulted in 100% correct identifications and 0% FDR, thus creating a more reliable database of markers (Table 13.3). To challenge the refined database DB-FP, a further seven *E. coli* isolates, including clinical isolates H112160280, H112160540, H112160541, E99518, EDL933, H10302 and NCTC12900 were analyzed. All seven were identified as *E. coli* with an FDR of 0% (Table 13.3). The number of matches ranged from 30 to 52 DB-FP matches, with a mean of 42 matches.

13.4.2 Demonstrating Capability to Delineate Pathotypes using *E. coli* 0104:H4 as an Exemplar

Using the proteomics approach described in this study, a wide range of proteins were identified, some of which could be used as markers for bacterial identification. To determine whether subspecies resolution, specifically *E. coli* pathovars, could be achieved using the

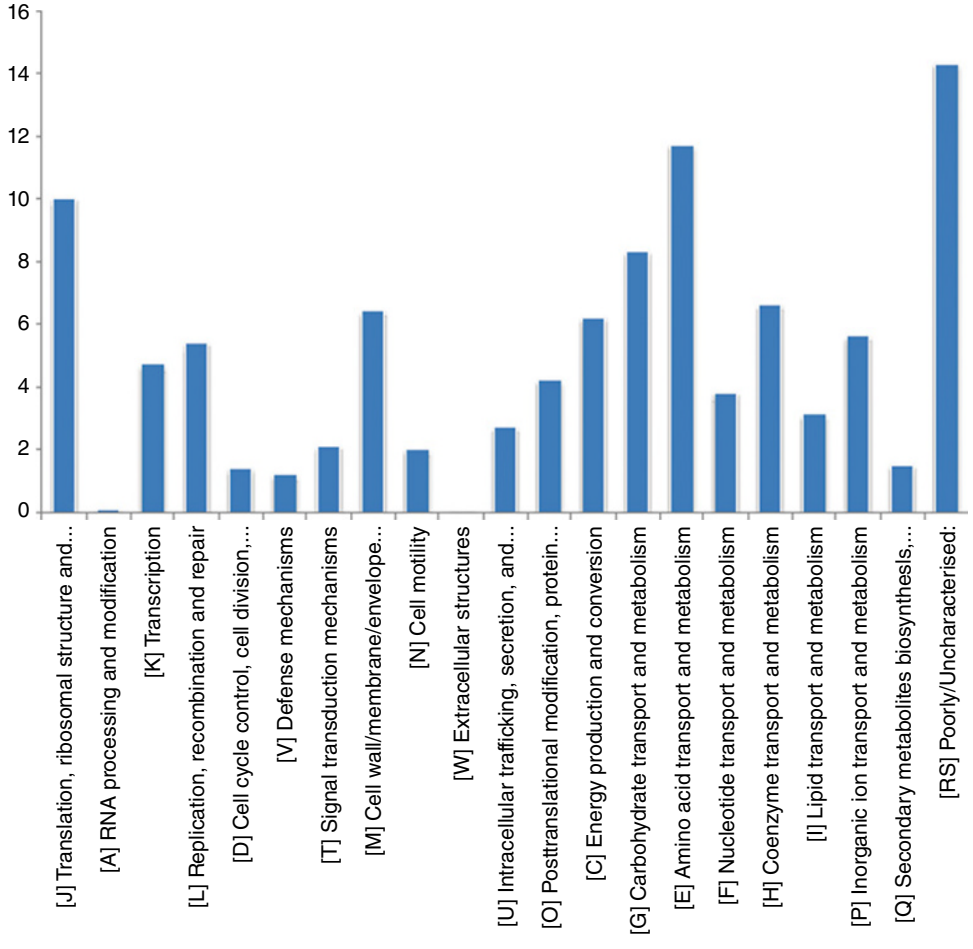


Figure 13.2 The distribution of COG functional categories, for all parent proteins, in the Enterobacteriaceae database, where: J = translation, ribosomal structure and biogenesis, A = RNA processing and modification, K = transcription, L = replication, recombination and repair, D = cell cycle control, cell division and chromosome partitioning, V = defence mechanisms, T = signal transduction mechanisms, M = cell wall/membrane, N = cell motility, W = extracellular structures, U = intracellular trafficking, secretion and vesicular transport, O = post-translational modifications, C = energy production and conversion, G = carbohydrate transport and metabolism, E = amino acid transport and metabolism, F = nucleotide transport and metabolism, H = coenzyme transport and metabolism, I = lipid transport and metabolism, P = inorganic ion transport and metabolism, Q = secondary metabolites biosynthesis, transport and catabolism and RS = poorly characterized/uncharacterized.

nano-LC-MS/MS described in this study, clinical and reference isolates of *E. coli* were selected. Seven strains, including three isolates from the 2011 outbreak in Germany (Brzuszkiewicz *et al.*, 2011; Mellmann *et al.*, 2011; Rasko *et al.*, 2011) designated H112160280, H112160540 and H112160541, as well as four strains representing well-characterized pathovars, E99518 (EAEC), EDL933 (EHEC), H10302 (EHEC) and NCTC12900 (non-toxigenic, shiga-like toxin (SLT) negative, EHEC), were investigated. All of these isolates, as previously described, could be identified to the species level with 0% FDR.

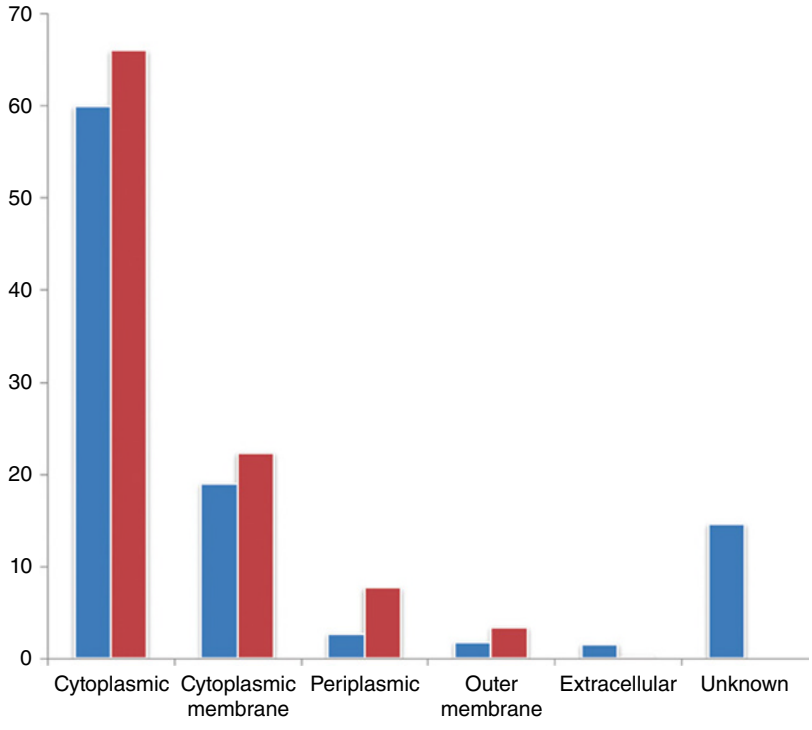


Figure 13.3 Percentage distribution of predicted subcellular locations for all parent proteins in the Enterobacteriaceae database (blue) versus *E. coli* K12 MG1655 (red).

Table 13.3 Comparison of identification methods: 16S rRNA, Biotyper and the DB-FP database. Poorly resolved identifications are in bold.

Accession:Strain name	16S rRNA id (RDP)	Biotyper id	DB-FP
NCTC10005: *E. cloacae*	*E. cloacae*	*E. cloacae*	*E. cloacae*
NCTC9001: *E. coli*	***Escherichia/Shigella***	***Escherichia/Shigella***	*E. coli*
H112160280: *E.coli*	***Escherichia/Shigella***	***Escherichia/Shigella***	*E. coli*
H112160540: *E.coli*	***Escherichia/Shigella***	***Escherichia/Shigella***	*E. coli*
H112160541: *E.coli*	***Escherichia/Shigella***	***Escherichia/Shigella***	*E. coli*
E99518: *E. coli*	***Escherichia/Shigella***	***Escherichia/Shigella***	*E. coli*
EDL933: *E. coli*	***Escherichia/Shigella***	***Escherichia/Shigella***	*E. coli*
H10302: *E. coli*	***Escherichia/Shigella***	***Escherichia/Shigella***	*E. coli*
NCTC12900: *E. coli*	***Escherichia/Shigella***	***Escherichia/Shigella***	*E. coli*
NCTC235: *M. morganii*	*M. morganii*	*M. morganii*	*M. morganii*
NCTC10020: *P. vulgaris*	***P. penneri/P. vulgaris***	*P. vulgaris*	*P. vulgaris*
NCTC11801: *P. rettgeri*	***P. rettgeri/P. vermicola***	*P. rettgeri*	*P. rettgeri*
NCTC12985: *S. boydii*	***Escherichia/Shigella***	*S. boydii*	*S. boydii*
NCTC10275: *Y. pseudotuberculsis*	***Yersinia***	*Y. pseudotuberculosis*	*Y. pseudotuberculosis*
NCTC04444: *S. enterica subsp. Enterica*	*S. enterica*	***Salmonella sp.***	*S. enterica*

Proteomic characterization of these isolates resulted in the identification of approximately 2000 proteins, ≥1 peptide ($n \geq 2$ technical replicates) and FDR <1%. Comparing the list to the VFDB compilation of virulence factors (VFs), virulence-related proteins were selected (Table 13.4). A total of 71 VFs were identified including SLT, serine proteases, intimin, aggregative proteins and resistance-associated proteins (antibiotic and heavy metal). Virulence-related proteins according to pathotype (Table 13.4), that is, EHEC, EAEC, EHEC and EAEC, in addition to outbreak (OB)-specific VFs, revealed

Table 13.4 Clustering of protein identifications identified by each sample cohort, where outbreak = OB strain, EHEC and EAEC represent OB strain-related pathovars. Numbers = unique peptide identifications per parent protein (white background = 0, pink = 1 and red = ≥ 2 peptides identified per parent protein).

	Outbreak strains' prefix H112160			EAEC	EHEC			
Identified proteins	**280**	**540**	**541**	**E99518**	**EDL933**	**H10302**	**NCTC12900**	
Per-activated serine protease autotransporter enterotoxin EspC	77	81	73	49	1	1	0	Shared
Uncharacterized membrane lipoprotein clustered with Tellurite resistance proteins TehA/TehB	54	39	44	26	0	1	1	Shared
Tellurium resistance protein TerD	24	24	20	11	9	7	4	Shared
Serine protease pic autotransporter	24	24	23	18	12	12	13	Shared
Outer membrane stress sensor protease DegQ, serine protease	22	21	20	10	5	7	8	Shared
Tellurium resistance protein TerD	16	13	13	13	6	5	5	Shared
21 kDa hemolysin precursor	16	15	9	14	10	13	0	Shared
Tellurium resistance protein	15	7	7	10	9	7	9	Shared
P pilus assembly/Cpx signalling pathway, periplasmic inhibitor/ zinc- resistance-associated protein	14	11	10	10	0	1	0	Shared
Tellurium resistance protein TerA	12	5	5	9	5	4	3	Shared
Tellurite resistance protein TehB	11	11	8	7	6	5	6	Shared
Membrane fusion protein of RND family multidrug efflux pump	11	9	8	7	3	4	5	Shared
Type III restriction enzyme domain protein	10	4	5	4	0	1	3	Shared
ABC-type multidrug transport system, ATPase component	7	10	8	4	3	2	3	Shared
Lactam utilization protein LamB	4	7	4	6	0	2	0	Shared
Beta-lactamase	3	1	1	1	2	2	2	Shared
Polymyxin resistance protein	3	2	0	1	2	1	1	Shared
Polymyxin resistance protein PmrJ	0	13	14	11	6	7	8	Shared

Table 13.4 (Continued)

	Outbreak strains' prefix H112160			EAEC	EHEC			
Identified proteins	**280**	**540**	**541**	**E99518**	**EDL933**	**H10302**	**NCTC12900**	
Outer membrane usher protein	0	0	0	11	0	0	0	EAEC
Chaperone protein Agg3D	0	0	0	2	0	0	0	EAEC
Enterohemolysin	0	0	0	2	0	0	0	EAEC
StcE	0	0	0	0	12	9	16	EHEC
Translocatedintimin receptor Tir	0	0	0	0	9	0	12	EHEC
Intimin (fragment)	0	0	0	0	7	0	0	EHEC
Outer membrane usher protein FimD	0	0	0	0	4	4	0	EHEC
Type III restriction enzyme	0	0	0	0	3	0	14	EHEC
Fimbrial protein	0	0	0	0	3	1	4	EHEC
Tir chaperone	0	0	0	0	3	1	1	EHEC
Non-LEE-encoded type III secreted effector	0	0	0	0	2	0	4	EHEC
EspM3 protein	0	0	0	0	2	1	2	EHEC
SLT 1A subunit (fragment)	0	0	0	0	2	2	0	EHEC
EspA	0	0	0	0	2	2	0	EHEC
EspB	0	0	0	0	2	1	0	EHEC
Beta-lactamase	0	0	0	0	1	1	4	EHEC
Non-LEE-encoded type III secreted effector	0	0	0	0	1	1	2	EHEC
Porcine attaching-effacing associated protein	0	0	0	0	0	0	0	EHEC
EspD	0	0	0	0	0	18	0	EHEC
Hemolysin	0	0	0	0	0	0	0	EHEC
Non-LEE-encoded type III secreted effector	0	0	0	0	0	0	2	EHEC
Outer membrane usher protein AggC	27	13	17	0	0	0	0	OB
Beta-lactamase	21	22	20	0	0	0	0	OB
IncI1 plasmid pilus assembly protein PilP	5	3	3	0	0	0	0	OB
Mercuric resistance operon regulatory protein	3	2	2	0	0	0	0	OB
Tellurite resistance protein TerB	2	1	1	0	0	0	0	OB
Beta-lactamase (fragment)	2	1	0	0	0	0	0	OB
AatC ATB binding protein of ABC transporter	2	3	1	0	0	0	0	OB
Chaperone protein AggD precursor	2	0	0	0	0	0	0	OB
Sulfonamide resistance protein (Fragment)	1	2	0	0	0	0	0	OB
Chaperone protein FimC	0	1	0	0	0	0	0	OB
Beta-lactamase	35	24	20	0	0	0	0	OB

(*Continued*)

Table 13.4 (Continued)

	Outbreak strains' prefix H112160			EAEC	EHEC			
Identified proteins	**280**	**540**	**541**	**E99518**	**EDL933**	**H10302**	**NCTC12900**	
Type III restriction-modification enzyme helicase subunit	20	18	12	10	0	0		OB & EAEC
14 kDa aggregative adherence fimbriae I protein	17	11	11	1	0	0	0	
Zinc resistance-associated	17	5	4	7	0	0	0	
Beta-lactamase class C and other penicillin-binding proteins	16	8	9	11	0	0	0	
Multiple antibiotic resistance protein MarR	12	11	12	4	0	0	0	
AatB	12	10	11	10	0	0	0	
Arsenical resistance operon repressor	6	0	2	1	0	0	0	
Chaperone protein AggD	4	4	2	4	0	0	0	
Tellurium resistance protein TerX	3	2	3	4	0	0	0	
putative phage inhibition, colicin resistance and tellurite resistance protein	2	1	1	3	0	0	0	
Copper resistance protein C	2	1	0	1	0	0	0	
Type III restriction-modification system methylation subunit	2	2	0	4	0	0	0	
Putative tellurium resistance protein	2	1	1	1	0	0	0	
Metal-dependent hydrolases of the beta-lactamase superfamily I; PhnP protein	0	1	1	1	0	0	0	
Per-activated serine protease autotransporter enterotoxin EspC	22	36	36	0	0	0	1	OB & EHEC
SLT II subunit A precursor	6	5	6	0	1	2	0	
SLT II subunit B precursor	3	3	4	0	1	1	0	
Chaperone FimC	0	2	1	0	1	0	1	

pathovar-specific groups. Eighteen VFs were present in all pathotypes, three were EAEC specific, 18 VFs were EHEC specific, the OB strains shared 14 with EAEC, four with EHEC and 11 were OB specific.

To further assess the resolution and accuracy for the identification of clinically relevant targets, the SLT was selected for analyses. The SLTs produced by some *E. coli* strains can lead to the severe, life-threatening disease haemolytic-uremic syndrome (HUS), which is characteristic of the pathotype EHEC (Kaper *et al.*, 2004). To determine whether the approaches described in this study can be used to differentiate *E. coli* strains on the basis of the presence or absence of SLTs, the SLT-negative strain NCTC12900 and the SLT-positive strains EDL933 and H10302 were analyzed. We were able to correctly

resolve the strains into SLT-positive and SLT-negative groups, whereby zero SLT-associated peptides were identified for strain NCTC12900, whereas four SLT-associated peptides were identified for strain EDL933 and five for H10302 (Table 13.4).

13.5 Discussion

Identification of isolates using partial 16S rRNA sequencing and MALDI-TOF MS varied in identification coverage. Using 16S rRNA sequencing, 78% of isolates resolved to the expected genus level and 51% to species identification. Over the last decade, most clinical laboratories have utilized 16S rRNA as its benchmark for species identification, successfully employing the PCR products used here for clinical diagnostics. However, in accordance with current investigations (Charnot-Katsikas *et al.*, 2014; Schulthess *et al.*, 2014), the present study reports a similar trend of greater accuracy of species identification using MALDI-TOF MS and places this technology at the forefront of microbial diagnostics. Furthermore, its speed of analysis, significantly lower cost, negligible sample preparation, minimal biomass required and automation to result compared to 16S rRNA is transforming the workflow in clinical laboratories.

One reason for the improved accuracy is that MALDI-TOF databases are well curated and rigorously controlled, particularly those provided by the vendors, and represent several thousands of spectra and multiple profiles for each species (see Chapter 2). By contrast, 16S rRNA databases, particularly public repositories such as NCBI, can be prone to error, in terms of sequence quality and annotation. To compensate, alternative specialist 16S rRNA databases have been implemented which are well curated and present sequence quality information, for example, RDP, SILVA and GreenGenes (Cole *et al.*, 2014; DeSantis *et al.*, 2006; Yilmaz *et al.*, 2014). Because MALDI-TOF MS largely targets the stable ribosomal complement of the cell's proteins and are in high copy number, most clinical samples can be analyzed directly without the need for prior protein extraction.

However, some taxa where 16S rRNA cannot delineate species also remain unresolved using MALDI-TOF MS, and higher-resolution techniques are required (Table 13.2 and Figure 13.1). In addition to low subspecies resolution, the current MALDI-TOF MS approaches and 16S rRNA identification approaches do not provide information regarding the pathogenic potential of a strain. As a result, additional tests are often needed to fill this gap, relying on a battery of molecular and phenotypic assays to supplement a taxonomic identification to better characterize an isolate. Characterization of pathogens is essential, that is, information relating to VFs such as toxin production or antimicrobial resistance. This in parallel with taxonomic identification dictates prognosis. Proteomics has the potential to consolidate many tests into a single assay, to identify an organism and characterize VFs, in addition to revealing the subtle mechanistic changes associated with the phenotype.

High-resolution mass spectrometers were used to perform GeLC-MS/MS analysis of the proteome of strains listed in Table 13.1. Protein mixtures were fractionated by SDS-PAGE followed by protein digestion using trypsin for nano-LC-MS/MS. Accurate determination of the peptide mass from MS1 spectra and amino acid sequence information from MS/MS spectra is used to better describe the peptide and thus the protein, enabling more accurate annotation of the proteome. Biases towards a specific subcellular location or functional category were shown to be minimal, demonstrating the potential to identify a wide range of proteins. As shown in Figures 13.2 and 13.3, this approach is

capable of identifying a much greater number of functional categories and is not limited to one subcellular location.

In this study, high-resolution MS platforms were used to profile and determine a list of potential species-specific markers, with specific emphasis on *E. coli*. Analysis of the test strains, using the DB-FP, demonstrated 100% correct identification (0% FDR) of the test *E. coli* strains (Table 13.4). Mirroring WGS, whole proteome profiling goes beyond the protein mass fingerprint employed by current MALDI-TOF MS identification systems. Tandem MS enables characterization of protein sequences, as well as increasing proteome coverage. The genome sequence of isolates from the *E. coli* serovar O104:H4 from an outbreak in Germany, 2011, were sequenced in depth using various platforms and annotated (Brzuszkiewicz *et al.*, 2011; Mellmann *et al.*, 2011; Rasko *et al.*, 2011). Comparative genomic studies highlighted VFs indicative of the outbreak strain, for example, serine proteases, SLT, heavy metal and antimicrobial resistance features (Loman *et al.*, 2012). Following the proteomic workflow in this study, we were able to corroborate it with comparative genomic analyses, identifying the same outbreak-specific features. In parallel, we characterized pathogenic strains of *E. coli* to the similar functional resolution as WGS, resolving pathovar-specific features, enabling subspecies resolution (pathotyping) of pathogenic *E. coli*. Although protein identifications varied between strains, proteins expected to be present in the outbreaks strains, EAEC and EHEC groups, were observed. The strain variations (Table 13.4) highlight the need to prioritize proteins/peptides which are conserved across each group for identification purposes, but for strain profiling demonstrates the need for further work to better characterize protein expression in these individual strains. In this study, the LC-MS/MS analyses took approximately 13 h per strain, which is comparable to current sequencing run times of 4 h to11 days. LC-MS/MS can be optimized to be gel free, greatly reducing the overall analysis time from 13 h to < 4 h per strain while maintaining similar protein numbers (Nagaraj *et al.*, 2011).

It is clear that proteomic characterization of microbial pathogens is comprehensive and with optimization can be performed in similar run times as WGS. This presents an exciting opportunity to integrate the proteome with the genome, which is ushering in a new field of proteogenomics for high-resolution characterizing of human pathogens. Confirming that the protein is present, expressed and mapped to a predicted gene provides key information on expression of virulence. Importantly, it also demonstrates that proteomics is able to mirror genomics and characterize outbreak strains in real time and highlight virulence-associated proteins. Protein quantitation, such as label-free approaches, can be performed and introduces a new dynamic to characterizing outbreak strains in real time, raising the possibility of a tandem taxonomic identification and a quantifiable phenotypic assay for the characterization of bacteria.

The iterative improvements to the database demonstrate the importance of optimization to ensure the greatest likelihood of identifying taxonomic and VF markers. Utilizing high-throughput genomic data integrated with proteomic datasets enabled a novel method to optimize proteomic marker databases. For future database optimization, it would be prudent to follow the MALDI-TOF MS approach that encompasses more replicates and more clinical strains to create MS/MS-specific curated database for identification and strain characterization. Although a limited dataset, bacterial identification was resolved to the subspecies of pathogenic *E. coli*, specifically 'pathotyping', and the proteomic datasets described in this study corroborate with the genomic studies performed during the outbreak (Brzuszkiewicz *et al.*, 2011; Mellmann *et al.*, 2011; Rasko *et al.*, 2011), but with the distinct advantage of

characterizing expressed features related to the observed phenotype. This greatly reduces the caveats associated with phenotypic predictions based solely on the genome sequence.

Part B

13.6 Highly Pathogenic Biothreat Agents

In 2001, the dispersal of letters spiked with *B. anthracis* spores through the US mail system highlighted the pressing need for rapid, robust and highly sensitive assays for the detection of high-risk pathogens such as *B. anthracis, Y. pestis*, *F. tularensis*, *B. pseudomallei, B. mallei* and the toxin-producing organism *C. botulinum*. These organisms are among the most dangerous and are classified as potential biological weapons due to their high mortality rate, impact on public health and ease of dissemination.

For example, *F. tularensis* poses a serious concern as a biological weapon mainly because it is one of the most infectious pathogenic bacteria known. Inhalation of as little as 10 organisms is sufficient to cause serious illness and death. An aerosol release of *F. tularensis* in a populated area would be expected to result in the abrupt onset of large numbers of cases of acute, non-specific respiratory febrile illness which would begin 3 to 5 days after exposure. In 1970, the World Health Organization (WHO) expert committee reported that if 50 kg of virulent *F. tularensis* was dispersed as an aerosol over a metropolitan area with a population of 5 million, there would be an estimated 250,000 incapacitating casualties, including 19,000 deaths.

It is important that suitable systems are in place for monitoring and detecting the deliberate release of high-risk pathogens (e.g. contaminated water or food) and for early detection in exposed individuals. At present, many assays (mainly PCR based) exist, but these are often highly specific to a particular organism rather than targeting a range of biothreat agents. In addition, many of these high-risk pathogens share close genetic relationships with other bacteria. For example, *B. anthracis* is closely related to *B. thuringiensis* (an insect pathogen), and it is difficult to clearly distinguish between them.

A good detection system requires specific targets for various organisms of interest that are derived from stable markers with specificity to multiple strains and the ability to distinguish them from closely related species.

A common feature of biothreat agents is that they contain proteins/peptides; even the potent neurotoxin (BoNT) produced by *C. botulinum* is a protein. Proteomics coupled with MS has emerged as a rapid technique for sensitive characterization of peptides and proteins in complex mixtures. Proteomics coupled with MS provides a robust strategy for biomarker discovery by allowing a more diverse range of samples to be tested and does not necessarily require microbial samples to be cultured. It enables a large proportion of the proteome to be screened, from which unique peptides (biomarkers) can be identified for the species of interest based on the presence or absence of the peptide in closely related organisms. Due to its sensitivity, MS-based methods have been demonstrated to detect organisms in a complex background, needing only 10^4 cells to be present, translating to attomoles of marker material. Although MS-based proteomic analyses enable a large proportion of the proteome to be screened, it is dependent on expression of the protein. Unique biomarkers may appear specific to a particular organism due to their absence in closely related strains, but they may be encoded on the genome at the DNA and not expressed.

Targeted nanoLC-MS/MS methods to rapidly quantify peptide/protein markers have emerged as a valuable tool to bridge the gap between biomarker discovery and verification. The specificity and sensitivity of this approach have been enhanced by the introduction of stable isotope-labelled peptide internal standards to compensate for variation in sample preparation, such as recovery and the influence of differential matrix effects. Because fragmentation patterns and retention times of the native tryptic peptide are identical to its isotope-labelled counterpart, and the absolute amount of internal standard is known, not only can the selected peptides be identified from protein extracts, but also their precise abundances.

Although MS-based proteomic analyses enable a large proportion of the proteome to be screened, the resultant markers are restricted to those that are expressed and of sufficient abundance. Complementary molecular approaches can be applied utilizing whole genome sequence data and direct sequencing of specific targets. The integration of both methodologies for marker discovery improves the robustness of identifying and characterizing candidate biomarkers, ensuring species specificity independent of gene expression.

Bioinformatic analysis aided by scripts helps deal with the large amount of data generated by MS and for comparisons between large numbers of samples. In addition, the vast amount of genome sequence data now available means that any potential biomarker can be validated by in silico analysis and compared to the genome sequences of the species of interest as well as closely related species to ensure that the predicted marker is unique.

Due to the impact of genetic diversity, it is essential to provide a second, confirmatory identification for the detection of a high-risk pathogen within a sample. Complementation with molecular assays such as PCR and direct sequencing of the target region can provide the confirmation needed to confidently determine the presence of a particular organism with stable genetic markers, providing a more robust target for molecular assays.

Taking each high-priority biothreat agent in turn, *B. anthracis*, *Y. pestis*, *F. tuarensis*, *C. botulinum*, *B. pseudomallei* and *B. mallei* each had to be grown in conditions favourable for high yields of both genetic and protein material. Therefore, a good understanding of the organism and its taxonomy is needed to ensure (1) successful growth of the organism and that (2) representative strains are chosen which enable discovery and validation of the range of markers specific to the species.

Please note that these organisms are highly pathogenic, and suitable specialist facilities are needed to grow and handle them safely.

13.7 *Bacillus anthracis*

The genus *Bacillus* comprises a heterogeneous collection of gram-positive, rod-shaped bacteria that share the ability to form endospores under aerobic or facultative anaerobic conditions. *Bacillus anthracis* is an endospore-forming, toxin-producing bacterium and is the causative agent of anthrax, an acute and often fatal disease in humans and other mammals. Its spore-forming capability and highly pathogenic nature, particularly inhalation anthrax, has made it one of the most likely biological warfare agents. Although there are phenotypic assays which can assist in differentiating *B. anthracis* from other bacilli, due to the high genetic relatedness with other species, mainly species collectively known as the *B. cereus* group, it has proved a challenge to do so rapidly and consistently (Misra *et al.*, 2012).

13.7.1 Methods: Strain Panel

Eight *Bacillus anthracis* strains (NCTC 109, NCTC 1328, NCTC 2620, NCTC 5444, NCTC 7752, NCTC 7753, NCTC 8234 and NCTC 10340) and seven reference strains representing the *B. cereus* group of species (*B. cereus* NCTC 11143, NCTC 11145, ATCC 10987; *B. thuringiensis* DSM 6029, DSM 2046; *B. weihenstephanensis* DSM 11821 and *B. mycoides* NCTC 7586) were used.

13.7.2 Whole Cell Protein Extraction

All strains were cultured on nutrient agar plates overnight at 37 °C for 18 h. After this, the cells were suspended in 10 ml of nutrient broth to an OD_{600} of approximately 1.0. 15 µl of this was used to inoculate 30 ml of broth in a sterile 50 ml centrifuge tube. The tubes containing broth cultures were sealed and incubated with shaking at 150 rpm for 18 h at room temperature.

30 ml of culture was centrifuged at 12,000 g for 20 min at 4 °C to harvest cells. Broth supernatant was removed by repeated aspiration with a 5 ml pipette, and the pellet and remaining broth were transferred to a 1.5 ml microcentrifuge tube. The cells within the microcentrifuge tubes were centrifuged at 12,000 g for 10 min at 4 °C to re-pellet cells. Excess broth was removed by aspiration with a 1 ml pipette. 1 ml of 10 mM Tris-HCl, pH 8.0, containing 1 mM EDTA and 1 mM PMSF, was added to the microcentrifuge tube, and the cells were washed using a 1 ml pipette. The cells were pelleted by centrifugation at 12,000 g for 10 min at 4 °C.

The washed, pelleted cells were re-suspended in a lysis solution comprising 500 µl of sucrose buffer (0.5 M sucrose, 20 mM maleic acid/KOH, pH 6.5, 20 mM $MgCl_2$) containing 6 µg/µl lysozyme and 1 mM PMSF. Enzymatic lysis was done for 1 h at 37 °C. After lysis, the resultant cell mixture was pelleted and washed as follows: centrifugation at 3000 g for 10 min at 4 °C and one wash in 1 ml of 0.5 M sucrose, 20 mM maleic acid/KOH, pH 6.5, 20 mM $MgCl_2$ and 1 mM PMSF. The washed cell mixture was collected by centrifugation at 3000 g for 10 min at 4 °C. The resultant pellet was re-suspended in 100 µl of a solubilization cocktail of 30 mM Tris-Cl pH 8.5, 7 M Urea, 2 M Thiourea, 4% CHAPS and 40 mM DTT (60 mg/ml) and left at room temperature for 30 min. The solubilized suspension was clarified by centrifugation at 21,000 g for 30 min at 21 °C, and the supernatant was filtered through 0.2 um anopore vectra spin filters at 8000 g. The supernatant containing cellular proteins ('protein extracts') was removed and placed in 1.5 ml tubes and stored at −20 °C.

> As highlighted above, *B. anthracis* is pathogenic and before the extracts could be safely handled outside of specialist containment laboratory (C.L.3), they had to be proved to be 'non-viable'.

Viability studies were performed on *B. anthracis* extracts. 15 µl of protein extract was re-suspended in 15 ml nutrient broth and on nutrient agar and incubated at 37 °C for 7 days. 15 µl of broth was plated onto nutrient agar and incubated at 37 °C for 48 h. All plates and broths were examined for growth that resembles that of *B. anthracis*. Only samples that show negative growth were transferred out of the C.L.3 laboratory.

13.7.3 One-Dimensional SDS-PAGE and In-Gel Digestion of Bacterial Proteins

Note: Four biological replicates were prepared for 1D SDS-PAGE analysis.

The protein concentration for each extract was determined by the Bradford assay using BSA as a standard.

On the basis of the Bradford quantitation, 10 µg of total protein was taken from each extract and separated on a 10% Bis-Tris gel using 2-(N-morpholino)-ethanesulfonic acid or 3-(N-morpholino)-propanesulfonic acid SDS running buffer in accordance with the manufacturer's instructions and stained with colloidal Coomassie. Each gel lane was cut into 10 equal pieces and destained with 50% (w/v) methanol for 3 × 20 min, dehydrated using 100% acetonitrile for 10 min and dried for 5 min. Gel pieces were rehydrated and proteins reduced by addition of 10 mM DTT and subsequently alkylated by addition of 55 mM iodoacetamide. In-gel digestion of proteins was performed with 10 ng μl^{-1} porcine trypsin overnight at 37 °C. Tryptic peptides were extracted with 2% acetonitrile/0.1% trifluoroacetic acid (TFA) for 1 h with gentle agitation. The resultant peptide mixture was stored at −80 °C until needed.

13.7.4 In-Solution Protein Digestion Directly from Protein Extracts

By Bradford quantitation, 20 µg of total protein was precipitated with 100% ice-cold acetone and kept overnight at −20 °C. The precipitates were pelleted by centrifugation at 21,000 g at 4 °C, washed three times with ice-cold 90% methanol for 5 min, and re-pelleted at 21,000 g at 4 °C. The washed pellets were re-suspended in 50% TFA/50% 50 mM ammonium bicarbonate containing 10 mM DTT and incubated at 60 °C for 30 min. Each washed pellet was treated with 55 mM iodoacetamide for 45 min at room temperature in the dark. The resultant protein mixtures were diluted fivefold with 25 mM ammonium bicarbonate and digested with sequence grade-modified trypsin at a sample-to-protease ratio of 50:1 overnight at 37 °C. Proteolysis was quenched by addition of TFA to a final concentration of 0.1%.

Note: For internal standardization and validation of these procedures, 5 ng of horse myoglobin was spiked into each sample directly after quantitation.

13.7.5 1-D Nanoflow LC-MS/MS, Data-Dependent and Targeted MS Analysis

Peptides were separated online by RP HPLC on an Ultimate 3000 (Dionex, UK). All LC-MS/MS experiments were performed on a Thermo Finnigan LTQ Orbitrap equipped with a nanoelectrospray ion source and nano-LC pico-tip (20 um ID 10 um tip diameter).

Peptides were loaded and desalted on a RP trap column (C18, 300 um i.d. × 3 mm) with a flow rate of 25 µl min^{-1} using 2% acetonitrile, 0.1% formic acid normal phase solvent. Analytical separation was achieved on a RP nano-column C18, 75 um i.d. × 15 cm at a flow rate of 0.3 µl min^{-1} and mobile phase gradient of 2%–30% acetonitrile for 60 min and 30%–50% for a further 30 min. Data was acquired in data-dependent mode using the Thermo Finnigan Xcalibur software (version 2.0.6). The precursor ion scan MS spectra (*m/z* 450-1600) were acquired in the Orbitrap with a resolution R = 60,000 at *m/z* 400, with the number of accumulated ions being 5×10^5. The six most intense ions were isolated and fragmented and detected in the linear ion trap (number of accumulated ions: 3×10^4). Use of data-dependent LC-MS/MS methodology dynamic exclusion was set to a duration of 30 s. Initially, angiotensin was spiked into each

tryptic digest, providing an internal standard for the analytical platform. A peptide mixture consisting of five tryptically digested yeast proteins was used to ensure that sampling efficiency was reproducible over the course of the entire analysis. The lock mass option was enabled for accurate mass measurements in MS mode. Three polydimethylcyclosiloxane ions generated in the electrospray process from ambient air (protonated $(Si(CH_3)_2O)_6$, *m/z* 445.120025, were used for internal recalibration in real time (providing mass stability of less than 2 ppm throughout the entire analysis without the need to stop and recalibrate the instrument midway). Only mass tags, produced using Thermo Electron Protein calculator software, were generated for peptides of interest. These peptide sequences were entered into the program, and masses corresponding to their 2+ and 3+ charge states were calculated, accounting for variable oxidation of methionine and deamidation of asparagine and static modification of cysteine by carboxyamidomethylation.

Note: Five technical replicates were performed using LC-MS/MS.

13.7.6 Bioinformatic Workflow for Biomarker Detection

13.7.6.1 In-House Protein Database Setup

The UniProt database was supplemented with additional *Bacillus* sequences from unfinished genomes submitted to the NCBI database (Accession numbers: NZ_AAEK00000000, NZ_ABQK00000000, NZ_ABQL00000000, NZ_ABQM00000000, NZ_ABQN00000000, NZ_AAOX00000000, NZ_AAAC00000000, NZ_ACMW00000000, NZ_ABLT00000000, NZ_ABLH00000000, NZ_ABKF00000000, NZ_ABDM00000000, NZ_ABDN00000000, NZ_ABDM00000000, NZ_ABCZ00000000, NZ_ABCF00000000, NZ_AAJM00000000, NZ_AAES00000000, NZ_AAEN00000000, NZ_AAEO00000000, NZ_AAEP00000000, NZ_AAER00000000 and NZ_AAEQ00000000). Redundancy in the NCBInr database and UniProt/Swiss-Prot databases was removed using the program CD-HIT (version 3.1.2).

13.7.7 Protein/Peptide Marker Identification

DTA files were generated from tandem mass spectra, with charge states assigned using BioWorks version 3.3. De-isotoping was not performed. MS/MS samples were analyzed using Sequest to search an in-house version of the UniProt/Swiss-Prot database and NCBInr database. In silico trypsin digestion of protein sequences in the database was implemented in Sequest. Sequest search parameters were set to a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 10.0 ppm. Oxidation of methionine and the iodoacetamide derivative of cysteine plus de-amidation of asparagine were specified as variable modifications. Peptide identifications were accepted if they could be established at less than 0.01 probability value as specified by the BioWorks browser (version 3.3.1 SP 1) and Xcorr values of 1.5, 2.0 and 2.5 for charge states 1, 2 and 3, respectively. Protein identifications were accepted with at least one identified peptide. The data generated from the Sequest searches from each strain was collated and parsed in to a PostgreSQL (version 8.1.18) database, using in-house Perl scripts. Unique *Bacillus anthracis*–specific peptides were initially identified in silico, through NCBI BlastP analysis (parameters adjusted for short sequences) against the non-redundant NCBI database (release 161), supplemented with in-house short-length peptide sequences derived from LC-MS/MS analysis. The program Scaffold (version 2.6) was used to compare the data from the strains analyzed in this study, to validate the spectral

quality of protein/peptide identifications and compare LC-MS/MS datasets from each sample. This approach provides an experimentally verified list of candidate peptide markers that can accurately identify the species of interest was determined. This list of candidate peptide markers were further validated using molecular approaches (below).

13.7.8 Procedure for DNA Extraction

Strains were prepared similar to the protein preparation methods, but there were slight differences which helped increase the yield of high-quality DNA. The growth, preparation and extraction of DNA was performed as follows.

Strains of *Bacillus* were initially cultured on nutrient agar overnight at 37 °C for 18 h. From this plate, cells were suspended in 10 ml of nutrient broth to give an OD_{600} of around 1. 30 µl of this was used to inoculate 30 ml of nutrient broth. These were grown for 24 h with shaking at 30 °C.The OD_{600} was checked, and if it was between 0.7 and 1, DNA extraction was performed. 30 ml of culture was centrifuged at 10,000 g for 10 min at 4 °C to harvest cells, and the supernatant was removed. The cell pellet was washed 3 times with 1 ml 1 × TE buffer containing 1 mM PMSF, followed by centrifugation at 8000 g for 10 min at 4 °C. The cell pellet was re-suspended in 300 µl 1 × TE and 300 µl lysozyme at 50 mg/ml. The suspension was incubated for 1 h at 37 °C followed by centrifugation at 3000 g for 10 min, and the supernatant was removed. The pellet was re-suspended in 750 µl 1 × TE, 90 µl 10% SDS and 15 µl proteinase K and incubated at 37 °C for 2 h. Extracts were clarified by centrifugation at 21,000 g for 30 min at 21 °C, and the supernatant was collected.

Viability studies were performed on *B. anthracis* extracts. 15 µl of the supernatant was re-suspended in 15 ml nutrient broth and on nutrient agar and incubated at 37 °C for 7 days. 15 µl of broth was plated onto nutrient agar and incubated at 37 °C for 48 h. All plates and broths were examined for growth that resembled that of *B. anthracis*. Only samples that showed negative growth were transferred out of the C.L.3 laboratory.

13.7.9 DNA Extraction

CTAB/NaCl was heated at 65 °C for 30 min. Next, 300 µl of 5 M NaCl was added to the supernatant and mixed well, followed by the addition of 240 µl of the preheated CTAB NaCl to each lysate. The samples were mixed thoroughly and incubated at 65 °C for 10 min. 800 µl of chloroform/iso-amyl alcohol (24:1v:v) was added and the sample vortexed and centrifuged for 10 min at 12,000 g. The upper layer of extract was removed and placed in a clean Eppendorf tube. An equal volume of Tris-buffered phenol chloroform was added and the sample centrifuged for 10 min at 12,000 g, and the upper part of the extract was again removed and placed in a clean Eppendorf tube. An equal volume of isopropanol was added and mixed on a shaker for 10 min until the DNA was visible. The sample was centrifuged at 13,000 g for 1 min to pellet the DNA and the supernatant removed. The pellet was washed twice with 70% ethanol, followed by centrifugation at 13,000 g for 1 min and the supernatant removed. The DNA was left to air-dry and then re-suspended in 50–100 µl of molecular-grade water.

13.7.10 Genetic Validation of Candidate Peptide Biomarkers

For each candidate peptide the nucleotide sequence of the marker was obtained from the fully sequenced *B. anthracis* str. Ames ancestor genome (Genbank Accession

AE017334.2) in FASTA format. In addition, for each candidate peptide, nucleotide sequences of the corresponding protein were obtained from publicly available genome sequences for members of the *B. cereus* group in FASTA format (AE017334.2, AE016879.1, CP001598.1, AE017225.1, AE017194.1, AE016877.1, CP000001.1, CP000227.1, CP001186.1, NC_005957.1, NC_008600.1 and NC_010184.1). Nucleotide sequences were translated into their corresponding amino acid sequence in Bioedit v7.0.4.1 software. The peptide biomarker was aligned with the protein sequences using the multiple alignment algorithm clustalW as implemented in Bioedit v7.0.4.1 software (Figure 13.1). The aligned protein sequences were back-translated to the corresponding nucleotide sequence, and the consensus sequence of the alignment was determined. Using the consensus sequence determined for each marker, PCR primers were designed with a Tm between 45 °C and 60 °C and G + C content between 30% and 60% within areas of sequence conservation flanking the biomarker region. The program Oligoanalyzer 3.0 was used to ensure that the Tm and G + C content of the primer sequences fell within the above criteria and that primers did not form homo/heterodimers. Target biomarker regions were amplified using a 50 µl reaction volume containing 25 µl GoTaq® Green Master Mix, 10 pmol each of the forward and reverse primer for the target of interest (target-specific primer sequences are listed in Table 13.1), 1 µl template DNA and the volume made up to 50 µl using nuclease-free water. PCR amplification was performed using a touchdown PCR protocol with the following cycling conditions: 1 cycle of 94 °C for 2 min; 10 cycles of 94 °C for 1 min, 60 °C for 1 min decreasing by 1 °C to 50 °C for each cycle, 72 °C for 2 min; 25 cycles of 94 °C for 1 min, 50 °C for 1 min, 72 °C for 2 min; and 1 cycle of 72 °C for 10 min and held at 4 °C. Amplified samples were run on precast 1.2% agarose E-gels following the manufacturer's instructions and visualized using a UV transilluminator to check for the presence of a PCR product and verify that the product was of the predicted size by comparison with a 100 bp DNA ladder. PCR products were purified using the AMPure system following the manufacturer's instructions on the Biomek NxP robot.

Cleaned PCR products and target-specific PCR primers listed in Table 13.1 were used for sequencing with the ABI Big Dye Terminator Kit v3.1. DNA sequencing reactions were purified with the CleanSeq magnetic beads on the Biomek NxP robot. Automated sequence detection was performed on a 48-capillary ABI 3730 genetic analyzer. Sequence trace files for each strain were acquired in.abi format and the forward and reverse sequences assembled using Bionumerics v4.01 software. The nucleotide sequence data obtained was aligned against the corresponding biomarker nucleotide sequence as described previously. The resulting alignment was manually analyzed in nucleotide and amino acid sequence format to determine if:

1) The peptide biomarkers were unique to the species of interest.
2) The biomarkers were genetically stable (e.g. the DNA sequence of the biomarker contains no silent mutations in the species of interest; see Figure 13.4).

Eight *Bacillus anthracis* strains (NCTC 109, NCTC 1328, NCTC 2620, NCTC 5444, NCTC 7752, NCTC 7753, NCTC 8234 and NCTC 10340) and seven reference strains representing the *B. cereus* group of species (*B. cereus* NCTC 11143, NCTC 11145, ATCC 10987; *B. thuringiensis* DSM 6029, DSM 2046; *B. weihenstephanensis* DSM 11821 and *B. mycoides* NCTC 7586) were used.

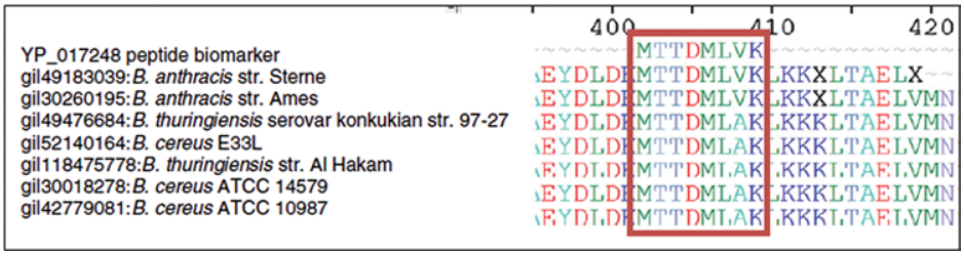


Figure 13.4 An example of a peptide biomarker, from the protein YP_017248 (boxed in red) aligned against amino acid sequences for a selection of members from the B. cereus group.

Table 13.5 Primer sequences used for PCR amplification and DNA sequencing of each target biomarker region.

Target biomarker	Approximate amplicon size (bp)	Primer sequences
YP_016767	730	YP_016767_R 5′-TGACGAAATCAATGGAATTTCT-3′ YP_016767_R 5′-TCATGTCCTTGCTCAACTCCTT-3′
YP_020743	780	YP_020743_F 5′-CAACTAGAGTTCAAGAGTCAAAGTT-3′ YP_020743_R 5′-CTTTTAATCCAGCTGCCCTT-3′
YP_016916	770	YP_016916_F 5′-TACAAACAACGAACGGACCACT-3′ YP_016916_R 5′-TCTCATAGTCTTTTTCGAATGACA-3′
YP_017487	930	YP_017487_F 5′-ATGGGCTGATGAGCTCGTAAA-3′ YP_017487_R 5′-AAATGGCTCTTCCCAAGCAA-3′
YP_020804	490	YP_020804_F 5′-CCAAAAGTATTTATGGATTACGAAA-3′ YP_020804_R 5′-TTCAATGAGTAAGTCACCCGTT-3′
YP_019807	300	YP_019807_F 5′-AACGYAGTTCAATGAACTCTTGG-3′ YP_019807_R 5′-GCCRTGYACCATAGTATATGC-3′
YP_021007	450	YP_021007_F 5′-TTGGWCCATCAAGTTCACATACAGC-3′ YP_021007_F 5′-ATTGARGCTACHGCTGCWATAGC-3′
YP_020576	500	YP_020576_F 5′-ATGAAGCCTTTTCTATCAGC-3′ YP_020576_R 5′-TTGATAAATAGCCGTAATGATAGACC-3′
YP_019000	1000	YP_019000_F 5′-ATTGTAACHGAGCGTACRAARTGG-3′ YP_019000_R 5′-ATGTGTTTTTCGTAGCCATTGC-3′
YP_016771	1470	YP_016771_F 5′-TTGATAGTGATGTGCCGTGTCC-3′ YP_016771_R 5′-GTTTGCAGGTAGTTTGGCAGTACG-3′
YP_018907	300	YP_018907_F 5′-TTAGAAGGGATGACAGATGC-3′ YP_018907_R 5′-CGAGTTACAGAACCTATAGAACG-3′
YP_021350	980	YP_021350_F 5′-TTCTCGTATATCCAACGATGG-3′ YP_021350_R 5′-GAATTCGTTAACTTCTGTACMGC-3′

Table 13.5 (Continued)

Target biomarker	Approximate amplicon size (bp)	Primer sequences
YP_018345	830	YP_018345_F 5′-ACAAACATTAACAGCATGCGT-3′ YP_018345_R 5′-ACCATTTGTGGAGTTTGGTTT-3′
YP_020375	580	YP_020375_F 5′-TTGGAGYGGRAAGAWTTGTCC-3′ YP_020375_R 5′-AATTCACCACGTGTYACAAGRCG-3′
YP_018337	870	YP_018337_F 5′-GCATCATTAATTGAAAAAGGTGC-3′ YP_018337_R 5′-CGAATTTTCGAATGTAATTCTTCC-3′
YP_017248	500	YP_017248_F 5′-ATACACCTCATACATTGCWGAAGG-3′ YP_017248_R 5′-TAATTCATCACTAACTCAGCTGTGAGC-3′
YP_019177	600	YP_019177_F 5′-GAGCAATTCCAACTTCYACHTTAGG-3′ YP_019177_R 5′-ATCATRAAGRTGTCCTTCVTCC-3′
YP_022460	870	YP_022460_F 5′-AATGATGTTGCGTACTCACC-3′ YP_022460_R 5′-AGTGCAGGATGTGTATAACC-3′
YP_020399	600	YP_020399_F 5′-TTATGCTACAAAGGCTCTCG-3′ YP_020399_R 5′-ATGACCTAAAYGATAATACTCACACGCC-3′
YP_020336	500	YP_020336_F 5′-TGTWGCWATCGGTTATTCAAATGG-3′ YP_020336_R 5′-AGGTACRATCGTTGAYATTCARYTACC-3′
YP_018595	200	YP_018595_F 5′-GCTACATTGCARCATTACAATCC-3′ YP_018595_R 5′-CAGRTTRTATYYCTGTTCACC-3′
YP_022433	290	YP_022433_F 5′-AACGAAGGTAAACGGATTACG-3′ YP_022433_R 5′-CGATGGAGTTAAACCATAAGG-3′
YP_016683	530	YP_016683_F 5′-AAGTTCATCTTTGYGAGCAATGTGC-3′ YP_016683_R 5′-CTTCATCCATGGACTAATYGC-3′
YP_022180	1010	YP_022180_F 5′-AAGGTACAAGCAATTCTTACGG-3′ YP_022180_F 5′-TGGTAATCTACATGATAAGCAGC-3′
YP_021575	200	YP_021575_F 5′-AGTACGACATGYCCRATTAGTCAYGG-3′ YP_021575_R 5′-TTCCARTGMGAYGTATKCCATACTACC-3′
YP_020435	1000	YP_020435_F 5′-ATGTACCGAAGGATATTCCAATGG-3′ YP_020435_F 5′-TACGGCATTAGGATTGTTACCAATGG-3′
YP_020438	Part 1: 1310	YP_020438_1F 5′-CCAAGAAAATGAACATGTTAAACC-3′ YP_020438_1R 5′-CCTTCATCATATGCCTCAATTTGG-3′
	Part 2: 1140	YP_020438_2F 5′-ATGCGAAGGTTATTAAAGTTGCATGC-3′ YP_020438_2R 5′-AGTTTGTATAGTCTTCGGAGTTCC-3′

13.8 Summary of Results

Table 13.6 Functional categorization of *B. anthracis*–specific peptide markers (* = peptide derived from *B. anthracis* unique proteins, § = plasmid origin, † = no suitable primer region found and ‡ = possessed a silent mutation).

Locus tag	Peptide biomarker	DNA biomarker	Protein function
YP_016767	LLVSENIIR	TTACTAGTTTCTGAAAAT ATTATTCGC	Gluconate operon transcriptional repressor
YP_020743*	EANEGLTVLEK	GAAGCTAATGAAGGGTTA ACAGTTCTTGAGAAG	Hypothetical protein
YP_016916	DIDHIGINR	GATATCGATCACATTGGT ATTAATCGT	ATP-dependent DNA helicase PcrA
YP_019173†	DWEAVHEYTSD SDVMK	GATTGGGAAGCTGTTCAT GAGTATACATCAGATAGC GATGTTATGAAA	Acetyltransferase, GNAT family
YP_017487	SADLVQGLVDD AVEK	TCTGCTGACTTGGTTCAA GGCTTAGTTGACGACGCT GTTGAAAAA	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
YP_020804	LVSIGELQPDGNR	CTAGTATCAATCGGAGAGCTT CAGCCAGATGGAAATCGT	Pyruvate carboxylase
YP_019807	DTYDAAMEIVK	GACACTTATGATGCTGCA ATGGAGATTGTAAAG	Isochorismatase family protein
YP_021007‡	MGQVACQLFR	ATGGGGCAAGTTGCTTG TCAGCTGTTTCGT	L-serine dehydratase, iron-sulfur-dependent, beta subunit
YP_020576*	AIHEQLEAVEGGLR	GCTATTCATGAGCAATTGGAA GCAGTGGAAGGGGGACTAAGA	Aspartate kinase I
YP_019000	QLSDVAEEDVNR	CAGTTAAGTGATGTTGCA GAAGAAGATGTAAATCGG	3-hydroxyisobutyryl-CoA hydrolase
YP_016771	MENLQIGVVGV GVMGK	ATGGAAAATTTACAAATT GGAGTTGTTGGTGTGGGC GTTATGGGAAAA	6-phosphogluconate dehydrogenase
YP_018907	PMNMDIFK	CCGATGAACATGGATAT TTTTAAA	Heat shock protein, Hsp20 family
YP_021350	LVDLIASHLPIK	TTAGTGGATTTAATTGCTT CTCACTTGCCAATTAAA	ATP-dependent protease La 1
YP_018345	EFAELQEQIDYIAK	GAATTTGCAGAATTACAAGAAC AAATTGATTACATTGCTAAA	Flagellin
	ASQNTQDGMSLIR	GCTTCACAAAATACACAAGACG GGATGTCATTAATCCGT	
YP_020375	DDITGGWYEAAIR	GATGATATAACGGGTGGTTGG TACGAAGCAGCCATTCGT	N-acetylmuramoyl-L-alanine amidase
YP_018337*	SEDLPVSAEIR	TCAGAAGATCTTCCTGTT AGTGCAGAGATTAGA	TPR/glycosyl transferase domain protein
	ASEATVLLLIVG PGEK	GCTTCCGAAGCAACTGTA CTATTGTTAATAGTTGGA CCAGGAGAAAAA	

Table 13.6 (Continued)

Locus tag	Peptide biomarker	DNA biomarker	Protein function
YP_017248	MTTDMLVK	ATGACAACGGACATGT TGGTTAAG	NAD-dependent epimerase/dehydratase
YP_019177	DGESFASGHIEK	GATGGAGAGAGTTTTGCT TCTGGTCACATCGAGAAA	Hydroxyacylglutathione hydrolase
YP_022460§	SEITDLIGIR	TCAGAGATTACAGATTTA ATTGGAATTAGA	RelA/SpoT domain-containing protein
	NPDYYIDVNNYK	AATCCTGATTATTACATCG ACGTTAATAATTATAAA	
	NNDLFNFK	AATAATGATTTATTTAAT TTCAAG	
	AGEQFVIDFK	GCCGGCGAACAATTTG TTATTGATTTTAAA	
YP_017265†	QGVEAELDFLR	CAAGGGGTAGAGGCTGA ACTAGATTTTTTACGA	Hypothetical protein
YP_020399*	DIQIYTELLAGK	GACATTCAAATTTATACC GAACTACTAGCAGGGAAA	Prophage LambdaBa01 TPR domain-containing protein
YP_020336	AKEWYENTVL	GCAAAAGAATGGTATGA AAATACAGTACTA	Phospholipase/ carboxylesterase family protein
YP_018595	EFQPLENCVK	GAATTCCAACCATTAGAA AACTGTGTGAAA	Nitroreductase family protein
YP_022433§	LYSTADLAEE LEMTK	TTATATTCAACTGCTGATTTA GCAGAGGAATTAGAGATGAC AAAG	ImpB/MucB/SamB family protein
YP_016683	NLENQLSEYR	AATCTTGAAAATCAGCTTAGT GAGTATAGA	Unknown
YP_022180*	ALLTYASYFK	GCGCTACTTACTTATGCATCG TATTTTAAA	Hypothetical protein
YP_021575	AVPAYCLYVQDAR	GCAGTGCCAGCGTATTGCTTAT ACGTACAAGATGCGAGA	Putative cytoplasmic protein
YP_020435*	DPFDVYEEIR	GATCCTTTTGATGTATAT GAAGAGATAAGA	Hypothetical protein
	QIEDSLLSDDG EFHLK	CAAATTGAAGATTCATTACTAT CTGATGATGGTGAGTTTCAT TTGAAA	
YP_020438*	LMLSILGAQGT EEDLNK	CTTATGTTATCAATTTTAGGAG CTCAAGGAACAGAAGAAG ATTTAAATAAA	Prophage LambdaBa01, acyltransferase, putative
	QIIIIPGIMGSK	CAAATAATAATTATACCTG GAATAATGGGAAGTAAA	
	DWFPVMMSAEK	GATTGGTTTCCAGTTATG ATGTCTGCAGAGAAA	
	NVLNVFSTDSAG NAEATK	AATGTATTGAATGTATTTTCA ACGGATAGTGCTGGAAAT GCAGAAGCTACTAAA	
YP_018571†	DSSTSVNVYLGK	GACTCTAGTACTAGTGTA AATGTTTATCTTGGAAAA	Putative S-layer protein

13.9 *Yersinia pestis*

The genus *Yersinia* encompasses a group of gram-negative, rod-shaped bacteria that are facultative anaerobes. The genus *Yersinia* contains three pathogenic species: *Y. pestis*, the causative agent of plague, and the enteric food- and water-borne pathogens *Y. pseudotuberculosis* and *Y. enterocolitica.* DNA–DNA hybridization studies and 16S rRNA sequence comparisons indicate that *Y. pestis* is highly related to *Y. pseudotuberculosis. Y. pestis* is described as being a clone that evolved from *Y. pseudotuberculosis* 1,500 to 20,000 years ago and has been subdivided into three biovars (Antiqua, Medievalis and Orientalis).

Y. pestis is primarily a disease of rodents or other wild mammals that is usually transmitted by fleas. *Y. pestis*, the etiologic agent of plague, exists in three forms: bubonic, septicaemic and pneumonic, and is considered to be a potential bioweapon because it is relatively easy to acquire from the environment and can be effectively dried and dispersed in an aerosol form.

13.10 Method: Strain Panel

Eight *Y. pestis* strains (NCTC 5924, NCTC 5923, NCTC 10029, NCTC 10030, NCTC 144, NCTC 10329, NCTC 2868 and NCTC 2028) and eight reference strains representing closely related species *Y. enterocolitica* (NCTC 11598 and 11600), *Y. frederiksenii* (NCTC 11470), *Y. intermedia* (NCTC 11469), *Y. kristensenii* (NCTC 11471) and *Y. pseudotuberculosis* (NCTC 10275, 8315 and 8487) were used.

13.10.1 Procedure for Whole Cell Protein Extraction

Strains were grown on Columbia blood agar for 24 h. A 1 × 10 µl loop of cells was re-suspended in 1 ml of L6 lysis buffer (guanidium isothiocyanate boom extraction buffer) in a microcentrifuge tube. The tube was vortexed to ensure an even suspension and incubated at room temperature for 45 min.

Viability studies were performed on *Y. pestis* extracts. 100 µL of the extract was plated onto Columbia blood agar, incubated at 37 °C and checked for growth at 24 and 48 h.

The cell suspensions were centrifuged at 6000 rpm for 30 min, and the supernatant collected and stored at −80 °C for further analysis. The supernatant was treated with the GE 2D clean-up kit following the manufacturer's instructions. Three biological replicates were prepared for 1D PAGE analysis

13.10.2 One-Dimensional SDS-PAGE and In-Gel Digestion of Bacterial Proteins

The procedure is the same as that described for *B. anthracis.*

13.10.3 One-Dimensional Nanoflow LC-MS/MS, Data-Dependent and Targeted MS Analysis

The procedure is the same as that described for *B. anthracis.*

13.10.4 Bioinformatic Workflow for Biomarker Detection

The identification of *Y. pestis*–specific biomarkers was carried out as described for *B. anthracis*. However, the UniProt database was supplemented with additional *Yersinia* sequences from unfinished genomes submitted to the NCBI database (Accession numbers: NZ_ABCD00000000, NZ_ABAT00000000, NZ_AAYV00000000, NZ_AAYU00000000, NZ_AAYT00000000, NZ_AAYS00000000, NZ_AAYR00000000, NZ_AAUB00000000, NZ_AAOS00000000, NZ_AALF00000000, NZ_AALE00000000, NZ_AALC00000000, NZ_AALD00000000, NC_004839, NC_004836, NC_009139, NC_006323, NC_002120, NC_012208, NC_004837, NZ_ADDC00000000, NC_010377, NC_005570, NC_005017, NC_002144, NC_004835, NC_004564 and NC_011759).

13.10.5 Genetic Validation of Peptide Biomarkers

In silico genetic validation was performed on all *Y. Pestis*–specific biomarkers. For candidate peptide, the nucleotide sequence of the marker was obtained from the fully sequenced *Y. pestis* biovar Antiqua genome (Genbank Accession CP000308.1) in FASTA format. For each candidate peptide, nucleotide sequences of the corresponding protein were obtained from publicly available genome sequences for members of the species *Y. pestis* and closely related species *Y. enterocolitica* and *Y. pseudotuberculosis* in FASTA format (Genbank Accession: AM286415.1, CP000901.1,CP000308.1, AL590842.1, CP001585, CP001589, AE009952.1, CP000305.1, CP000668.1, CP001593.1, CP001608, AE017042.1, CP000720.1, BX936398.1, CP001048.1 and CP000950.1).

Nucleotide sequences of the proteins and the peptide biomarker were translated into their corresponding amino acid sequence in Bioedit v7.0.4.1 software. The peptide biomarker was aligned with the protein sequences using the multiple alignment algorithm clustalW as implemented in Bioedit v7.0.4.1 software. The resulting alignment was manually analyzed in nucleotide and amino acid sequence format to determine if:

1) The peptide biomarkers were unique to the species of interest.
2) The biomarkers were genetically stable (e.g. the DNA sequence of the biomarker contains no silent mutations in the species of interest).

13.11 Summary of Results

Table 13.7 Summary of *Y. pestis*–specific peptides. In silico genetic validation of the biomarkers found all markers to be genetically stable.

Locus tag	Peptide biomarker	DNA biomarker	Protein function
YP_650749	TDKEGVFHT EWMA	ACAGATAAAGAAGGTGTATT CCACACTGAGTGGATGGCG	6-phosphogluconate dehydrogenase
YP_652810	QIDALLEEAA AQLLR	CAGATTGACGCACTGTTGG AAGAGGCGGCAGCGCAG CTACTCCGT	Pantoate beta-alanine ligase
YP_652827	AEAAPAAAGGGL PGILPWPK	GCCGAAGCTGCACCAGCAGCG GCTGGCGGCGGCCTGCCGGG CATATTGCCTTGGCCAAAA	Dihydrolipoamide acetyltransferase

13.12 *Fransicella tularensis*

The genus *Francisella* encompasses a group of pathogenic gram-negative, strictly aerobic bacteria that are facultative intracellular parasites of macrophages. *F. tularensis* is the causative agent of tularaemia, a serious disease that presents in two main forms: pneumonic and ulceroglandular. As one of the most infectious pathogenic bacteria known, it poses a serious concern as a biological weapon due to its high virulence and ease of spread by aerosol (inhalation of as few as 10 organisms can cause disease). Currently, four known subspecies of *F. tularensis* that differ in virulence and geographical distribution are recognized: *F. t. tularensis* (type A), *F. t. holarctica* (type B), *F. t. mediasiatica*, and *F. t. novicida*, with *B. tularensis* subsp. *tularensis* being the most virulent.

13.13 Method

13.13.1 Strain Panel

Two clinical strains of *F. tularensis* and one reference strain of *F. philomiragia* subsp. *philomiragia* (DSMZ 7535) were used.

13.13.2 Procedure for Whole Cell Protein Extraction

Francisella tularensis was initially cultured on Columbia blood agar, at 37 °C for 48 h. From this plate, 1 × 10 µl loop of cells was re-suspended in 20 ml of nutrient broth. The OD_{600} was adjusted to approximately 0.1–0.2, and the starting broth culture was incubated for 18 h at 37 °C with constant shaking. After ~18 h, when the OD_{600} was measured to be ~1.0, a purity plate was prepared and incubated for 48 h at 37 °C. 20 ml of culture was centrifuged at 10,000 g for 10 min at 4 °C to harvest cells. Broth supernatant was removed and the pellet washed once with 1× TE buffer containing 1 mM PMSF. Cells were centrifuged at 10,000 g for 10 min at 4 °C to re-pellet them. The cell pellets were re-suspended in 1.5 ml of 1× TE buffer containing 6 µg/µl lysozyme and 1 mM PMSF and incubated for 1 h at 37 °C. Cells were collected by centrifugation at 3000 g for 10 min at 4 °C. The cell pellets were washed twice in 1 ml 1× TE buffer containing 1 mM PMSF. Cells were collected by centrifugation at 3000 g for 10 min at 4 °C. Cells were frozen at −20 °C until the purity of the culture was confirmed.

The cell pellet was re-suspended in 100 µl of a solubilization cocktail (30 mM Tris-Cl pH 8.5, 7 M Urea, 2 M Thiorurea, 4% CHAPS and 70 mM DTT) by reciprocation with a pipette. The suspension was incubated for 30 min at room temperature and clarified by centrifugation at 21,000 g for 30 min at 21 °C. The supernatant was filtered through 0.2 µm Anopore vectra spin filters at 8000 g until the filtrate ran through. The supernatant containing cellular proteins was removed and placed in 1.5 ml tubes. Extracts were stored at −20 °C until completion of the viability study.

Viability studies were performed on *F. tularensis* strains. 50 µl of protein extract was re-suspended in 20 ml nutrient broth, spotted on a CBA plate and incubated at 37 °C for 48 h. All plates and broths were examined for growth that resembles that of *F. tularensis*. Only samples that showed negative growth were used for further analysis.

Three biological replicates were prepared for 1D PAGE analysis.

13.13.3 One-Dimensional SDS-PAGE and In-Gel Digestion of Bacterial Proteins

The procedure is the same as that described for *B. anthracis*.

13.13.4 One-Dimensional Nanoflow LC-MS/MS, Data-Dependent and Targeted MS Analysis

The procedure is the same as that described for *B. anthracis*.

13.13.5 Bioinformatic Workflow for Biomarker Detection

The identification of *F. tularensis*–specific biomarkers was carried out as described for *B. anthracis*. However, the UniProt database was supplemented with additional *Francisella* sequences from unfinished genomes submitted to the NCBI database (Accessions numbers: NZ_ABYY00000000, NZ_ABXZ00000000, NZ_ABRI00000000, NZ_ABAH00000000, NZ_AAYF00000000, NZ_AAYE00000000, NZ_AAYD00000000, NZ_AAUD00000000, NZ_AASP00000000 and NZ_ABSS00000000).

13.13.6 Genetic Validation of Peptide Biomarkers

In silico genetic validation was performed on all *F. tularensis*–specific biomarkers.

For each candidate peptide, the nucleotide sequence of the marker was obtained from the fully sequenced *F. tularensis* subsp. *tularensis* FSC198 genome (Genbank Accession AM286280.1) in FASTA format. For each candidate peptide, nucleotide sequences of the corresponding protein were obtained from publicly available genome sequences for members of the species *F. tularensis* including the subspecies *tularensis*, *holarctica*, *mediasiatica* and *novicida* and closely related species *F. philomiragia* in FASTA format (Genbank Accession: CP000439.1,CP000937.1, CP000803.1, AM233362.1, CP000437.1, CP000915.1, AM286280.1, CP001633, AJ749949.2 and CP000608.1). Nucleotide sequences of the proteins and the peptide biomarker were translated into their corresponding amino acid sequence in Bioedit v7.0.4.1 software. The peptide biomarker was aligned with the protein sequences using the multiple alignment algorithm clustalW as implemented in Bioedit v7.0.4.1 software.

The resulting alignment was manually analyzed in nucleotide and amino acid sequence format to determine if:

1) The peptide biomarkers were unique to the species of interest.
2) The biomarkers were genetically stable (e.g. the DNA sequence of the biomarker contains no silent mutations in the species of interest).

13.14 Summary of Results

Table 13.8 Summary of *F. tularensis*–specific peptides and subspecies-specific peptides. In silico genetic validation of the biomarkers found that all markers were genetically stable.

Locus tag	Peptide biomarker	DNA biomarker	Protein function	Specificity
YP_666502	ATINQLVNNPR	GCAACTATAAATC AGTTGGTGAACA ACCCTCGC	30S ribosomal protein S12	Specific to all subspecies
YP_666528	AQQICQTCN VDPTVK	GCGCAACAAATTT GCCAAACTTGCA ATGTAGATCCAAC TGTCAAA	30S ribosomal protein S13	Specific to all subspecies
	IKDLSEEQ VESLR	ATTAAAGATTTATC AGAAGAACAAGT TGAATCTTTAAGA		
ZP_02274832	EMGGLPDAIV VIDGNK	GAAATGGGCGGT TTACCTGATGCAA TCGTTGTTATTGA TGGCAACAAA	30S ribosomal protein S2	Specific *F. tularensis* subsp. *holarctica*
YP_666530	GNTGATLL ELLESR	GGTAATACCGGTG CTACATTGTTAGA GCTATTAGAAT CAAGA	30S ribosomal protein S4	Specific to all subspecies
	RMYGILEGQFK	CGTATGTATGGTA TTTTAGAAGGTC AATTTAAA		
YP_169371	VAAEILEAVEGR	GTGGCTGCTGAA ATTCTCGAAGCT GTAGAGGGTAGA	30S ribosomal protein S7	Specific to all subspecies
	ALESVSPMVEVK	GCGTTGGAAAGT GTTAGCCCAATG GTGGAAGTTAAG		
YP_666760	ITVNDESAAA AVPEIVK	ATTACAGTTAATGA TGAGTCAGCTGCT GCTGCTGTACCTG AGATTGTTAAA	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	Specific to all subspecies
YP_514377	EEALATLLNIMQ APVTK	GAAGAGGCACTTG CTACATTACTTAAT ATTATGCAAGCAC CAGTTACTAAG	50S ribosomal protein L10	Specific *F. tularensis* subsp. *holarctica*
	GLTVNQMTSLR	GGTTTGACTGTT AATCAAATGAC TTCATTAAGA		
YP_666506	GGIPGSVGGDII VTPAVK	GGTGGTATTCCTG GTTCAGTTGGTGG AGATATTATCGTT ACTCCAGCTGT GAAA	50S ribosomal protein L3	Specific to all subspecies

Table 13.8 (Continued)

Locus tag	Peptide biomarker	DNA biomarker	Protein function	Specificity
YP_513727	LFGSVGIAEV AEAVSNQSGK	CTATTTGGTTCTG TAGGTATAGCTGA AGTTGCTGAAGCT GTTTCTAACCAAT CTGGCAAA	50S ribosomal protein L9	Specific *F. tularensis* subsp. *holarctica*
YP_667260	VSQDIFDQLNK	GTGTCTCAGGAT ATTTTTGATCA GCTTAATAAA	Adenylate kinase	Specific to all subspecies
YP_666636	MNIENYLS ETLAK	ATGAATATAGAAAA TTATTTATCAGAAA CTCTTGCAAAG	Arginyl-tRNA synthetase	Specific to all subspecies
YP_512829	EDDLLFFGAGK	GAGGATGATTTAT TATTCTTTGGTG CTGGTAAG	Aspartyl-tRNA synthetase	Specific *F. tularensis* subsp. *holarctica*
YP_667062	SIPINTLIPIK	AGTATTCCTATCAA TACATTAATACCT ATTAAA	Biotin synthase	Specific *F. tularensis* subsp. *tularensis*
YP_666215	EDEEYSFGLPLK	GAAGATGAGGAAT ACTCTTTTGGTTT ACCGTTAAAA	Chromosomal replication initiator protein DnaA	Specific to all subspecies
YP_514346	SAGGIILTGN AQEKPSQGEVVA VGNGK	TCTGCTGGTGGAAT TATCTTAACTGGTAA TGCTCAAGAGAAAC CTAGCCAAGGTGA GGTTGTTGCTGTT GGTAATGG TAAA	Co-chaperonin GroES	Specific *F. tularensis* subsp. *holarctica*
YP_514179	EIAESEITSEQILR	GAAATTGCTGAGTC AGAAATTACCTCAG AACAAATTTTAAGA	DNA-directed RNA polymerase omega subunit	Specific *F. tularensis* subsp. *holarctica*
ZP_02274323	LIPAGTGLAVR	TTAATACCAGCAG GTACTGGTCTAG CAGTAAGA	DNA-directed RNA polymerase, beta subunit	Specific *F. tularensis* subsp. *holarctica*
YP_666504	FVDEVVGG VVPK	TTTGTTGATGAGG TTGTTGGTGGTGT AGTTCCTAAA	Elongation factor G	Specific to all subspecies
	GVQAVLDGVVR	GGTGTTCAAGCAG TTCTTGATGGTGT GGTTAGA		
	FEPLDEVDE NGEAK	TTTGAGCCTTTGGA TGAAGTTGATGAGA ACGGTGAAGCTAAA		
	MEMIEAAAEA SEELMEK	ATGGAGATGATCG AGGCGGCTGCAG AAGCTTCAGAAGA GCTTATGGAGAAA		
	YLEGGELSEDEI HQGLR	TATCTTGAGGGTG GTGAACTTTCTGA AGATGAGATTCAT CAAGGTCTGCGT		

(*Continued*)

Table 13.8 (Continued)

Locus tag	Peptide biomarker	DNA biomarker	Protein function	Specificity
YP_666268	GEVATDLTSPIEK	GGTGAAGTGGCAAC TGATTTAACTTCAC CTATCGAAAAA	F0F1 ATP synthase subunit alpha	Specific to all subspecies
YP_667055	IAVNTFVDLGK	ATTGCAGTAAATAC ATTTGTAGATTTAG GCAAA	Glutamate-1-semialdehyde-2, 1-aminomutase	Specific to all subspecies
	ISQPGFFDELGAK	ATCTCACAACCAGG GTTCTTTGATGAGC TTGGAGCTAAA		
YP_666536	YSDCINTSIQMK	TACTCAGATTGTAT CAATACATCAATTC AGATGAAG	Heat shock protein 90	Specific *F. tularensis* subsp. *tularensis*
YP_513194	ATVYTAYNNNP QGSVR	GCAACTGTATATAC AGCATACAATAATA ACCCACAAGGAAG TGTAAGA	Lipoprotein	Specific *F. tularensis* subsp. *holarctica*
YP_667413	AICAAIDNAIK	GCAATCTGTGCAG CTATTGACAATG CAATCAAA	Phospho-glyceromutase	Specific *F. tularensis* subsp. tularensis
YP_666911	SDFMIELDLQK	TCTGACTTTATGAT TGAGCTTGATCTA CAGAAG	Preprotein translocase subunit SecA	Specific to all subspecies
	TPLIISGASDDR	ACACCACTTATCAT ATCAGGTGCCTCA GATGATAGA		Specific *F. tularensis* subsp. *tularensis*
YP_666257	LASATITEVDLSK	CTAGCTAGTGCAA CTATTACAGAAGT GGATCTTTCTAAG	Ribosome-binding factor A	Specific to all subspecies
YP_513533	EREVISEILAEK	GAAAGAGAAGTAA TTTCTGAGATTTT AGCTGAAAAA	Shikimate kinase I	Specific *F. tularensis* subsp. *holarctica*
YP_763652	ISGIDALEIAEK	ATATCTGGGATAGA TGCTCTAGAAATA GCTGAGAAA	tRNA modification GTPase TrmE	*Specific F. tularensis subsp. holarctica*

13.15 *Clostridium botulinum*

The genus *Clostridium* comprises a heterogeneous collection of gram-positive, anaerobic, rod-shaped bacteria that share the ability to form endospores. *C. botulinum*, the causative agent of the paralytic illness botulism, is defined by its ability to produce a potent neurotoxin, of which seven toxin types exist (A–G). Defining *C. botulinum* on the basis of this single phenotypic trait has resulted in phenotypic and genotypic heterogeneity within the species. Strains of *C. botulinum* can be subdivided into four distinct phenotypic groups, as also confirmed by phylogenetic analysis of the 16S rRNA gene. In addition to

this heterogeneity, a number of the other *Clostridium* species share a close genetic relationship with each of these groups, and strains of species other than *C. botulinum* have been found to express the neurotoxin (e.g. *C. barati*, *C. butyricum*) (Al-Shahib *et al.*, 2010). The grouping of *C. botulinum* strains is summarized below.

Properties	Group I	Group II	Group III	Group IV
Toxin types	A,B,F	B,E,F	C,D	G
Disease host	Human	Human	Animal/fish	Unknown
Phenotype	Proteolytic	Non- proteolytic	Non- proteolytic	Proteolytic
Phylogenetically related species	*C. sporogenes* *C. tetani*	*C. butyricum* *C. barati* *C. beijerinkii* *C. perfringens*	*C. novyi* *C. haemolyticum*	*C. subterminale*

13.16 Method

13.16.1 Strain Panel

Eight *C. botulinum* strains belonging to group 1 (Toxin A: NCTC 13319, NCTC 2916, NCTC 7272 and A HO6506; Toxin B: NCTC 751, NCTC 3807 and NCTC7273; Toxin F: NCTC 10281), one *C. botulinum* strain belonging to group 2 (Toxin E: NCTC 82660) and five *C. botulinum* strains belonging to group 3(Toxin C: NCTC 3723, MPRL 3493, NCTC 8549 and MPRL 4240 and Toxin D: NCTC 8265) were used. Thirteen non–*C. botulinum* strain (*C. sporogenes*: NCTC 275, NCTC 533 and NCTC 534; *C. difficile*: NCTC 630; *C. butyricum:* NCTC 7423; *C. perfringens*: NCTC 8238 and NCTC 3110; *C. tetani*: NCTC 5405; *C. beijerinkii*: NCTC 11920; *C. hylemonae*: DSM 15053; *C. nexile*: DSM 1787; *C. subterminale*: DSM 6970; *C. ramosum*: DSM 1402) were included.

13.16.2 Procedure for Whole Cell Protein Extraction

C. botulinum strains were cultured on *C. botulinum* isolation agar under anaerobic conditions at 37 °C for 48 h, whereas non–*C. botulinum* strains were grown under the same conditions but on fastidious anaerobic agar. From each culture, 1 × 10 µl of cells was inoculated into a 20 ml of Trypticase-Peptone-Glucose-Yeast Extract Broth. The turbidity of the starting culture was adjusted to give an OD_{600} of approximately 0.3 before incubation under anaerobic conditions at 37 °C for 24 h. After ~24 h, when the OD_{600} was measured to be ~ 2–2.5, 20 ml of culture was centrifuged at 10,000 g for 10 min at 4 °C to harvest the cells. Broth supernatant was removed, and the cell pellets were washed with 1 ml 1× TE buffer containing a 1× protease inhibitor cocktail. The cells were centrifuged at 10,000 g for 10 min at 4 °C to re-pellet the cells and the supernatant removed. The wash step was repeated twice. The cell pellets were treated with 1 ml of 1× TE buffer containing 6 µg/µl lysozyme and a 1× protease inhibitor cocktail and incubated for 1 h at 37 °C on a heating block. Cells were collected by centrifugation at 8000 g for 10 min at 4 °C and the supernatant removed. The cell pellets were washed with 1 ml 1× TE buffer containing a 1× protease inhibitor cocktail. The cells were centrifuged at

10,000 g for 10 min at 4 °C to re-pellet the cells and the supernatant removed. The wash step was repeated twice. The pellet was re-suspended in 100 µl of a solubilization cocktail (30 mM Tris-Cl pH 8.5, 7 M Urea, 2 M Thiorurea, 4% CHAPS and 70 mM DTT), and solubilization allowed to occur for 30 min at room temperature. The suspension was clarified by centrifugation at 21,000 g for 30 min at 21 °C. Extracts were filtered through 0.2 µm anopore vectra spin filters by centrifugation at 8000 g and 4 °C, and the filtrate was transferred to fresh 1.5 ml tubes.

Viability studies were performed on all *C. botulinum* extracts. 15 µl of protein extract was re-suspended in 15 ml nutrient broth and on nutrient agar and incubated at 37 °C for 7 days. 15 µl of broth was plated onto nutrient agar and incubated at 37 °C for 48 h. All plates and broths were examined for growth that resembles that of *C. botulinum*. Only samples that show negative growth were analyzed by 1D PAGE.

Three biological replicates were prepared for 1D PAGE analysis.

13.16.3 One-Dimensional SDS-PAGE and In-Gel Digestion of Bacterial Proteins

The procedure was as described for *B. anthracis* with the following deviation: 10 µg of protein extract was run on a 4%–12% Bis-Tris gel, and each lane was cut into 12 bands that were in-gel digested with trypsin as described.

13.16.4 1-D Nanoflow LC-MS/MS, Data-Dependent and Targeted MS Analysis

Digested peptides were separated online by RP HPLC on an Ultimate 3000 (Dionex, UK). All LC-MS/MS experiments were performed on a Thermo Finnigan LTQ Orbitrap equipped with a nanoelectrospray ion source and nano-LC pico-tip (20 um ID 10 um tip diameter).

Separations were performed on a nano analytical C18 column (75 µm id × 15 cm, 3 µm) using a 45-min linear gradient of 5% to 45% solvent B (90% CH_3CN/0.1% formic acid) versus solvent A (2% CH_3CN/0.1% formic acid), then to 90% B for an additional 5 min. Data was acquired in data-dependent mode using the Thermo Finnigan Xcalibur software (version 2.0.6). The precursor ion scan MS spectra (*m/z* 440–2000) were acquired in the Orbitrap with a resolution R = 60000 at *m/z* 400 with the number of accumulated ions being 5×10^5. The six most intense ions detected in the preceding survey scan were dynamically selected and subjected to CID in the linear ion trap to generate MS/MS spectra.

13.16.5 Bioinformatic Workflow for Biomarker Detection

MS data was generated in the form of.RAW files which contain all the spectra detected from the LC-MS/MS analysis for each sample were used. Each replicate was examined separately. The database search algorithms Mascot and Sequest Bioworks version 3.3 were used. The same parameters were applied for both algorithms and used to search against the NCBInr database. In silico trypsin digestion of protein sequences in the database was implemented in Mascot and Sequest. Search parameters were set for carbamidomethylation of cystine, oxidation of methionine, missed cleavage sites = 2 and a peptide mass tolerance ± 10 ppm. When using Mascot, sequences that had the specific Mascot score >20 were reported. A decoy database was used to calculate the FDR in both Mascot and Sequest. The decoy and target databases were concatenated, and the FDR was then calculated on the basis of the true and false positive identifications. The output from each algorithm was analyzed, scripts were used to remove duplicate

peptides from each replicate, and the data converted to the FASTA format. For each replicate, BLASTp in WU-BLAST 2.0 was adjusted for short sequences and used to identify specific markers. For each BLASTp output, if the description states *C. botulinum*, the script looks for 100% sequence identity to the query sequence. If this is satisfied, then the length of the query sequence is determined, and an exact match in the alignment is searched. A search for any match in output is then made, and if there are no conflicts, the results indicate the presence of a biomarker. Peptides identified from each replicate were compared to find the peptides conserved among all three replicates. To ensure that the peptides are unique to *C. botulinum* and not any other species, a comparison of peptides was made between *C. botulinum* and non–C*. botulinum* species. The peptides that were shared with these strains were eliminated, and a refined list of markers was produced. A consensus of toxin-specific and group-specific peptides was produced by comparing the biomarkers identified by both search algorithms.

13.16.6 Procedure for DNA Extraction

Additional strains were included in the DNA panel (*C. barati*, *C. haemolyticum, C. subterminale* and *C. novyi*). For these and the non–C*. botulinum* strains listed in the strain panel, DNA extracts were prepared using the ZYMO Bacterial/Fungal DNA extraction kit following the manufacturer's instructions.

For *C. botulinum* strains in the strain panel, the following steps were performed. Broth cultures were prepared as described for the whole cell protein extraction. 1 ml of the broth culture was taken and transferred to a 1.5 ml microcentrifuge tube. The tube was centrifuged at 8000 g for 10 min and the supernatant removed. DNA extraction was performed on the cell pellets using the Maxwell® 16 system. Viability studies were performed on the DNA extract as described for the whole cell protein extract.

13.16.7 Genetic Validation of Peptide Biomarkers

Candidate biomarkers for various *C. botulinum* strains were split into three groups depending on which phenotypic/phylogenetic cluster they were found to correspond to on the basis of the BLASTp output:

Group 1: Proteolytic *C. botulium* toxin types A, B and F
Group 2: Non-proteolytic *C. botulium* toxin types B, E and F
Group 3: *C. botulium* toxin types C and D

Due to the broad genetic diversity within the genus *Clostridia* and the diversity between *C. botulinum* strains derived from different phylogenetic clusters, strains in the test panel were divided into different PCR groups for genetic validation studies. These PCR groups were based on the phylogenetic cluster the *C. botulinum* strains in the test panel were derived from and the non- *C. botulinum* strains that were most closely related to them (see Table 13.9).

These PCR groups were as follows:

PCR group 1: Group 1 *C. botulinum* strains, *C. sporogenes* and *C. tetani*
PCR group 2: Group 2 *C. botulinum* strains, *C. perfringens*, *C. butyricum, C. barati* and *C. beijerinkii*
PCR group 3: Group 3 *C. botulinum* strains, *C. novyi* and *C. haemolyticum*

Table 13.9 Primer sequences used for group-specific PCR amplification and DNA sequencing of each target biomarker region.

Target biomarker	Approximate amplicon size (bp)	Primer sequences
PCR group 1		
YP_001255947	490	YP_001255947_F 5′- ACT WCC ART AGC TAT MCC WAA TGG -3′ YP_001255947_R 5′- CCW GTY TTR CCT TCC TTA CG -3′
YP_002805802	150	YP_002805802_F 5′- CTA CAG TAA GAC CTG GTG TTA TGG -3′ YP_002805802_R 5′- CCA CCA GAT ACR ATA AAG TCT GC -3′
YP_002802667	140	YP_002802667_F 5′- CAG CTA TAG GAG TAG GAC CTG G -3′ YP_002802667_R 5′- CAA TTA CTG ATT GYT CYG AKG C -3′
YP_002863717	680	YP_002863717_F 5′- AGA TGA YGC WGC AGG ATT AGC -3′ YP_002863717_R 5′- GTT WGC TTG WGC AAG CAT WGC -3′
PCR group 2		
YP_001922388	570	YP_001922388_F 5′- ACA AGC TGG TWT AGC TGG AGC -3′ YP_001922388_R 5′- TTC CAC CTA CTG CTG TRA ATT CTC C -3′
YP_001307555	750	YP_001307555_F 5′- AGA VGT AGT TAT WGY AAG YGC -3′ YP_001307555_R 5′- ATW CCT GAW GCA TTA CCW GC -3′
YP_001922189	220	YP_001922189_F 5′- ATW GCT TAY GAA CCA ATC TGG -3′ YP_001922189_R 5′- AGC TCC ACC AAC TAA WGC -3′
PCR group 3		
ZP_02620662	520	ZP_02620662_F 5′- GGA TGT MGK GAA AAR CCT AGA GG -3′ ZP_02620662_R 5′- CWA CAT TYC CAC AWC YAA ATA TTC C -3′
YP_878395	350	YP_878395_F 5′- AGA TTT CAA AGG CGT TTG GGT ATT CG -3′ YP_878395_R 5′- CCA GTK TGT AAT CTW GCW GC -3′
ZP_04863386	470	ZP_04863386_F 5′- CAG GAC TTG GWG CTA TWG G -3′ ZP_04863386_R 5′- CTC ATA GCT GCA TAA ACW CC -3′
ZP_04863198	560	ZP_04863198_F 5′- ATG TCT TAG CCA AAA ATG GAA GG -3′ ZP_04863198_R 5′- AGC ACA TCC GTC AAA TTS TTG C -3′

For each candidate peptide, the nucleotide sequence of a group 1 marker was obtained from the fully sequenced *C. botulinum* toxin type A. (ATCC 3502) genome (Genbank Accession AM412317.1) in FASTA format. For group 2, the fully sequenced *C. botulinum* E3 str. Alaska E43 and *C. botulinum* B str. Eklund 17B genomes (Genbank Accession CP001078.1 and CP001056.1, respectively) were used. For group 3, the fully sequenced genomes *C. botulinum* C str. Eklund and *C. botulinum* D str. 1873 (Genbank Accession ABDQ00000000 and ACSJ00000000, respectively) were used. For each candidate peptide, nucleotide sequences of the corresponding protein were obtained from publicly available genome sequences for members of each PCR group (described above) in FASTA format (Genbank Accession: Group 1: CP000726.1, AM412317.1, CP000727.1, CP001581.1, CP000962.1, CP000939.1, CP001083.1, CP002011 and CP000728.1; Group 2: CP001056.1 and CP001078.1; Group 3: ABDQ00000000, ACSJ00000000, ACSC00000000, ABDO00000000, CP000721.1, AM180355.1, CP000382.1, CP000246.1, CP000312.1, BA000016.3, AE015927.1, ABDT00000000, ACOM00000000 and ABKW-00000000). Nucleotide sequences were translated into their corresponding amino acid sequence in Bioedit v7.0.4.1 software. The peptide biomarker was aligned with the protein sequences using the multiple alignment algorithm clustalW as implemented in Bioedit v7.0.4.1 software. The aligned protein sequences were back-translated to the corresponding nucleotide sequence and the consensus sequence of the alignment determined. PCR primers were designed as described for *B. anthracis*. PCR and direct sequencing were performed as described for *B. anthracis* using the primers listed in Tables 13.5, 13.6). Sequences trace files were analyzed and alignments for each biomarker produced as described for *B. anthracis*.

The resulting alignments were manually analyzed in nucleotide and amino acid sequence format to determine if:

1) The peptide biomarkers were unique to a specific toxin type or phylogenetic group.
2) The biomarkers were genetically stable (e.g. the DNA sequence of the biomarker contains no silent mutations in the species of interest).

13.17 Summary of Results

Group 1: 6 group-specific peptide biomarkers (1 biomarker specific to toxin type A)
Group 2: 153 group-specific peptide biomarkers (53 biomarkers specific to toxin type E)
Group 3: 3 group-specific peptide biomarkers (43 specific to toxin type D and 1 specific to toxin type C) (Table 13.10).

13.18 *Burkholderia pseudomallei* and *B. mallei*

The genus *Burkholderia* refers to a group of virtually ubiquitous gram-negative, motile, obligately aerobic rod-shaped bacteria. *B. pseudomallei*, an environmental saprophyte, is endemic to southeast Asia and northern Australia and the causative agent of the fatal

Table 13.10 A summary of *C. botulinum* toxin-specific and group-specific markers that were genetically validated by PCR and direct sequencing (* = no suitable primer region found, [$] = experimentally validated stable marker, [†] = genetically stable by in silico analysis, [◊] = silent mutations present, [‡] = PCR amplification/sequencing not achieved for all relevant *C. botulinum* strains).

Locus tag	Peptide biomarker	DNA biomarker	Protein function	Specificity
Group 1				
YP_001255947[$]	ALVNNMVTGVNEGYVK	GCTTTAGTAAACAACATGGTAACA GGAGTTAATGAAGGATACGTTAAA	Ribosomal protein L6	Specific to toxin type A
YP_002805802[†‡]	VALSEDEIR	GTAGCATTAAGTGAAGATG AGATTAGA	Electron transfer flavoprotein, alpha subunit/FixB family protein	Specific to group 1
YP_002802667[†‡]	MAVSSVVLSK	ATGGCTGTAAGCTCAGTGGTA CTTTCTAAA	Aldehyde-alcohol dehydrogenase	Specific to group1
YP_002863717 [◊]	LEHTIANLDNSAENL QAAESR	TTAGAACACACAATAGCAAACT TAGATAATTCAGCTGAAAACTTA CAAGCAGCGGAATCAAGA	Flagellin	Specific to group 1
Group 2				
YP_001922388[$]	LTADSITTLDGIK	TTAACAGCGGATTCAATAACAAC ACTTGATGGAATAAAA	Cell wall binding repeat protein	Specific to toxin type E
YP_001307555[$]	FKDEIVPVMIPQR	TTTAAAGATGAAATAGTTCCAGT TATGATTCCACAAAGA	acetyl-CoA acetyltransferase	Specific to group 2
YP_001922189[$]	QANETILAIR	CAAGCTAATGAAACAATATTA GCAATCAGA	Triosephosphate isomerase	Specific to group 2
Group 3				
ZP_02620662[‡]	DIVILGSGDVALLMAR	GATATAGTTATTTTAGGTTCGGGTG ATGTGGCACTGCTTATGGCACGA	Sarcosine oxidase alpha subunit	Specific to toxin type C
ZP_04863386[‡]	LVIGTLQGR	TTAGTTATAGGAACATTAC AAGGAAGA	Purine nucleoside phosphorylase I, inosine and guanosine-specific	Specific to toxin type D
ZP_04863198[‡◊]	IDVPEGTDPLYR	ATTGACGTTCCAGAAGGAACAGA TCCATTATATAGA	(R)-2-hydroxyglutaryl-CoA dehydratase subunit beta	Specific to group 3
YP_878395[‡◊]	LGVELTAVLLGNK	CTAGGAGTAGAATTAACTGCAGT TTTACTTGGAAATAAG	Electron transfer flavoprotein	Specific to group 3

human and animal disease melioidosis. Unlike *B. pseudomallei*, the closely related species *B. mallei* is not an environmental pathogen, and its ability to survive in the environment is limited. *B. mallei* is the causative agent of glanders, a disease most commonly affecting equids. *B. pseudomallei* and *B. mallei* are genetically very similar, sharing more than 99% 16S rRNA sequence identity and can be difficult to distinguish. Both species are characterized by the United States Centres for Disease Control as a category B biowarfare agent. These organisms are less familiar to medical and laboratory personnel compared to other biothreat agents such as *Yersinia pestis*, *Bacillus anthracis* and *Francisella tularensis*, and both organisms have many features that make them attractive candidates for bioterrorism such as their ability to enter via the inhalation route, severity of disease and the fact that there is no vaccine available. *B. thailandensis* is a non-virulent closely related species to *B. pseudomallei*, sharing similar biochemical profiles.

13.19 Method

13.19.1 Strain Panel

Four *B. pseudomallei* strains NCTC 4845, 1688 and wild-type 204, one *B. mallei* strain NCTC 12938, one *B. thailandensis* strain DSM 13276 and representatives of the *B. cepacia* complex (*B. cepacia* LMG 2161, *B. multivorans* LMG 13010, *B. cenocepacia* LMG 18863, LMG 16604, LMG 212161, *B. stabilis* LMG 14294, *B. vietnamiensis* LMG 10929, *B. dolosa* LMG 18941, *B. ambifaria* LMG 11351, *B. anthina* LMG 20983, *B. pyrrocinia* LMG 14191 and *B. gladiola* NCTC 12378) were used.

13.19.2 Procedure for Whole Cell Protein Extraction

Isolates were grown on nutrient agar plates for 48 h at 30 °C. 2 × 10 µl loops of cells were re-suspended in 100 µl of distilled water. Samples were boiled for 1 h in a water bath and cooled to room temperature.

Viability was determined for boiled extracts of *B. pseudomallei* and *B. mallei* by taking 10 µl of the extract, plating onto nutrient agar, and inoculating 20 ml of nutrient broth with 10 µl of boiled extract. Broths and agar were incubated at 30 °C for 48 h and observed for growth. 10 µl of the incubated broths was plated onto nutrient agar and incubated at 30 °C for 48 h. Samples were stored at −20 °C until the viability test was confirmed as negative.

Cells were harvested from boiled extracts by centrifugation at 8000 g for 10 min at 4 °C. Supernatant was removed, and the cell pellets were washed with 500 µl 1× TE containing 1× protease inhibitor cocktail. Cells were harvested by centrifugation at 8000 g for 10 min at 4 °C. The cell pellets were re-suspended in 500 µl 1× TE containing 1× protease inhibitor cocktail and 6 mg/ml lysozyme and incubated at 37 °C for 1 h. Cells were harvested by centrifugation at 8000 g for 10 min at 4 °C and the supernatant removed. The pelleted cells were washed using 500 µl 1× TE containing 1× protease inhibitor cocktail and the cells collected by centrifugation at 8000 g for 10 min at 4 °C. Supernatant was removed, and the wash step was repeated twice. 500 µl of lysis solution

(7 M urea, 2 M thiourea, 4% CHAPS, 20 mM Tris, 70 mM DTT) was used to re-suspend the cell pellet, and glass beads were added to approximately a third of the volume. The cells were lysed for 60 s using the Fast Prep system. Unbroken cells were collected by centrifugation at 21,000 g for 30 min at 21 °C, and the supernatant was collected and transferred to fresh 1.5 ml tubes.

13.19.3 One-Dimensional SDS-PAGE and In-Gel Digestion of Bacterial Proteins

The procedure was as described for *C. botulinum*.

13.19.4 One-Dimensional Nanoflow LC-MS/MS, Data-Dependent and Targeted MS Analysis

The procedure was the same as that described for *C. botulinum*.

13.19.5 Bioinformatic Workflow for Biomarker Detection

The procedure was the same as that described for *C. botulinum*.

13.19.6 Procedure for DNA Extraction

DNA extracts were prepared using the ZYMO Bacterial/Fungal DNA extraction kit following the manufacturer's instructions. Strains belonging to *B. mallei* and *B. pseudomallei* were boiled, and viability studies were performed as described for the whole cell protein extraction before carrying out a DNA extraction on the cell pellet.

13.19.7 Genetic Validation of Peptide Biomarkers

The procedure was as described for *B. anthracis*, with the following exceptions:

1) The fully sequenced genome *B. pseudomallei* 668 and *B. mallei* ATCC 23344 (Genbank Accession: CP000570.1and CP000011.1, respectively) was used for obtaining the nucleotide sequence of the peptide biomarker for *B. mallei* and *B. pseudomallei*, respectively.
2) Nucleotide sequences of the corresponding protein sequences were taken from fully sequenced genomes belonging to *B. ambifaria* (Genbank Accession: CP000440.1 and CP001025.1), *B. cenocepacia* (CP000378.1, CP000458.1, AM747720.1 and CP000958.1), *B. mallei* (Genbank Accession: CP000011.1, CP000545.1, CP000547.1 and CP000525.1), *B. multivorans* (Genbank Accession: CP000869.1 and AP009387.1), *B. pseudomallei* (Genbank Accession: CP000572.1, CP000124.1, CP000570.1 and BX571965.1), *B. thailandensis* (Genbank Accession: CP000085.1) and *B. vietnamensis* (Genbank Accession: CP000616.1) (See 'Summary of Results', Table 13.11 and Table 13.12).
3) The procedure for PCR and direct sequencing was performed as described for *B. anthracis* using the primers listed in Table 13.7 and DNA from strains belonging to the *Burkholderia* strain panel.

Table 13.11 Primer sequences used for PCR amplification and DNA sequencing of each target biomarker region.

B. mallei			
Locus tag	**Approximate amplicon size (bp)**	**Primer sequences**	
YP_103702	310	YP_103702_F	5′- GAA CAT CCC GAC CTA CCT GAT -3′
		YP_103702_R	5′- GTS GCG AGC TTC AGC TTG AA -3′
YP_104460	330	YP_104460_F	5′- GCG AAT ACG ARC ACA TYT CG -3′
		YP_104460_R	5′- CTT CTG GTT GAT GAT CGY GTC G -3′
YP_104003	400	YP_104003_F	5′- ATG AAG ATT GCG ATT GCY GG -3′
		YP_104003_R	5′- ARC TTC AKC GTG ACG TTS ACG -3′
B. pseudomallei			
YP_001066449	160	YP_001066449F	5′- GGR CAT GCT GTT CTT CAC C -3′
		YP_001066449R	5′- TCG TSG CGT ACY GRT ACA GG -3′
YP_001075630	170	YP_001075630F	5′- GTC GCT CGA CAA GGT GAT GG -3′
		YP_001075630R	5′- TTC AGG TAA TAS AGA TAC TCS GC -3′
YP_001061921	400	YP_001061921F	5′- GCS ATC AAC GGC TAY GCG ATG G -3′
		YP_001061921R	5′- AAC AGR TCG ACG AAG CGY TCG -3′
YP_001061919	550	YP_001061919F	5′- ATG ATG GGY GCG AAG AAY CAY GC -3′
		YP_001061919R	5′- ATC GGR ATG TTG ATS CCR ACC TGG -3′
YP_001074630	210	YP_001074630F	5′- TAC GAR ATC TAC ATG GTB KCS GAC G -3′
		YP_001074630R	5′- TTG TGV ASC ATC GTG TAC G -3′
YP_001060796	410	YP_001060796F	5′- ATC GCG RSG AAY CTG CTG AAG G -3′
		YP_001060796R	5′- TCY TCT GRC CGA TCG CRT CG -3′

13.20 Summary of Results

Table 13.12 Summary of *B. mallei*– and *B. pseudomallei*–specific peptides subjected to PCR and direct sequencing (* = no suitable primer region found, § = genetically stable marker, † = genetically stable by in silico analysis, ‡ = amplification not achieved for all *B. pseudomallei* strains). Markers were not found to contain any silent mutations by in silico analysis.

B. mallei			
Locus tag	**Peptide biomarker**	**DNA biomarker**	**Protein function**
YP_103702§	SIINDPIVNIAR	TCGATCATCAACGATC CGATCGTCAACATCG CGCGC	Bifunctional glucokinase/ RpiR family transcriptional regulator
YP_104460§	VLDALGNPIDGK	GTGCTCGACGCGCT CGGCAATCCGATCG ACGGCAAG	ATP synthase subunit alpha
YP_104003§	TGSSQLGQDA GAFLGK	ACGGGCTCGTCGCA GCTCGGCCAGGACG CGGGCGCGTTCCTC GGCAAG	Dihydrodipicolinate reductase
B. pseudomallei			
YP_001075502*	AAAAQPVLTVR	GCGGCGGCGGCGCA GCCGGTGCTGACCG TTCGC	Hypothetical protein
YP_001064058*	AFADAILQAAHL	GCGTTCGCCGACGCG ATCCTGCAGGCCGCG CACCTG	Alpha/beta hydrolase family protein
YP_001066449§	DNLGQAVVG GIVTGR	GACAATCTGGGGCAG GCCGTGGTGGGCGG GATCGTCACCGGTCGG	Hypothetical protein
YP_001075630‡	IDLGLAPTPPR	ATCGACCTCGGGCTC GCGCCCACGCCGC CGCGC	Hypothetical protein
YP_001075510*	IDPVAIEAAIGR	ATCGATCCGGTCGC GATCGAAGCCGCGAT CGGCCGT	Putative nucleoside 2-deoxyribosyltransferase
YP_001061921‡	IGLVEEVVDAGR	ATCGGCCTCGTCGAG GAGGTGGTCGACGC CGGCCGC	Enoyl-CoA hydratase
YP_001061919‡	VNAGAEAGTD VGPLVSR	GTCAACGCGGGCGCG GAAGCGGGCACCGAC GTCGGCCCCCTG GTGTCGCGC	Methylmalonate-semialdehyde dehydrogenase
	VLGLIETGEQEGAR	GTGCTCGGCCTCATC GAGACCGGCGAACA GGAAGGCGCGAGG	
YP_106976*	VLVVIDTAYIR	GTGCTGGTGGTTATC GACACGGCTTACAT CAGG	Hypothetical protein

Table 13.12 (Continued)

Locus tag	Peptide biomarker	DNA biomarker	Protein function
YP_001074630[§]	DTYDAVMTLVK	GACACGTACGACGCG GTGATGACGCTCG TGAAG	Hydrolase, isochorismatase family
YP_001060796[‡]	DVFDAALLEQAPR	GACGTGTTCGACGCG GCCCTCCTCGAGCA AGCGCCGCGC	Dehydrogenase
YP_001075559*	LTPIPLLLVHGPNA	CTCACGCCGATTCCG TTGCTGCTCGTGCAT GGGCCGAACGCC	Universal stress family protein
	VLFATDGSPIAAR	GTGCTGTTCGCGACC GACGGCAGCCCGATC GCCGCGCGC	
	WVEGSVSEPLLAR	TGGGTCGAGGGCAG CGTGTCCGAGCCGCT GCTCGCGCGG	
YP_001062606*	VDVTVETAIVDLAK	GTCGACGTGACCGTC GAAACGGCGATCGTC GATCTCGCGAAG	Universal stress family protein

13.21 Biomarker Detection Sensitivity and Quantification

Using *B. anthracis* as a model, peptide biomarkers were verified and their absolute quantifications were obtained using their corresponding stable isotope-labelled peptides as internal standards. It was expected that the retention time and fragmentation pattern of the native peptide would be identical to that of the corresponding stable isotope-labelled standard peptide producing highly specific and highly sensitive analysis.

13.22 Method

Four peptide markers from the panel of *Bacillus anthracis*–specific peptide biomarkers were selected for further verification and quantification and are shown below:

- SADLVQGL[C13N15]VDDAVEK
  - SADLVQGLVDDAVEK
- MDVDML[C13N15]SNR
  - MDVDMLSNR
- AIGAEL[C13N15]DQLVK
  - AIGAELDQLVK
- LVSIGEL[C13N15]QPDGNR
  - LVSIGELQPDGNR

Five *B. anthracis* strains (NCTC 109, NCTC 2620, NCTC 5444, NCTC 7752 and NCTC 7753) and four reference strains representing the *B. cereus* group of species (*B. cereus* NCTC 11143, *B. thuringiensis* DSM 2046, *B. weihenstephanensis* DSM 11821 and *B. mycoides* NCTC 7586) were used for assessing the sensitivity and specificity of biomarker detection.

13.22.1 Preparation of Stable Isotope-Labelled Peptides

Stable isotope-labelled internal standard peptides were commercially synthesized and supplied >95% pure in 5 × 1 nmol quantities and kept at −80 °C until use. A 5 pmol/µl stock solution was prepared by dissolving in 20 µl of 10% (v/v) aqueous formic acid to which 180 µl of 0.1% (v/v) aqueous formic acid was then added and gently swirled to mix. A working solution containing 1 pmol/µl of each internal standard peptide was prepared by combining the stock solutions and diluted using 0.1% aqueous formic acid.

13.22.2 Preparation of Samples for Absolute Quantification

Protein concentration for extracts was determined by the Bradford assay using BSA as a standard. Protein extracts (10 µg) were separated on 4%–12% SDS/PAGE gels and stained with Coomassie blue. Each gel lane was excised into 12 pieces and destained with 50% (v/v) methanol for 3 × 20 min, dehydrated using 100% acetonitrile for 10 min and dried for 5 min. Gel pieces were rehydrated and proteins reduced by addition of 10 mM DTT and subsequently alkylated by addition of 55 mM iodoacetamide. In-gel digestion of proteins was performed with 50 µl (10 ng µl^{-1}) of porcine trypsin containing isotope-labelled internal standards (60 fmol µl^{-1}) overnight at 37 °C. Tryptic peptides were extracted with aqueous TFA (0.1%, v/v, 50 µl) for 1 h with gentle agitation. The TFA (0.1%, 50 µl) extraction step was repeated. Extracts from the two TFA (0.1%) extractions were combined and stored at −80 °C until further analysis.

13.22.3 One-Dimensional Nanoflow LC-MS/MS, Data-Dependent MS Analysis

NanoLC-MS/MS analysis was performed in triplicate using the method described previously. However, the following deviations were used:

1) Analytical separation was achieved on a RP nano-column C18, 75 um i.d. × 15 cm at a flow rate of 0.3 µl min^{-1} as described previously, but with a gradient of 10%–40% acetonitrile in 45 min, increased to 90% acetonitrile and held for 7 min before it returned to the initial condition of 10% acetonitrile.
2) The precursor ion scan was acquired (m/z 440–2000).

13.22.4 Data Analysis

The native peptide and its corresponding isotope-labelled internal standard were observed in the extracted ion chromatograms as peak doublets with the same retention (Figure 13.2) time but separated by a mass of $\Delta m = 3.5$ (doubly charged product ions;

Table 13.13 Selected peptides, their corresponding stable isotope-labelled standard peptides and retention time.

Peptide	Monoisotopic mass of the peptide	Product ion $[M+2H]^{2+}$ (m/z)	Retention Time (min)
MDVDMLSNR	1080.5	540.8	16.8
MDVDML*SNR	1087.5	544.2	
AIGAELDQLVK	1156.7	578.8	22.2
AIGAEL*DQLVK	1163.7	582.3	
LVSIGELQPDGNR	1397.7	699.4	18.4
LVSIGEL*QPDGNR	1404.7	702.9	
SADLVQGLVDDAVEK	1558.8	779.9	27.4
SADLVQGL*VDDAVEK	1565.8	783.4	

* ^{13}C and ^{15}N labelled amino acid leucine (L*). Mass increase 7 Da (*e.g.,* six ^{13}C plus one ^{15}N) in comparison with standard amino acid.

Table 13.13). Quantification was done manually from nanoLC-MS/MS data after integrating the pair of peaks for the isotope-labelled internal standard and its native counterpart in an extracted ion chromatogram (see Figure 13.5). Baselines were manually and carefully adjusted to obtain the most accurate area under the peak. The extracted ion chromatogram can be obtained by extraction of a specific mass. Native peptide levels were subsequently quantified by multiplying the ratios of the areas under the curve for each respective peak doublet by the known amounts of isotope-labelled peptide standards added (Table 13.14).

13.23 Summary of Results

Sensitivity testing was also performed for *C. botulinum*. Two peptide biomarkers specific to Group I *Clostridium botulinum* strains (Type A, B and F) were used:

- EAEYIF[C13N15]GNFGK
  - EAEYIFGNFGK
- YANIADYL[C13N15]SLGGK
  - YANIADYLSLGGK

In addition, three *Clostridium botulinum* strains (toxin type A: NCTC 13319; toxin type B: NCTC 3807; toxin type C NCTC 2510) and three non-*botulinum Clostridium* strains (*C. sporogenes*: NCTC 275; *C. difficile*: NCTC 630; *C. tetani*: NCTC 5405) were also included (Table 13.15).

Samples were processed and analyzed as previously described for *B. anthracis*. However, the native peptide and its corresponding isotope-labelled internal standard were observed in the extracted ion chromatograms as a peak doublets with the same retention time but separated by a mass of $\Delta m = 3.5$ for the peptide pair YANIADYL[C13N15]SLGGK and YANIADYLSLGGK and $\Delta m = 5.0$ for the peptide pair EAEYIF[C13N15]GNFGK and EAEYIFGNFGK.

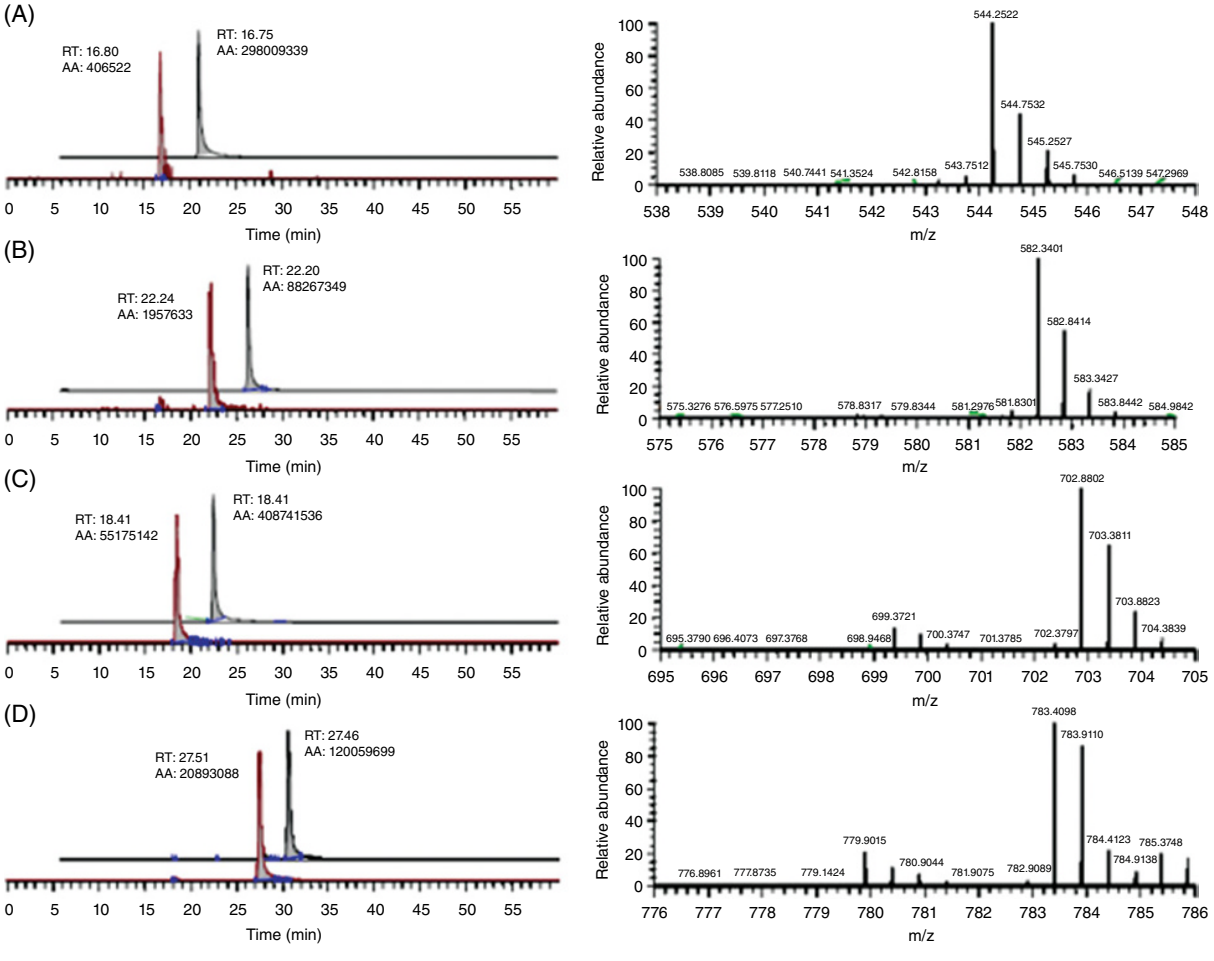


Figure 13.5 Four extracted ion chromatograms and MS/MS product ions of selected native peptides (in red) and the corresponding stable isotope-labelled standard peptides (in black) detected in *B. anthracis* protein extract. (A) Peptides MDVDMLSNR and MDVDML*SNR, (B) peptides AIGAELDQLVK and AIGAEL*DQLVK, (C) peptides LVSIGELQPDGNR and LVSIGEL*QPDGNR, and (D) peptides SADLVQGLVDDAVEK and SADLVQGL*VDDAVEK.

Table 13.14 Abundance of target peptides in selected *B. anthracis* strains. Each value is the mean value ± S.E. of three technical replicates.

	***B. anthracis* strains**				
	NCTC109	**NCTC2620**	**NCTC5444**	**NCTC7752**	**NCTC7753**
	Abundance of selected peptides (pg) in 10 μg of protein extract				
MDVDMLSNR	4.50 ± 0.23	8.24 ± 0.65	10.80 ± 1.05	12.65 ± 0.82	5.30 ± 0.11
AIGAELDQLVK	15.20 ± 0.24	47.44 ± 0.29	38.50 ± 0.52	77.12 ± 2.75	44.10 ± 1.01
LVSIGELQPDGNR	173.8 ± 6.85	423.3 ± 14.20	391.6 ± 27.89	548.4 ± 17.83	247.7 ± 4.45
SADLVQGLVDDAVEK	185.3 ± 40.65	210.6 ± 8.92	358.4 ± 2.63	307.6 ± 25.92	110.7 ± 8.43

13.24 Assay Sensitivity in Relation to Bacterial Cell Numbers

The capture and lysis of low-abundant cells was investigated to determine if proteins could be detected from samples with lower cell counts. Water represents a possible contamination source for many of these high-risk pathogens and was used as a background matrix for preliminary studies.

13.24.1 Method

C. sporogenes NCTC 275 was used as a test organism for developing and optimizing the assay.

13.24.2 Preparation of Cell Dilutions

C. sporogenes NCTC 275 was grown as described previously for *C. botulinum*.

The number of cells in 20 ml of broth with an OD_{600} 2–2.5 was determined by diluting 1 ml broth 1000-fold and determining the number of cells in 10 μl using a haemocytometer. An OD_{600} range 2–2.5 was calculated as corresponding to ~1.3×10^{10}–1.86×10^{10} *C. sporogenes* cells in a 20 ml broth. Cells were pelleted by centrifugation at 8000 g for 10 min and washed with distilled water three times to remove broth. Te pelleted cells were re-suspended in 1 ml distilled water to be used as the background matrix (the total number of cells ranged from 1.3×10^{10} to 1.86×10^{10} as shown above, which was calculated to produce ~299 μg of protein material based on a Bradford assay of whole cell extracts of this strain). A serial dilution of re-suspended cells: 10^{-1}–10^{-10} (100 μl in 1 ml) was prepared.

13.24.3 Cell Lysis Procedure

Four different approaches were investigated:

1) Boiling method: Samples boiled for 1 h 30 min.
2) Freeze/thawing: Freeze for 2 min, and thaw for 5 min for 3 cycles.

Table 13.15 Selected peptides, their corresponding stable isotope-labelled standard peptides, retention time and abundance of target peptides in selected *Clostridium botulinum* strains.

				***C. botulinum* strains**	
				NCTC13319	**NCTC3807**
Peptide	**Monoisotopic mass of the peptide**	**Production $[M+2H]^{2+}$ (m/z)**	**Retention time (min)**	**Abundance of selected peptides (pg) in 10 μg of protein extract**	
EAEYIFGNFGK	1273.6	637.8	22.7	37.8 ± 7.47	26.3 ± 7.77
EAEYIFΔGNFGK	1283.6	642.8			
YANIADYLSLGGK	1383.7	692.9	26.2	1160.9 ± 5.62	623.8 ± 48.90
YANIADYL*SLGGK	1390.7	696.4			

^{13}C and ^{15}N labelled amino acid leucine (L*). Mass increase 7 Da (e.g., six ^{13}C plus one ^{15}N) in comparison with standard amino acid. $^{\Delta}$ ^{13}C and ^{15}N labelled amino acid phenylalanine (FΔ). Mass increase 10 Da (e.g., nine ^{13}C plus one ^{15}N) in comparison with standard amino acid.

3) Sonication: Six cycles of probe sonication for 30 s on and 30 s off at maximum power.
4) Pressure Cycling Technology (PCT) Barocycler treatment: PCT was conducted at 35,000 psi for a total of 30 cycles, where each cycle consisted of 15 s at 35,000 psi and 15 s at ambient pressure.

13.24.4 Capture of Cells and Protein Material

Samples were spun through Amicon Ultra 0.5 ml centrifugal filters with a molecular weight cut-off of 3 kDa at 10,000 g to retain proteins and whole cells, and the filtrate was discarded. The filter column was treated with 500 µl standard lysis solution (SLS) for 30 min at room temperature and spun at 10,000 g until the solution ran through (SLS: 30 mM Tris-Cl pH 8.5, 7 M Urea, 2 M Thiorurea, 4% CHAPS and 70 mM DTT). The filter was washed with 100 µl SLS at 10,000 g until the solution ran through.

13.24.5 Trypsin Digestion on Filters

To remove salts and detergents, proteins captured were washed three times using ammonium bicarbonate buffer (100 mM, 500 µl × 3). Following each wash, the bicarbonate buffer was removed by centrifugation (14,000 g, ~30 min × 3). To perform disulfide reduction, reducing buffer (200 µl) containing 100 mM dithiothreitol (DTT) in the ammonium bicarbonate buffer (100 mM) was added into to each filter column. Reduction was carried out at 60 °C for 1 h. Reduction was stopped by removing excess DTT by centrifugation (14,000 g, ~30 min). Sulfhydryl alkylation was performed by adding alkylation buffer (200 µl) containing iodoaceamide (100 mM) in the ammonium bicarbonate buffer (100 mM) to the filter column. Alkylation was carried out at room temperature, in the dark for 1 h. To stop the alkylation, excess of iodoaceamide was removed by centrifugation (14,800 g, ~30 min), which was followed by washing steps using the ammonium bicarbonate buffer (500 µl × 3, ~30 × 3 min). Digestion buffer containing trypsin (10 ng/µl) in the ammonium bicarbonate buffer containing acetonitrile (80:20, v/v) was freshly prepared prior to trypsin digestion. To each filter column, 30 µl of digestion buffer was added. Trypsin digestion was carried out in a 37 °C incubator overnight. The filter columns were centrifuged at 14,000 g for ~30 min to collect tryptic peptides. Peptide samples were kept at −80 °C until nanoLC-MS/MS analysis. Analytical separation was achieved on a RP nano-column C18, 75 um i.d. × 15 cm at a flow rate of 0.3 µl min^{-1} as described previously, but with a gradient of 0%–25% acetonitrile in 54 min, increased to 50% acetonitrile in 37 min, then to 90% acetonitrile and held for 5 min before a return to the initial condition of 0% acetonitrile.

13.25 Summary of Results

So far, 10^{-6} to 10^{-10} dilution yielded no spectra from peptides derived from *C. sporogenes*, which could be due to too low cell numbers. 10^{-4}–10^{-5} dilution yielded reasonably good-quality spectra from peptides derived from *C. sporogenes*. All lysis methods investigated yielded *C. sporogenes* proteins that were detected in the 10^{-5} dilution, which equates to ~10^{5} cells. The results indicate that the sensitivity assay described is able to detect

peptides from ~10^5 cells in a matrix of water. On the basis of the protein concentration calculations for the whole cell extracts of *C. sporogenes* NCTC 275, 10^5 cells equates to ~2.99 ng of protein.

13.26 Spiked Samples

The specificity and sensitivity of the selected peptide markers were assessed by a spiking experiment using a serial dilution of a protein extract from the organism of interest spiked into a mixture of protein extracts from closely related members. This was tested with (1) *B. anthracis* and (2) *C. botulinum* samples.

13.27 Method

Protein extracts (10 µg) containing a serial dilution of *B. anthracis* NCTC 109 (2.0, 1.5, 1.0, 0.75, 0.5, 0.25, 0.1 and 0.0 µg) in a background of *B. cereus* containing *B. cereus, B. thuringiensis, B. weihenstephanensis* and *B. mycoides* (1:1:1:1, w/w/w/w) were separated on 4%–12% SDS/PAGE gels, and in-gel trypsin digestion was performed in the presence of isotope-labelled internal standards (60 fmol μl^{-1}) overnight at 37 °C, as described previously.

Protein extracts (10 µg) containing a serial dilution of *C. botulinum* type A strain NCTC 13319 (2.0, 1.5, 1.0, 0.75, 0.5, 0.25, 0.1 and 0.0 µg) in a background of *C. sporogenes* and *C. difficile* were separated on 4%–12% SDS/PAGE, gels and in-gel trypsin digestion was performed in the presence of isotope-labelled internal standards (120 fmol μl^{-1}) overnight at 37 °C, as described in the previous section.

1-D nanoflow LC-MS/MS, data-dependent MS analysis and data analysis were performed as described previously.

13.28 Summary of Results

For *B. anthracis*, none of the selected peptides was detected from the background protein mixture. Peptides LVSIGELQPDGNR and SADLVQGLVDDAVEK were detected when 0.1 µg of protein extract from *B. anthracis* NCTC109 was spiked into 9.90 µg of *B. cereus* protein extracts with signal-to-noise ratios of 75 and 16, respectively (Figure 13.6). Peptides MDVDMLSNR and AIGAELDQLVK gave signal-to-noises of 12 and 22, respectively, when 1.25 µg and 0.75 µg of protein extracts of *B. anthracis* NCTC109 were spiked into 8.75 µg and 9.25 µg of *B. cereus* protein extracts, respectively.

For *C. botulinum*, none of the selected peptides was detected from the background protein mixture. Peptide YANIADYLSLGGK was detected when 0.1 µg of protein extract from *C. botulinum* A NCTC13319 was spiked into 9.90 µg of *C. sporogenes* and *C. difficile* protein extracts with signal-to-noise ratios of 32. Peptide EAEYIFGNFGK gave a signal-to-noise ratio of 10 when 2.0 µg of protein extract from *C. botulinum* A NCTC13319 was spiked into a protein background containing 8.0 µg of protein extract from *C. sporogenes* and *C. difficile* (Table 13.18).

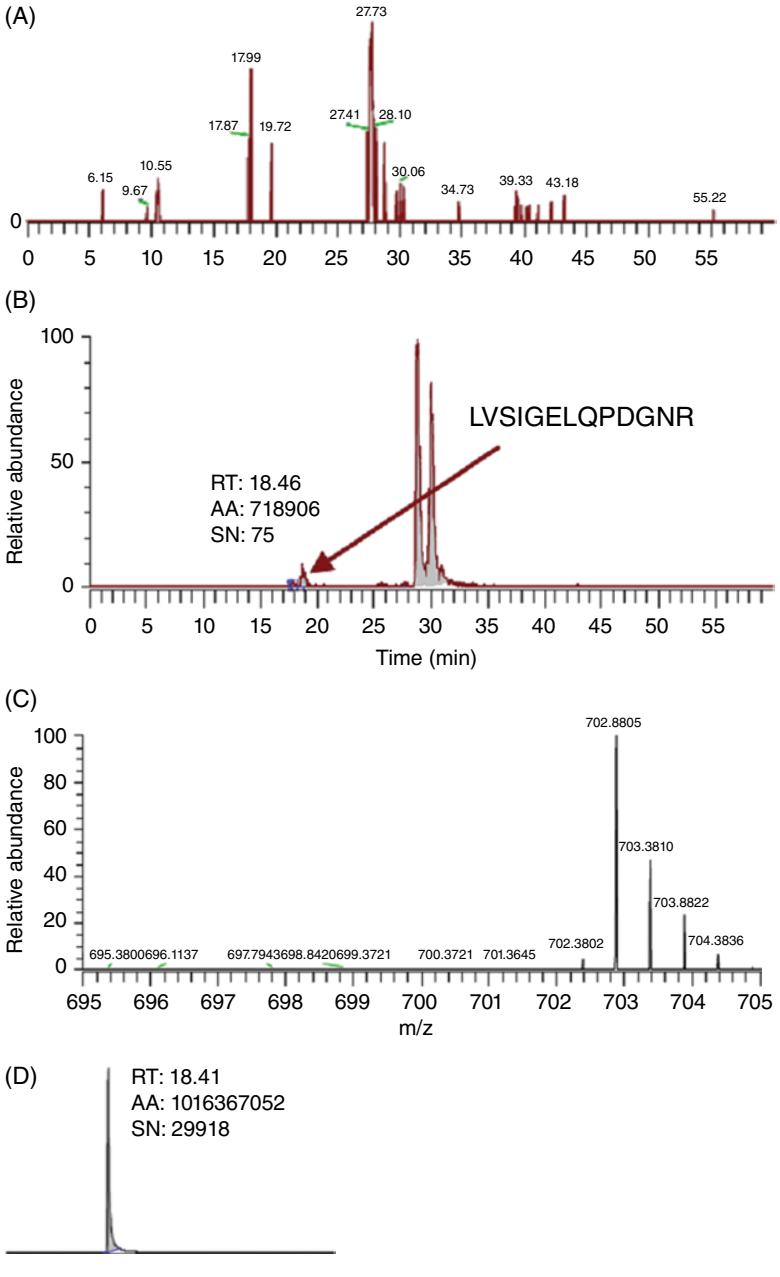


Figure 13.6 Detection of peptide LVSIGELQPDGNR from 0.1 µg of protein extract from *B. anthracis* NCTC109 spiked into a background of 9.90 µg of protein extract from *B. cereus* (*B. cereus*, *B. thuringiensis*, *B. weihenstephanensis* and *B. mycoides*). (A) No selected peptide (LVSIGELQPDGNR) is detected in background of *B. cereus*; (B) Peptide LVSIGELQPDGNR is detected with a signal-to-noise ratio of 75; and (C) MS spectrum of peptide LVSIGELQPDGNR and its corresponding stable isotope-labelled peptide internal standard LVSIGEL*QPDGNR; and inset (D) is the extracted ion chromatogram of the internal standard peptide.

13.29 Spiked Cells

Spiked cell samples were prepared in water using a 1 in 10 serial dilution (100 µl in 1 ml) of *C. botulinum* NCTC 13319 cells (starting material of ~10^8 cells) in a fixed background of *C. sporogenes* NCTC 275 (10^{10} cells). These cell numbers equate to ~157 µg and 299 µg of protein material, respectively, based on previous calculations of the protein concentration of whole cell extracts using a Bradford assay.

13.30 Method

1) Proteins were extracted using the procedure for whole cell protein extraction described for *C. botulinum*.
2) Protein extracts (10 µg) were separated on a 4%–12% Bis-Tris gel, and in-gel trypsin digestion was carried out in the presence of the internal standards described previously.
3) Data-dependent MS/MS was performed using an exclusion list containing 500 *C. sporogenes* peptides, in order to reduce the interference from *C. sporogenes* peptides.

13.31 Summary of Results

Peptide YANIADYLSLGGK was detected from the 10^{-2} dilution (~10^6 *C. botulinum* NCTC 13319 cells in the background of 10^{10} *C. sporogenes* NCTC 275 cells), which equated to ~1.57 µg of *C. botulinum* proteins in a total of ~1074 µg of protein material.

Peptide EAEYIFGNFGK was detected when 10^8 cells of *C. botulinum* NCTC 13319 cells were spiked into a background of 10^{10} *C. sporogenes* NCTC 275 cells), which equated to ~157 µg of *C. botulinum* proteins against a background of 744 µg total protein content.

13.32 *B. anthracis* Spore Analysis

Purified heat-inactivated spores were obtained for analysis of the spore proteome and detection of *B. anthracis* biomarkers.

13.33 Method

1) Protein extraction was performed using the PCT system under the same conditions described previously.
2) 10 µg of protein extract from the spores was separated on a 4%–12% Bis-Tris gel, and in-gel trypsin digestion was performed as described previously.

Table 13.16 Summary of *B. anthracis*–specific peptides.

Protein names	Peptide sequence
Inosine-uridine preferring nucleoside hydrolase family protein [*Bacillus anthracis* str. A0488]	ATSNAAYLLQLAGR
Small, acid-soluble spore protein B [*Bacillus anthracis* str. Ames]	LAVPGAESALDQMK LVSLAEQQLGGFQK YEIAQEFGVQLGADATAR
Hypothetical protein BA_1237 [*Bacillus anthracis* str. Ames]	AAALTVAQIS FFLSLGTPANIIPGSGTAVR IVPVELIGTVDIR MFSSDCEFTK SNVIGTGEVDVSSGVILINLNPGDLIR
Spore coat protein Z [*Bacillus anthracis* str. Ames]	AGAPFEAFAPSANLTSCR VESVDDDSCAVLR
Hypothetical protein BA_1173 [*Bacillus anthracis* str. Ames]	FIQQGADAVMK
30S ribosomal protein S7 [*Bacillus anthracis* str. Ames]	LANEILDAANNAGASVK
Hypothetical protein pxo1_137 [*Bacillus anthracis*]	VPGIIIAVDK
Hypothetical protein BA_4189 [*Bacillus anthracis* str. Ames]	ANADLTLAEIV VLYLEATSEK
Ribosome-binding factor A [Bacillus anthracis str. Ames]	IDTLLHEINK VGFVTVTDVQVSR
Hypothetical protein BAH_2616 [*Bacillus anthracis* str. A0442]	IEAIEALLT ITTGSNNYAGTVSSVTCDVVK LANSAGVVTVIISVCK
Hydrolase, alpha/beta fold family protein [*Bacillus anthracis* str. A1055]	ILQAAANEDVTK LVLVDAVSPSFVK

13.34 Summary of Results

Overall, 113 *Bacillus*-specific proteins (269 peptides) were identified. Of the 113, 10 proteins (23 peptides) were identified as *B. anthracis* specific (Table 13.8 and Table 13.16).

13.35 Assay Sensitivity in Relation to Bacterial Spore Numbers

13.36 Method

1) Serial dilutions of spores in water (starting material ~10^9 spores, which was found to produce ~375 µg of protein material) containing 10^8, 10^7, 10^6 and 10^5 spores were prepared.
2) Cell lysis was carried out using the PCT system as previously described.

3) *B. anthracis* spores and proteins were captured using Amicon Ultra 0.5 ml centrifugal filters, and trypsin digestion was performed on filters as described previously.
4) Data-dependent MS/MS was performed using two inclusive lists:
   a) 269 *Bacillus*-specific peptides
   b) 24 *B. anthracis*–specific peptides

13.37 Summary of Results

A total six *Bacillus*-specific proteins were identified from peptides derived from 10^5 *B. anthracis* spores (Table 13.17). On the basis of the protein content of an extract from ~10^9 spores, 10^5 spores equates to ~37.5 ng of spore protein.

13.38 Summary of Results for Biomarker Detection Sensitivity

It is vital that any biomarker detection assay has high specificity and sensitivity as well as the ability to detect biomarkers from a range of sample types.

The table below (Table 13.18) summarizes the preliminary data for biomarker detection in two model organisms (*B. anthracis* and *C. botulinum*). Various aspects were

Table 13.17 Summary of *B. anthracis*–specific proteins and peptides that were detected from 10^5 *B. anthracis* spores.

Protein names	Peptide sequence
Hypothetical protein BAM_2606 [Bacillus anthracis str. A0465]	ITTGSNNYAGTVSSVTCDVVK
S-layer protein EA1 [Bacillus anthracis str. Ames]	ADLYDTLTTK AEAAQFIALTDK ASAAVIFTK DNAQAYVTDVK INIGTVLELEK LDLNVSTTVEYQLSK LGDVTVSQTSDSALPNFK LSADDVTLEGDK TGVVAEGGLDVVTTDSGSIGTK VYSDPENLEGYEVESK
Spore coat protein B [*Bacillus thuringiensis* str. Al Hakam]	VEGILQDVSCDFVTLIVK
Spore coat protein B [*Bacillus cereus* W]	VGELVSLGK
Spore coat protein Z [*Bacillus anthracis* str. Ames]	VESVDDDSCAVLR
Hypothetical protein bthur0005_37820 [*Bacillus thuringiensis* serovar pakistani str. T13001]	EVKPQQPAVCNVLASISVGTELSLLSVK

Table 13.18 Summary of results for biomarker detection sensitivity and quantification.

			Results	
	Pathogen (strain)	**Biomarker**	**Amount of protein from target organism (μg)**	**Biomarker abundance (pg) in 10 μg of protein**
Biomarker detection	*B. anthracis* (NCTC 109, NCTC 2620, NCTC 5444, NCTC 7752, NCTC 7753)	MDVDMLSNR	**10**	**4.5–12.65**
		SADLVQGLVDDAVEK	**10**	**15–77**
		AIGAELDQLVK	**10**	**173–548**
		LVSIGELQPDGNR	**10**	**110–358**
	C. botulinum group 1 (NCTC 13319 toxin A, NCTC 3807 toxin B)	EAEYIFGNFGK	**10**	**26–38**
		YANIADYLSLGGK	**10**	**623–1161**
			Minimum ratio of whole cell protein extract (organism of interest/total protein) (μg) for biomarker detection	**Biomarker detection limit (fg)**
	B. cereus group spiked with *B. anthracis* (NCTC 109) (protein extracts)	MDVDMLSNR	1.25/10	165
		SADLVQGLVDDAVEK	0.1/10	107
		AIGAELDQLVK	0.75/10	224
		LVSIGELQPDGNR	0.1/10	198
	C. sporogenes/*difficile* spiked with *C. botulinum* NCTC 13319 (protein extracts)	EAEYIFGNFGK	0.1/10	201
		YANIADYLSLGGK	2/10	251
			Minimum no. of cells required for biomarker detection	***C. botulinum* total protein abundance (μg) in spiked sample**
	C. sporogenes spiked with *C. botulinum* NCTC 13319 (cells)	EAEYIFGNFGK	$\mathbf{10^8}$	**157 μg**
		YANIADYLSLGGK	$\mathbf{10^6}$	**1.57 μg**

(*Continued*)

Table 13.18 (Continued)

			Results	
	Pathogen (strain)	Biomarker	Amount of protein from target organism (μg)	Biomarker abundance (pg) in 10 μg of protein
Detection sensitivity			Minimum no. of cells required for detection of peptides from organism of interest	Minimum amount of protein required for detection of any peptide from organism of interest (ng)
	C. sporogenes cells (NCTC 275) in water	N/A	10^5	2.99
	B. anthracis spores in water	N/A	10^5	37.5

investigated such as the detection sensitivity of particular biomarkers, protein extractions from low-abundant cells and the analysis of spiked samples as well as the analysis of spores and cells in water (a possible source of contamination).

The data summarized in Table 13.18 highlights the sensitivity of biomarker detection. Detection limits vary for different markers due to the difference in their expression and therefore their abundance in the sample. The results show that detection is possible down to femtogram amounts and is also possible in spiked samples when other closely related species are present. The detection limits when analyzing diluted protein samples and spiked protein extracts was far more sensitive than when analyzing protein extracts from spiked cells and diluted cells. For example, detection of the EAEYIFGNFGK *C. botulinum* biomarker was achieved in a protein dilution that contained only 0.1 µg of *C. botulinum* proteins in a total of 10 µg of protein material from closely related species. In contrast, spiked cells of *C. botulinum* (10^8) in non–*C. botulinum* (10^{10}) that was calculated to equate to ~157 µg of *C. botulinum* in a total of 901 µg (1.7 µg in a 10 µg total used for SDS-PAGE) was the detection limit achieved. This differs by a factor of ~10, which indicates that the limitation on the sensitivity achieved may be due to the efficiency of the lysis method, and further optimization is required.

References

Al-Shahib, A., Misra, R., Ahmod, N., Fang, M., Shah, H. & Gharbia, S. (2010). Coherent pipeline for biomarker discovery using mass spectrometry and bioinformatics. *BMC Bioinformatics* **11**, 437.

van Belkum, A., Durand, G., Peyret, M., Chatellier, S., Zambardi, G., Schrenzel, J., Shortridge, D., Engelhardt, A. & Dunne, W. M. (2013). Rapid clinical bacteriology and its future impact. *Ann Lab Med* **33**, 14–27.

Brzuszkiewicz, E., Thürmer, A., Schuldes, J., Leimbach, A., Liesegang, H., Meyer, F.-D., Boelter, J., Petersen, H., Gottschalk, G. & Daniel, R. (2011). Genome sequence analyses of two isolates from the recent *ESCHERICHIA COLI* outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic Escherichia coli (EAHEC). *Arch Microbiol* **193**, 883–891.

Carbonnelle, E., Mesquita, C., Bille, E., Day, N., Dauphin, B., Beretti, J.-L., Ferroni, A., Gutmann, L. & Nassif, X. (2011). MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clin Biochem* **44**, 104–109.

Charnot-Katsikas, A., Tesic, V., Boonlayangoor, S., Bethel, C. & Frank, K. M. (2014). Prospective evaluation of the VITEK MS for the routine identification of bacteria and yeast in the clinical microbiology laboratory: Assessment of accuracy of identification and turnaround time. *J Med Microbiol* **63**, 235–241.

Chen, L., Xiong, Z., Sun, L., Yang, J. & Jin, Q. (2012). VFDB 2012 update: Toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res* **40**, D641–5.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R. & Tiedje, J. M. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res* **42**, D633–642.

Corina, D. L. & Sesardic, D. (1980). Profile analysis of total mycolic acids from skin corynebacteria and from named Corynebacterium strains by gas-liquid chromatography and gas-liquid chromatography/mass spectrometry. *J Gen Microbiol* **116**, 61–68.

Costas, M. (1990). Numerical analysis of sodium dodecyl sulphate-polyacrylamide gel electrophoretic protein patterns for the classification, identification and typing of medically important bacteria. In *Electrophoresis*, pp. 382–391.

Dauga, C. (2002). Evolution of the gyrB gene and the molecular phylogeny of Enterobacteriaceae: A model molecule for molecular systematic studies. *Int J Syst Evol Microbiol* **52**, 531–547.

Delmas, J., Breysse, F., Devulder, G., Flandrois, J.-P. & Chomarat, M. (2006). Rapid identification of Enterobacteriaceae by sequencing DNA gyrase subunit B encoding gene. *Diagn Microbiol Infect Dis* **55**, 263–268.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069–5072.

Dingle, T. C. & Butler-Wu, S. M. (2013). Maldi-tof mass spectrometry for microorganism identification. *Clin Lab Med* **33**, 589–609.

Everley, R. A, Mott, T. M., Wyatt, S. A, Toney, D. M. & Croley, T. R. (2008). Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and Shigella species. *J Am Soc Mass Spectrom* **19**, 1621–1628.

Fournier, P.-E., Drancourt, M., Colson, P., Rolain, J.-M., La Scola, B. & Raoult, D. (2013). Modern clinical microbiology: New challenges and solutions. *Nat Rev Microbiol* **11**, 574–585.

Shah, H. N., Keys, C., Gharbia S. E., Ralphson, K., Trundle, F., Brookhouse, I. & Claydon, M. (2000). The application of matrix-assisted laser desorption/ionisation time of flight mass spectrometry to profile the surface of intact. *Microb Ecol Health Dis* **12**, 241–246.

Ho, Y.-P. & Reddy, P. M. (2010). Identification of pathogens by mass spectrometry. *Clin Chem* **56**, 525–536.

Kaper, J. B., Nataro, J. P. & Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2**, 123–140.

Karas, M. & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299–2301.

Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M. & Shah, H. N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect Genet Evol* **4**, 221–242.

Loman, N. J., Misra, R. V, Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J. & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **30**, 434–439.

Mayberry, W. R. (1981). Dihydroxy and monohydroxy fatty acids in Legionella pneumophila. *J Bacteriol* **147**, 373–381.

Mellmann, A., Harmsen, D., Cummings, C. A., Zentz, E. B., Leopold, S. R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y. & other authors. (2011). Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS ONE* **6**, e22751.

Misra, R. V, Ahmod, N. Z., Parker, R., Fang, M., Shah, H. & Gharbia, S. (2012). Developing an integrated proteo-genomic approach for the characterisation of biomarkers for the identification of Bacillus anthracis. *J Microbiol Methods* **88**, 237–247.

Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S. & Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**, 548.

Paradis, S., Boissinot, M., Paquette, N., Bélanger, S. D., Martel, E. A., Boudreau, D. K., Picard, F. J., Ouellette, M., Roy, P. H. & Bergeron, M. G. (2005). Phylogeny of the Enterobacteriaceae based on genes encoding elongation factor Tu and F-ATPase beta-subunit. *Int J Syst Evol Microbiol* **55**, 2013–2025.

Pham, H. N., Ohkusu, K., Mishima, N., Noda, M., Monir Shah, M., Sun, X., Hayashi, M. & Ezaki, T. (2007). Phylogeny and species identification of the family Enterobacteriaceae based on dnaJ sequences. *Diagn Microbiol Infect Dis* **58**, 153–161.

Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R. & other authors. (2008). The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* **190**, 6881–6893.

Rasko, D. A., Webster, D. R., Sahl, J. W., Bashir, A., Boisen, N., Scheutz, F., Paxinos, E. E., Sebra, R., Chin, C.-S. & other authors. (2011). Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. *N Engl J Med* **365**, 709–717.

Sabat, A. J., Budimir, A., Nashev, D., Sá-Leão, R., van Dijl, J. M., Laurent, F., Grundmann, H. & Friedrich, A. W. (2013). Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro Surveill* **18**, 20380.

Schleifer, K. H. (2009). Classification of Bacteria and Archaea: Past, present and future. *Syst Appl Microbiol* **32**, 533–542.

Schulthess, B., Bloemberg, G. V., Zbinden, R., Böttger, E. C. & Hombach, M. (2014). Evaluation of the Bruker MALDI Biotyper for identification of Gram-positive rods: Development of a diagnostic algorithm for the clinical laboratory. *J Clin Microbiol* **52**, 1089–1097.

Shah, H. N. & Collins, M. D. (1983). Genus Bacteroides. A chemotaxonomical perspective. *J Appl Bacteriol* **55**, 403–416.

Shah, H. N., Keys, C. J., Schmid, O. & Gharbia, S. E. (2002). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and proteomics: A new era in anaerobic microbiology. *Clin Infect Dis* **35**, S58–64.

Stecher, B., Denzler, R., Maier, L., Bernet, F., Sanders, M. J., Pickard, D. J., Barthel, M., Westendorf, A. M., Krogfelt, K. A. & other authors (2012). Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc Natl Acad Sci U S A* **109**, 1269–1274.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. & Glöckner, F. O. (2014). The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res* **42**, D643–648.

14

Mapping of the Proteogenome of *Clostridium difficile* Isolates of Varying Virulence

Caroline H. Chilton,[1] Saheer E. Gharbia,[2] Raju V. Misra,[2] Min Fang,[3] Ian R. Poxton,[4] Peter S. Borriello[5] and Haroun N. Shah[3]

[1] *Leeds Institute for Biomedical and Clinical Sciences, University of Leeds, Leeds, UK*
[2] *Genomics Research, Public Health England, London, UK*
[3] *Proteomics Research, Public Health England, London and Department of Natural Sciences, Middlesex University, London, UK*
[4] *Department of Medical Microbiology, University of Edinburgh, Edinburgh, UK*
[5] *Veterinary Medicines Directorate, Surrey, UK*

14.1 Introduction

Clostridium difficile is a gram-positive, spore-forming anaerobic bacillus, and is the major infective agent of antibiotic-associated diarrhoea. *C. difficile* infection (CDI) is particularly problematic for hospitalized elderly populations, and can produce a variety of symptoms ranging from mild, self-limiting diarrhoea, to pseudomembraneous colitis and toxic megacolon. Since the association between antibiotic administration and *C. difficile* was first established (Chang, 1978; Larson *et al.*, 1978), CDI has emerged as an important nosocomial infection, in part due to the spread of highly epidemic strains. In particular, strains designated as PCR ribotype 027 caused large outbreaks in Quebec in the 2000s (Pepin *et al.*, 2004; McDonald *et al.*, 2005), and Stoke Mandeville in the United Kingdom in 2004/2005 (Smith, 2005). More recently, PCR ribotype 078 strains have caused clinical outbreaks, and have been associated with higher mortality than other PCR ribotypes (Walker *et al.*, 2013). Factors responsible for the epidemic potential of certain strains are poorly understood, and *C. difficile* continues to place a major burden on healthcare facilities worldwide. It has been estimated that the incremental costs associated with CDI range from £4577 to £8843 across Europe (Wiegand *et al.*, 2012) and from US$4846 to US$8570 in the United States (Ghantoji *et al.*, 2010).

This chapter will explore the use of genomics and proteomics to investigate factors responsible for the varying virulence of different *C. difficile* strains. We use our recent analysis of a range of *C. difficile* strains (Chilton *et al.*, 2014) as an example to demonstrate the power of combining these technologies in order to achieve a more complete picture of virulence determinants.

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

14.2 Virulence of *Clostridium difficile*

14.2.1 Virulence Factors

At first glance, the virulence of *Clostridium difficile* is a simple process, in that disease is mediated by two large clostridial toxins, toxin A and toxin B. These toxins are glycosyltransferases that disrupt the cytoskeleton and tight junctions of the cells, resulting in apoptosis (Just and Gerhard, 2004) and leading to diarrhoea, inflammation and epithelial tissue damage. Without these toxins, *C. difficile* does not cause disease. However, CDI outbreaks tend to be associated with certain highly epidemic strains (such as those of PCR ribotype 027), which by and large produce the same toxins as the non-epidemic strains. The relative role of toxin A and toxin B in the disease-causing process has been disputed (Lyras *et al.*, 2009; Kuehne *et al.*, 2010; Kuehne *et al.*, 2014), and toxin B positive, A negative strains have been responsible for CDI outbreaks (Alfa *et al.*, 2000). In addition, the role of an additional binary toxin produced by some strains remains ambiguous (Gerding *et al.*, 2014; Kuehne *et al.*, 2014), and the potential strain-to-strain differences in toxin regulation and amounts produced in vivo (Warny *et al.*, 2005; Freeman *et al.*, 2007; Dupuy *et al.*, 2008; Deneve *et al.*, 2009; Vohra and Poxton, 2011) remain unclear. The large variation in the toxinotypes of pathogenic *C. difficile* strains (Rupnik, 2008) indicates that factors other than toxin production contribute to pathogenicity and associated virulence.

The identity and role of other *C. difficile* virulence factors are less well understood. Adhesion to epithelial tissue may be a key early event in colonization, and is mediated by cell surface proteins. This has led to the identification of a number of surface proteins postulated to have a link to virulence, including adhesins, S-layer proteins, cell wall proteins and the flagella proteins (Faulds-Pain *et al.*, 2014) as well as a number of S-layer protein paralogs (Wright *et al.*, 2005). Other factors, such as antimicrobial susceptibility and resistance, may also affect the success and spread (hence epidemic potential) of a particular strain.

14.2.2 Variation between Strains

Clostridium difficile is an ubiquitous organism, present in the environment as well as the mammalian gastrointestinal tract. An enormous number of different strain types have been identified, although many of them are not associated with clinical disease. Interest in CDI was driven by the huge increase in cases seen in the mid 2000s (Pepin *et al.*, 2004; McDonald *et al.*, 2005; Smith, 2005). This rise in disease was associated with the emergence of highly epidemic clonal *C. difficile* strains, the most notorious being strains of PCR ribotype 027. Various reports have suggested that these strains may be associated with greater disease severity and mortality (Loo *et al.*, 2005), as well as increased levels of toxin production (Warny *et al.*, 2005), although whether these strains deserve the 'hypervirulent' label has been disputed. Ribotype 027 strains continue to dominate in the United States and Eastern Europe, whereas the United Kingdom and Western Europe have seen a decrease in infections due to these strains (Valiente *et al.*, 2014). More recently, strains of PCR ribotype 078 have caused serious outbreaks in the United Kingdom (Walker *et al.*, 2013), and have been associated with more severe infection. Ultimately, although many suggestions have been raised, a definitive reason for the epidemic potential and widespread dissemination of ribotype

027 (and to a lesser extent other epidemic strains) has not been identified, nor have any strain-specific virulence factors that may explain observed differences in disease severity of different strains.

14.3 Current Genomic and Proteomic Data

In recent years, genomic data has widely been used to identify differences between strains, and the huge increase in the availability of sequencing technologies has facilitated this. The genomes of different *C. difficile* strains from different isolates spanning the last four decades have been sequenced, including strain 630 (Sebaihia *et al.*, 2006), which has more recently been updated (Monot *et al.*, 2011) and two PCR 027 isolates, a recent, epidemic isolate R20291 and a 'historic' non-epidemic isolate, CD196 (Stabler *et al.*, 2009). These have respectively enabled genomic comparison studies of multiple isolates (Stabler *et al.*, 2006; He *et al.*, 2010; Dingle *et al.*, 2013), confirming the huge genomic variation of this pathogen. Epidemic PCR ribotype 027 strains have been shown to cluster into a tight clade (Stabler *et al.*, 2009). However, disease-associated isolates belong to multiple lineages, indicating that certain genetic elements may underlie virulence, and as a consequence of the highly dynamic nature of the genome, that these elements may be transferable by horizontal gene transfer and recombination (He *et al.*, 2010; Dingle *et al.*, 2013).

However, there remains a paucity of proteomic data for *C. difficile.* Studies have looked at cell surface proteins (Wright *et al.*, 2005), spore proteins (Lawley *et al.*, 2009), and the insoluble proteome (Jain *et al.*, 2010) of the 630 reference strain. Changes over time (Janvilisri *et al.*, 2012), and in response to antimicrobial peptides (McQuade *et al.*, 2012), have been reported for single strains, as has investigation of metronidazole resistance mechanisms (Chong *et al.*, 2014). A comparative analysis of the *C. difficile* secretome (Boetzkes *et al.*, 2012) has recently been reported; however, only one comparative analysis of proteomic differences between historic and recently emerged *C. difficile* isolates has so far beer described (Chen *et al.*, 2013), and this does not include a low-virulence strain. Crucially, detailed biological characterization data are rarely available in conjunction with genomic and proteomic analysis.

14.4 Comparison of Strains of Varying Virulence

In order to maximize the power of genomic and proteomic data in the elucidation of factors responsible for *C. difficile* virulence, we performed tandem genomic and proteomic analysis on *C. difficile* strains exhibiting different virulence profiles. We then compared the genomic and proteomic profiles to known phenotypic data for these strains in an attempt to identify factors that could explain the differences in virulence. Three strains were used in this study, designated B-1, Tra 5/5 and 027 SM. Comparative analysis of *C. difficile* strains in the hamster model of disease (Borriello *et al.*, 1987) has shown that strain B-1 is highly virulent, but strain Tra 5/5 is of lower virulence (demonstrated by the number of animals dying after challenge with *C. difficile*). The virulence of strain 027 SM has not been tested in the hamster model, but many studies have characterized epidemic, PCR ribotype 027 outbreak strains, and some studies have linked these strains to increased disease severity.

14.5 Genomic Analysis of *Clostridium difficile*

In general, current whole genome sequencing methods, such as the so-called 'next generation sequencing (NGS)' approaches, generate many short-length assembled sequences, referred to as contiguous sequences or 'contigs'. It is not unusual for bacterial genomes such as *C. difficile* to generate 10s or even 100s of contigs; the number of contigs may indicate the complexity of the genome. For example, when reconstructing the genome from many shorter sequencing reads, genomes that contain many repeat regions may lead to a higher number of contigs from a single sequencing run. Repeat regions vary in length. Some, such as rDNA, can be >5 kb, far longer than many of the current read lengths achieved by NGS approaches. In parallel with the read length, sequencing accuracy is vital if one is to faithfully capture the original genome sequence. At present, base calling accuracy can drop due to a variety of reasons, some of which are platform specific; for example, accurate base calling over long homopolymer regions has proved a challenge for Roche/454- and PGM-based sequencers, more so for Roche/454. The Illumina systems are not immune to error biases; although greatly improved, Illumina datasets were shown to suffer from higher (compared to other Roche and PGM) substitution error rates and an issue referred to as 'phasing', which leads to sequence quality deterioration towards the end of a read. One way to compensate for these error types is by sequencing many reads across the genome or loci of interest, i.e. generate a high read depth, from which a consensus sequence can be generated. Most current NGS methods rely on amplifying a population of cells, and the consensus represents the 'average' or most dominant sequence present in that population.

14.5.1 Using Roche's Flx and Junior

The three *C. difficile* strains (027 SM, B-1 and Tra 5/5) were sequenced by Roche, using the Roche FLX platform (raw reads are available in EMBL-ENA; ERP002519 (027 SM), ERP002520 (B-1) and ERP002521 (Tra 5/5)). The Roche (FLX and Junior) sequencing method, in brief, relies on fragmenting DNA, ligating adaptors and attaching the adaptor-ligated DNA fragments to DNA capture beads. The DNA-attached beads (DaB) are added to a water–oil emulsion, in which the DNA is amplified directly on the beads, via a process referred to as emulsion PCR. Each amplified DaB is placed into individual wells, in a PicoTitrePlate (PTP). The PTP serves as the sequencing platform to which the enzymes DNA polymerase, ATP sulfurylase and luciferase are added and used to sequence the DNA on each DaB. The PTP with enzymes are placed in to the sequencer (FLX or Junior) and using a series of microfluidics, nucleotides are washed over the PTP in a set order, via a process referred to as pyrosequencing. As the name suggests, it is light based, whereby every time a nucleotide is incorporated, a flash of light is generated. The on-board CCD camera records the flashes of light in the order the bases are incorporated, translating them into a nucleotide sequence.

The Roche FLX system can generate reads >700 bases in length and can achieve an average quality (Q40) of >99%. The raw reads were assembled for annotation using the Roche Newbler v2.5 analysis package, and a summary of the assembly metrics is presented in Table 14.1. The complete genomes of each strain were compared using the Mauve progressive algorithm (Darling *et al.*, 2010).

Table 14.1 Roche assembly metrics. A comparable number of open reading frames (ORFs) were identified in strains 027 SM (3896) and Tra 5/5 (3840), similar to the number identified in the 630 strain. However, a larger number of ORFs (4061) were identified in strain B-1. Extra-chromosomal data were generated for three plasmids in strain B-1, and one plasmid in strain Tra 5/5.

Strain	027 SM	Tra5/5	B-1
Coverage	56x	42x	37x
Number of assembled reads	1,578,926	840,932	674,628
Number of paired-end reads	454,956	66,174	102,750
Number of contigs (large/all)	131/235	169/1026	148/294
Number of scaffolds (bp)	11	12	17
Average scaffold length/largest scaffold (bp)	376,652/1,971,874	342,691/1,818,314	260,750/1,938,015
% of bases >Q40	99.93	99.78	99.87

14.5.2 PacBio Genomic Analysis

Ideally, when performing whole genome sequencing, a single complete contig per genomic feature which faithfully recreates the original genome, such as a chromosome, plasmid or phage, is sought. Although a challenge for short read sequencers, long read sequencing technologies such as the PacBio© (Pacific Biosystems) can generate reads >20 kb in length, enabling long length repeat regions to be captured by a single read, thus greatly reducing the number of contigs. The PacBio© sequencing method is based on single-molecule, real-time (SMRT®) technology, which is built upon the implementation of Zero-Mode Waveguides (ZMWs) and an immobilized DNA polymerase. A single strand of DNA acts as a template, interacting with the DNA immobilized DNA polymerase, to which phospholinked nucleotides are added, all within a ZMW 'well', where a number of ZMW wells are created in a single sequencing cell, referred to as an SMRT © cell. The bottom of the well is illuminated, and a pulse of light is observed with the addition of a nucleotide. In addition to capturing the nucleotide incorporation and thus the template sequence, the polymerase kinetics, that is, the interaction dynamics between the added nucleotide, template and polymerase allowed for the direct detection of DNA base modifications.

Using the three Roche sequenced strains, 027 SM, B-1 and Tra5/5, PacBio© sequencing was also performed. A minimum of 10 µg of high-quality DNA, where the extracted DNA ideally should be enriched for >60 kb in length, to help ensure an average of around 20 kb read lengths could be achieved. Post DNA extraction, the samples were prepared for sequencing, and the resultant read datasets, including modifications, were used to piece together the genome sequence via an in silico process referred to as 'de novo assembly'; that is, the genome is assembled independent of a reference sequence. The assembler used was the HGAP long-read assembler, which after using three SMRT cells worth of data, could assemble the chromosome for each of the three strains into a single contig, with no gaps. Due to the large volume of data generated, a >100 average fold coverage across the genome could be achieved, from which the consensus sequence had a quality (phred) score of >Q60, the equivalent of a possible error every 1,000,000 bases. In the context of the *C. difficile* genome (chromosome), which is approximately

Table 14.2 PacBio sequencing metrics.

Strain	Genome size (bases)	Chromosome contigs	Plasmid/Phage contigs	Average coverage	Genes	rRNA	Signal peptides	tRNA
A027	4,199,633	1		>100	3824	33	177	87
Tra5/5	4,128,151	1		>100	3741	35	169	90
Tra5/5	45,012		1	>100	67		3	
B-1	4,388,148	1		>100	4096	29	181	87
B-1	46,764		1	>100	74		3	
B-1	41,090		1	>100				
B-1	12,488		1	>100			2	

four megabases in length, the error rate would be equivalent to four possible errors across the entire chromosome. The ability to generate complete, highly accurate genomes provides a strong foundation for downstream studies such as proteomics.

The genomes were annotated, and the results are summarized in Table 14.2.

14.6 Proteomic Analysis of *Clostridium difficile*

For proteomic analysis, the whole-cell soluble proteome was extracted from *C. difficile* cultures grown on Columbia Blood agar for 64 h. Three different proteomic approaches were used to maximize coverage.

14.6.1 Two-Dimensional Reference Mapping

Two-dimensional gel electrophoresis (2D GE) separates mixtures of proteins by both isoelectric point and size, thereby achieving adequate separation of highly complex protein mixtures to allow visualization and identification of individual protein spots (Figure 14.1). In this study, a pH gradient of 4–7 was used for separation in the first dimension, and proteins were equilibrated with DTT and iodoacetamide before separation in the second dimension through a 10% acrylamide gel. Gels were stained with SYPRO® Ruby Protein Gel stain (BioRad) and Sigma Brilliant Blue G-colloidal (Sigma), allowing individual protein spots to be excised from the gel and subject to trypsin digestion. Digested peptides were then extracted and analyzed by MALDI-TOF mass spectrometry (reflection mode) to give an individual peptide mass fingerprint for each protein spot. These were compared to protein sequence data from the NCBI database using MASCOT to generate protein identifications. Large numbers of proteins from a complex protein extraction can be visualized and identified using this technique. Here, 107, 126 and 158 protein spots were identified for strains 027 SM, B-1 and Tra5/5, respectively, and 2D proteome reference maps could be created for each strain. An example reference map is shown in Figure 14.1.

One advantage of 2D GE is that it allows visualization of post-translational modifications. Cleavage or chemical modification (glycosylation/phosphorylation) following the translation process can generate proteins of different molecular weights or isoelectric points, leading to differential migration during the 2D separation process. In *C. difficile*, the

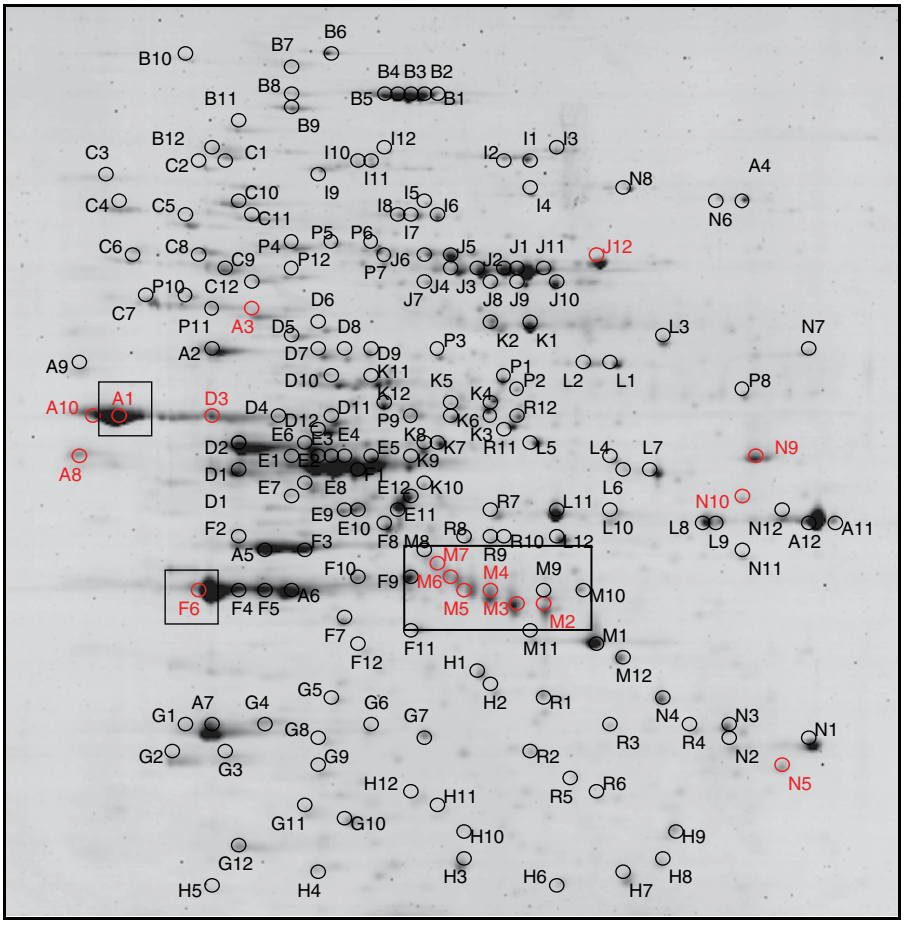


Figure 14.1 Spots picked for in-gel digestion and identification on the strain 027 SM reference map. Surface and virulence proteins are highlighted in red. The two spots corresponding to the S-layer protein (SlpA) are boxed (spots A1 and F6). A post-translational cleavage event creates separate high-molecular-weight and low-molecular-weight S-layer proteins. The series of spots corresponding to flagellin are also boxed (M2-M7). Here, post-translational glycosylation of flagellin leads to a series of protein spots which migrate differently.

S-layer protein (SlpA), is post-translationally cleaved to give high-molecular-weight-and low-molecular-weight protein subunits, which together form the S-layer protein lattice. This leads to two distinct, highly abundant spots on the 2D reference map, both identified as SlpA (Figure 14.1). Post-translational glycosylation of flagellin can also be visualized on the 2D reference map, where the glycosylation process creates a series of spots with a characteristic migration pattern, also reported for other organisms (Figure 14.1).

14.6.2 Differential In-Gel Electrophoresis (DIGE)

Coupled with the 2D GE reference mapping, this study utilized fluorescence labelling of protein extracts to perform DIGE, allowing a comparison of protein expression levels between different strains. DIGE allows three differently labelled protein samples to

be run together on the same gel, thereby removing gel-to-gel variation, and allowing accurate quantitative comparison of proteins between different samples. Typically, DIGE experiments are set up with a protein standard (a mix of all protein extracts compared in the experiment) run alongside two separately labelled protein samples. Each protein extract can then be compared to the standard to compare spot intensity (and hence expression levels) between samples. Following scanning at the relevant fluorescent wavelengths, differences between differently labelled samples can be clearly seen in different colours. In this study, DIGE identified a number of clear protein spot differences between different strains, as highlighted below for strains 027 SM and strain B-1 (Figure 14.2). Differences can be due to proteins present in one strain but not in another (Figure 14.2, box 1 and 2), or differences in the migration of proteins from different strains due to inherent variation in the molecular weight or isoelectric point arising either from the amino acid sequence (Figure 14.2, box 4), or from differences in post-translational modifications (Figure 14.2, box 3).

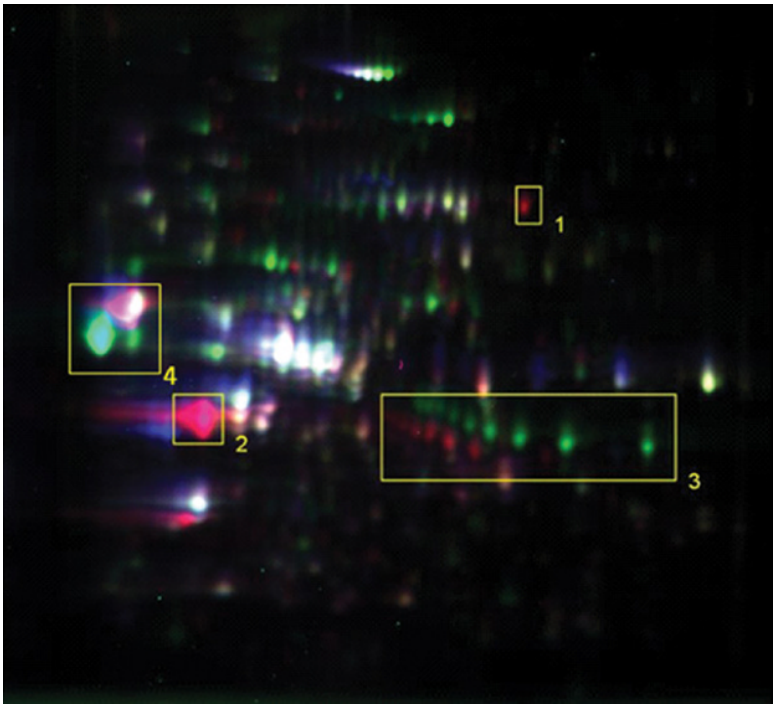


Figure 14.2 A DIGE gel image of strain 027 SM (Cy5, red) compared to strain B-1 (Cy3, green). The Cy2 channel (blue) is the internal standard containing a mix of protein extracts from all strains. The yellow boxes 1 and 2 highlight protein spots corresponding to surface proteins (Cwp2, SlpA) present in strain 027 SM but not strain B-1. Yellow box 3 highlights a series of spots identified as the same protein (flagellin) but modified post-translationally. The series of Cy3 and Cy5 labelled proteins are clearly visible as separate spots which have migrated differently within the gel, due to differences in the post-translational modification of this protein between the different strains. Yellow box 4 highlights the SlpA high-molecular-weight (HMW) protein, which shows differential migration between the strains. (Chilton *et al.*, 2014.)

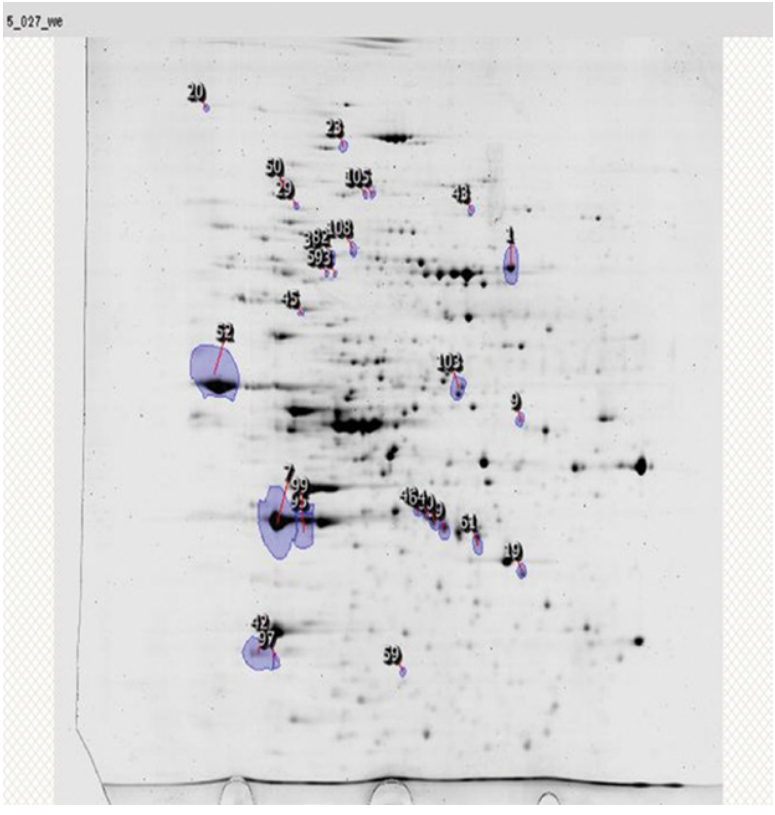


Figure 14.3 Proteins spots with significantly higher concentration in strain 027 SM. The proteins identified as up-regulated in strain 027 SM by correlation analysis were matched to the 'picking gel' used to create the strain A reference map. The numbers indicate the rank of the protein, with protein 1 showing the greatest difference between the strains.

Using specialist software, DIGE gels can be matched using the protein standards (run on each gel), enabling identification of protein spots with different intensities (expression levels) in different strains. Moreover, the DIGE gels can be matched to picking gels or reference maps to allow identification of these proteins of interest (Figure 14.3).

In this study, SameSpots software (progenesis) matched 453 proteins across the standards of all six DIGE gels, and identified 112 spots with a greater than twofold expression difference (ANOVA $p < 0.05$) between strains. By correlating these spots to the reference maps, five, ten and eight proteins showing significantly higher expression in strains 027 SM, B-1 and Tra 5/5 were identified (Figure 14.3).

14.6.3 One-Dimensional Gel Electrophoresis Coupled with LC-MS/MS

Although 2D reference mapping allows visual comparison of protein extracts and has the added bonus of enabling comparison of post-translational modifications, it can only be used to compare relatively abundant proteins. Lower-abundance proteins, which are not visible as individual spots on the gels, cannot be compared. Using higher-throughput techniques alongside reference mapping can greatly improve proteome coverage.

In this study, we used one-dimensional electrophoresis followed by LC-MS/MS using an LTQ Orbitrap (thermo-fisher) to increase protein detection. Protein extracts were separated by 1D gel electrophoresis (separation by mass only) on 10% acrylamide gels. Each lane was cut into sections of approximately 1 cm, and each gel section subject to trypsin digestion as with the 2D gel spots. Peptide mixtures were separated by HPLC using an Ultimate 3000 Dionex nano/capillary system coupled with a thermo LTQ Orbitrap, and protein identifications made by matching peptides to the non-redundant *C. difficile* database.

The 1D Gel electrophoresis LC-MS/MS workflow identified a total of 862 proteins, 734 (85%) of which had not been identified from the 2D reference maps. Only 332 (38%) of these were detected in all three strains. A total of 132 proteins (17%) were detected only in strain A, 161 (19%) were detected only in strain B-1 and 45 (5%) were detected only in strain Tra5/5.

14.7 Mapping the Proteogenome of *Clostridium difficile* to Phenotypic Profiles

Using the described range of genomic and proteomic analysis techniques, this study enabled mapping of phenotypic, genetic and expression data to investigate in detail individual aspects of virulence. Below we will discuss some aspects of virulence and how proteogenome mapping has highlighted strain-to-strain differences and given some insight into how these differences may be linked to virulence.

14.7.1 Toxin Expression

The two toxins, toxin A and toxin B, remain the most characterized and understood virulence factors for *C. difficile.* Toxin production is required for symptomatic disease, and consequently toxin levels have been used as a measure of virulence. Borriello *et al.* (1987) demonstrated that in vitro, the highly virulent strain B-1 produces significantly greater levels of toxin A than strain Tra 5/5, but the same maximum titre of toxin B (Borriello *et al.*, 1987). Many studies have characterized epidemic, PCR ribotype 027 outbreak strains and suggested that they may produce higher levels of toxin in vitro (Warny *et al.*, 2005). We used the gold-standard vero cell cytotoxicity assay to show that strain 027 and B-1 produced comparable levels of toxin in vitro and >10 times greater amounts of toxin that strain Tra 5/5 (Chilton *et al.*, 2014). This method does not distinguish between toxin A and toxin B, however.

Both toxins are encoded within the pathogenicity locus (PaLoc) along with three other ORFs postulated to encode transcription regulators and a holin-like protein (Figure 14.4). Genomic analysis confirmed that all three strains possess an intact pathogenicity locus, indicating the potential to produce both toxins A and B. Notably, genomic analysis also revealed that strain 027 SM contains an 18 bp deletion in the TcdC gene, absent in the other two strains. This deletion has been widely reported in ribotype 027 strains (Barbut *et al.*, 2007). It has been suggested that *TcdC* acts a negative regulator of toxin production, and that this deletion may lead to increased toxin production by ribotype 027 strains (Carter *et al.*, 2011). However, strain B-1 does not contain this deletion, yet produces comparable levels of toxin in vitro (Chilton *et al.*, 2014).

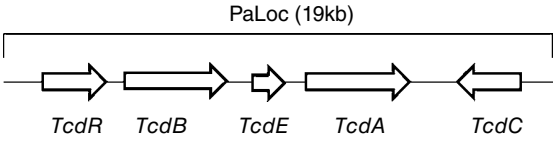


Figure 14.4 The *Clostrdium difficile* pathogenicity locus containing the two toxin genes (TcdB and TcdA), a σ-factor (TcdR) thought to positively regulate transcription of the toxin genes, a possible holin (TcdE) and a potential negative transcription regulator TcdC.

Strains B-1 and Tra 5/5 were genetically very similar in the PaLoc, yet showed ~10-fold difference in the toxin levels produced, suggesting that toxin production may be additionally controlled by factors outside of the PaLoc region. This is supported by recent work disputing the role of *TcdC* as a negative regulator of toxin production (Cartman *et al.*, 2012).

No toxins were identified on the 2D reference maps, probably due to the large size of these proteins hampering separation. However, using the 1D LC-MS/MS approach, toxin A was detected in all three strains, whereas toxin B was detected in strains 027 SM and B-1, but not strain Tra 5/5. The relative roles of toxin A and toxin B in the disease-causing process is not fully understood, but the apparent lower levels of toxin B produced by strain Tra 5/5 may contribute to the reported lower virulence of this strain, although the molecular reasons for reduced toxin production remain unclear.

14.7.2 Mucosal Adherence

Adherence of *C. difficile* to the intestinal mucosa locates the toxin-producing organism to the site of toxin action (intestinal epithelial cells), and so factors involved in mucosal adherence are often considered virulence factors, although the role adhesion plays in virulence is still not fully understood. Some studies have demonstrated that *C. difficile* strains showing variation in their ability to colonize and adhere to the gastrointestinal tract exhibit corresponding variation in virulence in the hamster model. Borriello *et al.* (1988) demonstrated that strain B-1 has a high level of adherence, and greater virulence compared to other strains. A number of cell surface proteins coordinating adhesion of the bacterial cell to the gut wall have been identified, although the full mechanism has not been elucidated.

S-layer (Surface layer) proteins have been linked to adhesion (Calabi *et al.*, 2002) and shown to be essential for virulence in some bacterial pathogens including *Aeromonas salmonicida* and *Campylobacter fetus* (Sara and Sleytr, 2000), and so S-layer proteins and their homologues were investigated in some detail during this study.

The S-layer of *C. difficile* is made up of two protein subunits, a HMW protein and a low-molecular-weight (LMW) protein, which are derived from a single gene, *slpA* (Calabi *et al.*, 2001) and cleaved post-translationally. During our analysis, SlpA protein was identified in all three strains and clearly identified on reference maps (Figures 14.1 and 14.2). The predicted pI and molecular weight of this protein was determined from the genetic sequence of *SlpA*, and showed variation between the strains (predicted pI = 4.83, 4.76 and 4.79, and predicted masses = 80,428 Da, 64,703 Da and 79,922 Da for strains 027 SM, B-1 and Tra 5/5, respectively). This variation is confirmed by differential migration of this protein during 2D DIGE (Figure 14.1, box 4). SlpA was present as

two separate spots on the reference maps of strain 027 SM and Tra 5/5, corresponding to the HMW and LMW mature proteins. Interestingly, however, only one SlpA protein subunit was observed on the B-1 reference map (Figure 14.1, boxes 2 and 4). Another major difference in the surface proteins of strain B-1 compared to the other two strains was in the lack of detection of SlpA homologue Cwp2 in this strain (Figure 14.2, box 1). To investigate the genetic basis for the differences in the expressed surface proteins of strain B-1, we compared the genetic organization of the SlpA containing operon using the online progressive Mauve algorithm (Darling *et al.*, 2010).

The *C. difficile* genome contains a number of *SlpA* homologues. Many of these ORFs cluster near *SlpA* in a locus containing 17 open reading frames thought to form an operon (Savariau-Lacomme *et al.*, 2003). We showed that the organization of this operon was very similar in strains 027 SM and Tra 5/5, but markedly different in strain B-1, which has undergone genetic rearrangement in this area, with an insertion of approximately 58 kb (Figure 14.5). This insertion has replaced the open reading frame encoding the S-layer homologue Cwp2, removing this gene from the B-1 genome, and thus providing a genetic reason for the lack of detection of this protein in this strain. The presence of this insertion, however, does not seem to block the transcription of downstream ORFs, with expression being confirmed by protein detection in many cases (SecA and Cwp84 were detected in all strains). A similar insertion into this genetic locus has recently been described in a number of strains by Dingle *et al.* (Dingle *et al.*, 2013).

This study highlighted genetic variation in the S-layer operon on strain B-1 data and indicates that strains containing this genetic variation may show corresponding variation in surface layer protein expression. A similar lack of LMW SLP protein has been identified in a strain designated 167 by Calabi and Fairweather (Calabi and Fairweather, 2002) The un-cleaved pre-protein of this strain had a predicted molecular mass of 62,312 Da, lower than that of other isolates, but similar to the predicted mass for strain B-1 (64,703Da). The authors proposed that the lack of LMW SLP in strain 167 could be due to rapid degradation of this subunit, and showed that this did not affect the pathogenicity or cellular viability of the strain. The differences in S-layer profiles, however, do not obviously correspond with virulence, as the two highly virulent strains B-1 and 027 SM show marked differences in S-layer expression profiles, whereas the lower-virulence Tra 5/5 strains show high homology with the 027 SM strain. These results correlate with many published analysis (Dingle *et al.*, 2013; Spigaglia *et al.*, 2011), and suggest that the relationship between expression of S-layer proteins and virulence is highly complex.

14.7.3 Flagella

Flagella proteins have been shown to contribute to bacterial virulence in a number of ways (Ramos *et al.*, 2004). They confer motility, allowing access of bacteria to mucosal tissues, and as with other cell surface structures, they can act as adhesins. Aflagellate strains have been reported to associate with caecal tissue at a slower rate than flagellate strains (Tasteyre *et al.*, 2001), which in turn may affect strain virulence. In addition, it has been suggested that flagella can play a role in sensing environmental conditions such as 'wetness' (Wang *et al.*, 2005) and hence may control expression of genes and virulence factors under certain conditions (Boin *et al.*, 2004).

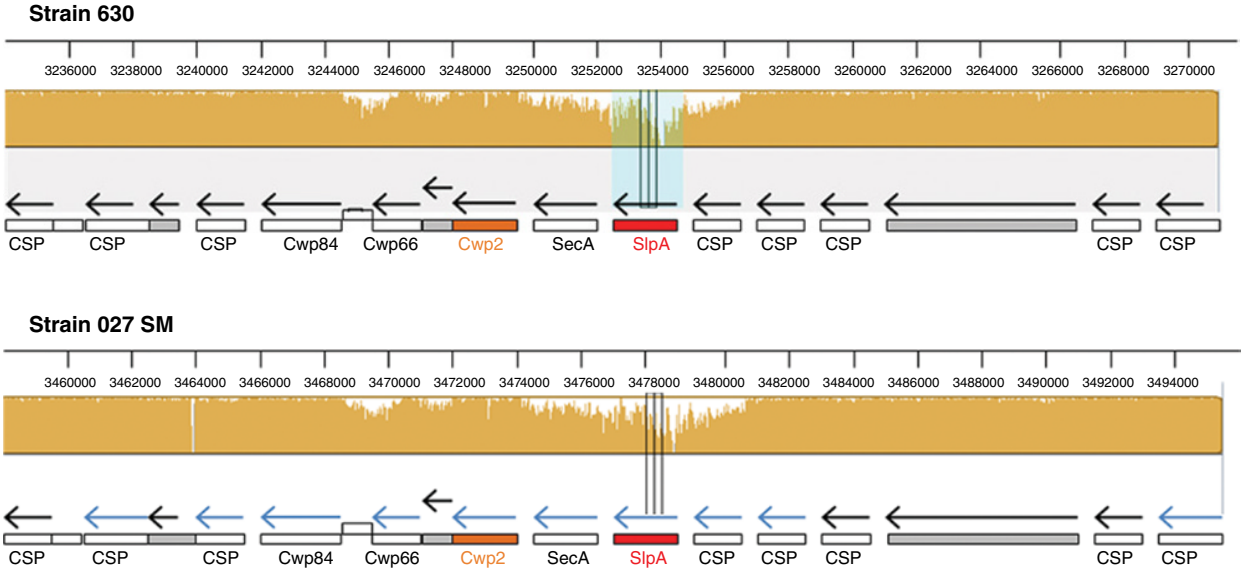


Figure 14.5 Mauve analysis comparing the organization of the SlpA operon for strains 027 SM, B-1, Tra5/5 and the 630 reference strain. Yellow bars represent areas of genetic homology between the strains, with the height of the bar representing the level of homology and ORFs being designated by bars and arrows below. The SlpA gene is shown in red in all strains, and highlighted in green in strain 630, and the gene encoding Cwp2 is shown in orange. Hypothetical or putative ORFs are shown in grey. Arrows above the genes show the direction of transcription; blue arrows indicate that the corresponding protein was detected in this analysis. CSP denotes genes encoding cell surface proteins (Slp homologues). In strain B-1, this genetic locus shows considerable differences from the other three genomes, with an insertion of approximately 58 kb. This insertion contains 50 ORFs, the majority of being putative and uncharacterized.

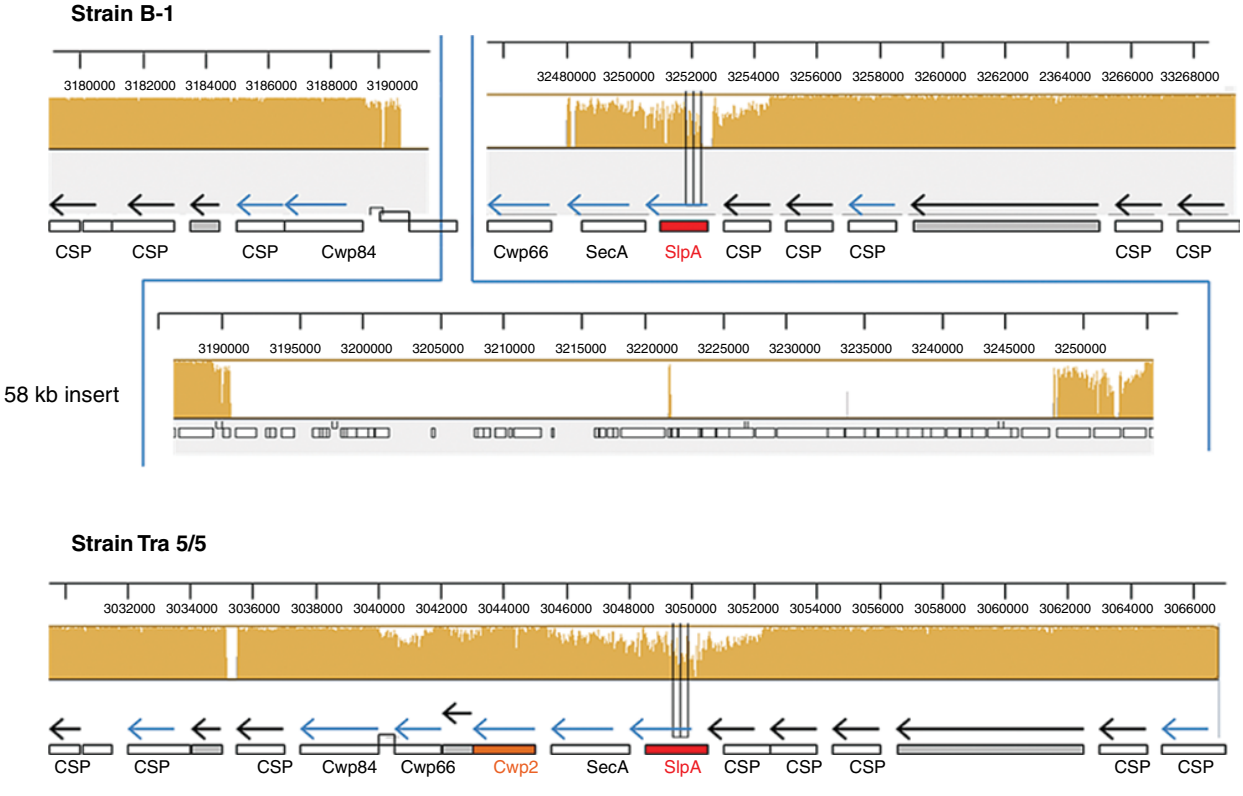


Figure 14.5 (Continued)

During this study, there was some variation in flagella proteins identified between the three strains. Flagellin (CD0239) was identified in all three strains, whereas the flagellar hook protein (CD0255) was identified in strains B-1 and 027 SM, but was not detected in strain Tra 5/5. The genes encoding both proteins were present in the genomes of all three strains, however, demonstrating that a lack of the corresponding gene is not the reason for lack of the flagella hook protein in strain Tra 5/5. Flagellin was identified as multiple distinct spots following two-dimensional gel electrophoresis (Figure 14.1), a phenomenon associated with post-translational modifications. Moreover, the pattern of these spots was different in different strains, as highlighted by DIGE analysis. Glycosylation of flagellin in *C. difficile* has been described (Twine *et al.*, 2009; Faulds-Pain *et al.*, 2014), and this glycosylation process has been shown to be important in many bacterial species, affecting motility, adhesion and pathogenesis (Logan, 2006). Differences in the flagellar-associated glycosyl transferases have been shown between epidemic 027 ribotypes and non-epidemic strains (Stabler *et al.*, 2006), and it has been confirmed that flagellin-linked glycans show divergence between different strains (Twine *et al.*, 2009). Variations in flagellin migration have been seen in *Campylobacter jejuni*, correlating with introduction of flagellar glycosyltransferase knock-out mutations(van Alphen *et al.*, 2008), raising the interesting suggestion that flagellin glycosylation could vary between the three strains investigated during this study. This in turn could affect structure and function of the flagella, altering the role played in motility and adhesion.

The putative flagellar glycan biosynthesis locus reported by Twine *et al.* (2009) was compared in detail using the progressive Mauve algorithm, and found to be conserved between the three strains, indicating that genetic variation in this region does not explain the differential migration patterns. Interestingly, the 15 kb insert into the B-1 S-layer operon described above has been linked with possible glycosylation functions (Dingle *et al.*, 2013), and therefore may play a role.

Strain B-1 has shown evidence of flagella and motility (Tasteyre *et al.*, 2000), and epidemic 027 ribotype strains have been shown to be motile (Stabler *et al.*, 2009), although the presence of flagella was not confirmed microscopically. Strain Tra 5/5, however, tested negative for motility (Chilton *et al.*, 2014), indicating that this strain may not possess functioning flagella. Previous comparative genomic assays have indicated that flagella production and motility may not be a common feature for *C. difficile* isolates, and that flagella gene loci are not necessarily conserved among virulent strains (Stabler *et al.*, 2006), although they appear to be conserved for the three strains investigated here. It has been reported by RT-PCR that flagellin is expressed in both flagellate and aflagellate strains (Tasteyre *et al.*, 2000), so it is possible that strain Tra 5/5 is aflagellate, despite the detection of flagellin. This would correlate with the observed lack of motility of this strain, and the lack of detection of the flagellar hook protein. Dingle *et al.* have shown that disruption of flagellin genes leads to lack of flagella expression and motility, but does not affect adherence to Caco-2 cells, or virulence in the hamster model (as determined by percentage survival) (Dingle *et al.*, 2011), suggesting that if Tra 5/5 is aflagellate, this would not explain an apparent reduced virulence. However, recent work by Aubry *et al.* indicates that some flagellar mutants exhibit reduced toxin production, indicating that the flagellar regulon modulates toxin production (Aubry *et al.*, 2012). This raises the interesting possibility that the lack of flagella detection in strain Tra 5/5 may be linked to the lack of toxin detection.

14.8 Antibiotic Resistance

C. difficile is notorious for being multidrug resistant, and CDI is closely associated with antibiotic use. During the 1970s, increases in CDI were attributed to clindamycin use, in the 1980s to cephalosporins and penicillins while in the 2000s fluoroquinolones have been implicated (10). Careful antimicrobial stewardship has been used to limit CDI risk. Therapeutic options for CDI treatment remain limited to metronidazole and vancomycin, with the recent addition of fidaxomicin. However, reports of reduced metronidazole efficacy, especially for severe cases, and reduced metronidazole susceptibility of clinical isolates raise questions about treatment efficacy. Genetic markers for certain antibiotic resistances have been collated into curate databases, which contain information on specific antimicrobial resistance determinants, based on years of integrated phenotypic and molecular studies. However, predicting antibiotic resistance from the genome or proteome remains a contentious area, not just for *C. difficile* but in clinical microbiology as a whole. This controversy is due to the complexity of resistance, as it is not simply an 'on or off' process but one with many potential 'states' of resistance, based on the many potential permutations of gene and gene expression of resistance determinants, which denote the resistance phenotype. The challenge lies with capturing this complexity, with the aim of informing the clinician whether an isolate is resistant to an antibiotic and if so how resistant. Today, the current gold standards remain phenotypic, such as minimum inhibitory concentration (MIC) testing. The MIC is the lowest concentration at which an antimicrobial inhibits the growth of a microorganism after overnight incubation, and the resultant data indicates both whether an isolate is resistant and if so, how resistant (see Table 14.3).

To determine the antibiotic resistance profiles for the three *C. difficile* strains, MIC tests were performed, and the profile is given below, where: Met = Metronidazole, Vanc = Vancomycin, Fidax = Fidaxomicin, Rif = Rifampicin, Moxi = Moxifloxacin, Clinda = Clindamycin, Imi = Imipenem, Chlor = Chlorophenicol and Tige = Tigecycline.

Table 14.3 To determine whether, based on MIC profiles like those above, an isolate is resistant or sensitive to a particular antibiotic, external bodies such as the British Society for Antimicrobial Chemotherapy (BSAC) offer guidance notes on MIC cut-offs, referred to as 'MIC break points'. For example, the MIC breakpoints for the commonly prescribed (for *C. difficile* infections) antibiotics, metronidazole and vancomycin, are 'sensitive ≤2 mg/l and resistant >2 mg/l'. In the context of the three described strains, A027, B-1 and Tra5/5, all three are sensitive and thus treatable using either metronidazole or vancomycin.

	Minimum inhibitory concentration of antimicrobial (mg/l)								
Strain	**Met**	**Vanc**	**Fidax**	**Rif**	**Moxi**	**Clinda**	**Imi**	**Chlor**	**Tige**
A027	2	1	0.125	0.004	16	1	4	4	0.03
B-1	1	2	0.015	<0.001	2	>64	2	8	0.03
Tra5/5	<0.125	1	0.03	0.002	2	2	2	4	0.03

14.9 Conclusion

As is true of many bacterial pathogens, the virulence of *C. difficile* is a truly multifactorial process, with many different factors and pathways contributing to the epidemic potential of different strains. The recent advances in genomic technologies have greatly increased our understanding of the genetic variation of this organism; however, in order to fully understand strain-to-strain variation, genetic analysis alone is insufficient. Proteomic analysis allows comparison of gene expression and variations in protein modification, which may be key in determining virulence potential. Other technologies not discussed here, such as transcriptomics and metabolomics can add additional important information to strain comparisons. In short, studies using combinations of the growing range of 'omics' technologies can gain greater insight into bacterial variations than those employing single technologies only.

References

Alfa, M. J. *et al.* 2000. Characterization of a toxin A-negative, toxin B-positive strain of Clostridium difficile responsible for a nosocomial outbreak of *Clostridium difficile*-associated diarrhea. *J Clin Microbiol.* **38**(7), pp. 2706–2714.

Aubry, A. *et al.* 2012. Modulation of toxin production by the flagellar regulon in Clostridium difficile. *Infect Immun.* **80**(10), pp. 3521–3532.

Barbut, F. *et al.* 2007. Prospective study of *Clostridium difficile* infections in Europe with phenotypic and genotypic characterisation of the isolates. *Clin Microbiol Infect.* **13**(11), pp. 1048–1057.

Boetzkes, A. *et al.* 2012. Secretome analysis of *Clostridium difficile* strains. *Arch Microbiol.* **194**(8), pp. 675–687.

Boin, M.A. *et al.* 2004. Chemotaxis in Vibrio cholerae. *FEMS Microbiol Lett.* **239**(1), pp. 1–8.

Borriello, S.P. *et al.* 1987. *Clostridium difficile* – a spectrum of virulence and analysis of putative virulence determinants in the hamster model of antibiotic-associated colitis. *J Med Microbiol.* **24**(1), pp. 53–64.

Calabi, E. *et al.* 2002. Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infect Immun.* **70**(10), pp. 5770–5778.

Calabi, E. and Fairweather, N. 2002. Patterns of sequence conservation in the S-Layer proteins and related sequences in *Clostridium difficile*. *J Bacteriol.* **184**(14), pp. 3886–3897.

Calabi, E. *et al.* 2001. Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Mol Microbiol.* **40**(5), pp. 1187–1199.

Carter, G. P. *et al.* 2011. The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of *Clostridium difficile*. *PLoS Pathog.* **7**(10), pe1002317.

Cartman, S. T. *et al.* 2012. Precise manipulation of the *Clostridium difficile* chromosome reveals a lack of association between the tcdC genotype and toxin production. *Appl Environ Microbiol.* **78**(13), pp. 4683–4690.

Chang, T. W. 1978. *Clostridium difficile* toxin and antimicrobial agent-induced diarrhea. *J Infect Dis.* **137**(6), pp. 854–855.

Chen, J. W. *et al.* 2013. Proteomic comparison of historic and recently emerged hypervirulent *Clostridium difficile* strains. *J Proteome Res.* **12**(3), pp. 1151–1161.
Chilton, C. H. *et al.* 2014. Comparative proteomic analysis of *Clostridium difficile* isolates of varying virulence. *J Med Microbiol.* **63**(Pt. 4), pp. 489–503.
Chong, P. M. *et al.* 2014. Proteomic analysis of a NAP1 *Clostridium difficile* clinical isolate resistant to Metronidazole. *PLoS ONE.* **9**(1), p10.
Darling, A. E. *et al.* 2010. progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE.* **5**(6), pe11147.
Deneve, C. *et al.* 2009. New trends in *Clostridium difficile* virulence and pathogenesis. *Int J Antimicrob Agents.* **33** Suppl 1, pp. S24–28.
Dingle, K. E. *et al.* 2013. Recombinational switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by large-scale whole-genome sequencing. *J Infect Dis.* **207**(4), pp. 675–686.
Dingle, T. C. *et al.* 2011. Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. *Infect Immun.* **79**(10), pp. 4061–4067.
Dupuy, B. *et al.* 2008. *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *J Med Microbiol.* **57**(Pt. 6), pp. 685–689.
Faulds-Pain, A. *et al.* 2014. The post-translational modification of the *Clostridium difficile* flagellin affects motility, cell surface properties and virulence. *Mol Microbiol.* **94**(2), pp. 272–289.
Freeman, J. *et al.* 2007. Effect of metronidazole on growth and toxin production by epidemic *Clostridium difficile* PCR ribotypes 001 and 027 in a human gut model. *J Antimicrob Chemother.* **60**(1), pp. 83–91.
Gerding, D. N. *et al.* 2014. *Clostridium difficile* binary toxin CDT: Mechanism, epidemiology, and potential clinical importance. *Gut Microbes.* **5**(1), pp. 15–27.
Ghantoji, S. S. *et al.* 2010. Economic healthcare costs of *Clostridium difficile* infection: a systematic review. *J Hosp Infect.* **74**(4), pp. 309–318.
He, M. *et al.* 2010. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc Natl Acad Sci USA.* **107**(16), pp. 7527–7532.
Jain, S. *et al.* 2010. Proteomic analysis of the insoluble subproteome of *Clostridium difficile* strain 630. *FEMS Microbiol Lett.* **312**(2), pp. 151–159.
Janvilisri, T. *et al.* 2012. Temporal differential proteomes of *Clostridium difficile* in the pig ileal-ligated loop model. *PLoS ONE.* **7**(9), pe45608.
Just, I. and Gerhard, R. 2004. Large clostridial cytotoxins. *Rev Physiol Biochem Pharmacol.* **152**, pp. 23–47.
Kuehne, S. A. *et al.* 2010. The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature.* **467**(7316), pp. 711–713.
Kuehne, S. A. *et al.* 2014. Importance of toxin A, toxin B, and CDT in virulence of an epidemic *Clostridium difficile* strain. *J Infect Dis.* **209**(1), pp. 83–86.
Larson, H. E. *et al.* 1978. *Clostridium difficile* and the aetiology of pseudomembranous colitis. *Lancet.* **1**(8073), pp. 1063–1066.
Lawley, T. D. *et al.* 2009. Proteomic and genomic characterization of highly infectious *Clostridium difficile* 630 spores. *J Bacteriol.* **191**(17), pp. 5377–5386.
Logan, S. M. 2006. Flagellar glycosylation – a new component of the motility repertoire? *Microbiology.* **152**(Pt. 5), pp. 1249–1262.

Loo, V. G. *et al.* 2005. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N Engl J Med.* **353**(23), pp. 2442–2449.

Lyras, D. *et al.* 2009. Toxin B is essential for virulence of *Clostridium difficile*. *Nature.* **458**(7242), pp. 1176–1179.

McDonald, L. C. *et al.* 2005. An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N Engl J Med.* **353**(23), pp. 2433–2441.

McQuade, R. *et al.* 2012. *Clostridium difficile* clinical isolates exhibit variable susceptibility and proteome alterations upon exposure to mammalian cationic antimicrobial peptides. *Anaerobe.* **18**(6), pp. 614–620.

Monot, M. *et al.* 2011. Reannotation of the genome sequence of *Clostridium difficile* strain 630. *J Med Microbiol.* **60**(Pt. 8), pp. 1193–1199.

Pepin, J. *et al.* 2004. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *CMAJ.* **171**(5), pp. 466–472.

Ramos, H. C. *et al.* 2004. Bacterial flagellins: Mediators of pathogenicity and host immune responses in mucosa. *Trends Microbiol.* **12**(11), pp. 509–517.

Rupnik, M. 2008. Heterogeneity of large clostridial toxins: Importance of *Clostridium difficile* toxinotypes. *FEMS Microbiol Rev.* **32**(3), pp. 541–555.

Sara, M. and Sleytr, U. B. 2000. S-Layer proteins. *J Bacteriol.* **182**(4), pp. 859–868.

Savariau-Lacomme, M. P. *et al.* 2003. Transcription and analysis of polymorphism in a cluster of genes encoding surface-associated proteins of *Clostridium difficile*. *J Bacteriol.* **185**(15), pp. 4461–4470.

Sebaihia, M. *et al.* 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet.* **38**(7), pp. 779–786.

Smith, A. 2005. Outbreak of *Clostridium difficile* infection in an English hospital linked to hypertoxin-producing strains in Canada and the US. *Eurosurveillance.* **10**(6), E050630.2.

Spigaglia, P. *et al.* 2011. The LMW surface-layer proteins of *Clostridium difficile* PCR ribotypes 027 and 001 share common immunogenic properties. *J Med Microbiol.* **60**(Pt. 8), pp. 1168–1173.

Stabler, R. A. *et al.* 2006. Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. *J Bacteriol.* **188**(20), pp. 7297–7305.

Stabler, R. A. *et al.* 2009. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* **10**(9), pR102.

Tasteyre, A. *et al.* 2001. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun.* **69**(12), pp. 7937–7940.

Tasteyre, A. *et al.* 2000. Phenotypic and genotypic diversity of the flagellin gene (fliC) among *Clostridium difficile* isolates from different serogroups. *J Clin Microbiol.* **38**(9), pp. 3179–3186.

Twine, S. M. *et al.* 2009. Motility and flagellar glycosylation in *Clostridium difficile*. *J Bacteriol.* **191**(22), pp. 7050–7062.

Valiente, E. *et al.* 2014. The *Clostridium difficile* PCR ribotype 027 lineage: a pathogen on the move. *Clin Microbiol Infect.* **20**(5), pp. 396–404.

van Alphen, L. B. *et al.* 2008. A functional *Campylobacter jejuni* maf4 gene results in novel glycoforms on flagellin and altered autoagglutination behaviour. *Microbiology.* **154**(Pt. 11), pp. 3385–3397.

Vohra, P. and Poxton, I. R. 2011. Comparison of toxin and spore production in clinically relevant strains of *Clostridium difficile*. *Microbiology.* **157**(Pt. 5), pp. 1343–1353.

Walker, A. S. *et al.* 2013. Relationship between bacterial strain type, host biomarkers, and mortality in *Clostridium difficile* infection. *Clin Infect Dis.* **56**(11), pp. 1589–1600.

Wang, Q. *et al.* 2005. Sensing wetness: A new role for the bacterial flagellum. *EMBO J.* **24**(11), pp. 2034–2042.

Warny, M. *et al.* 2005. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet.* **366**(9491), pp. 1079–1084.

Wiegand, P. N. *et al.* 2012. Clinical and economic burden of *Clostridium difficile* infection in Europe: a systematic review of healthcare-facility-acquired infection. *J Hosp Infect.* **81**(1), pp. 1–14.

Wright, A. *et al.* 2005. Proteomic analysis of cell surface proteins from *Clostridium difficile*. *Proteomics.* **5**(9), pp. 2443–2452.

15

Determination of Antimicrobial Resistance using Tandem Mass Spectrometry

Ajit J. Shah[1], Vlad Serafim[1], Zhen Xu[2], Hermine V. Mkrtchyan[2] and Haroun N. Shah[1]

[1] *Department of Natural Sciences, Middlesex University, London, UK*
[2] *School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*

15.1 Antibiotic Resistance Mechanisms

Antibiotic resistance is a rising health care problem. The significant surge in total number of bacterial pathogens that show multiple drug resistance in recent years has been vast. Several disease prevention and control bodies, including the World Health Organization (WHO), are seeing the infections by antibiotic resistant pathogens as a key global health threat which will have a major impact in global health management. With the introduction of every new class of antibacterial agents into the clinic, resistance follows. Initially the antibiotic is effective, but with time and as a result of selective pressures, a small population of organisms acquires or develops resistance to the antibacterial agent (Table 15.1). Resistance can occur regardless of exposure to an antibacterial agent, as a result of either mutations or the absorption of mobile genetic elements which harbour resistant genes. However, the increase in proportion and number of microorganisms that are resistant to antibiotics has been aided by abuse and misuse of antibiotics in both clinic and in livestock farming and the discharge of these agents into the environment. Furthermore, the spread of these resistant organisms has been facilitated by the speed and increase in the amount of intercontinental travel.

There are several different mechanisms that bacteria use to resist the action of antibiotics. They produce enzymes that hydrolyze or modify antibiotics to render them inactive (Bush, 1999; Shakil *et al.*, 2008; Smith and Baker, 2002); produce efflux pumps that transport antibiotics out of the organism (Blair *et al.*, 2014); reduce the permeability of the outer membrane, which restricts antibiotic access (Sutterlin *et al.*, 2014); mutations in genes coding for penicillin binding proteins (PBPs) reduce the binding affinity of antibiotics to the active site (Sun *et al.*, 2014); and they alter metabolism (Zapun *et al.*, 2008). Classic methods of antibiotic sensitivity testing such as the disk diffusion assay provide a means of classifying the organism as susceptible, intermediately susceptible or resistant. However, this test does not reveal the minimum inhibitory concentration (MIC) required; it simply provides a qualitative measurement. On the other hand, the E-strip test is a quantitative method used for measuring antibiotic resistance. The major drawback of this technique is that a separate strip is required for each antibiotic, and

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

Table 15.1 Timeline of the discovery, introduction and first observed resistance of antibiotics.

Antibiotic	Year of discovery	Year of introduction	Year resistance observed	Mechanism of action	Activity or target species
β-lactams	1928	1936	1942	Inhibition of cell wall biosynthesis	Broad-spectrum activity
Sulphonamides	1932	1936	1942	Inhibition of dihydropteroate synthetase	Gram-positive bacteria
Aminoglycosides	1943	1946	1946	Binding of 30s ribosomal subunit	Broad spectrum
Tetracyclines	1944	1952	1950	Binding of 30s ribosomal unit	Broad spectrum
Cephalosporins	1948	1964	1966	Inhibition of cell wall biosynthesis	Broad spectrum
Pleuromutilins	1950	1979	1982	Binding of 50s ribosomal unit	Gram positive
Macrolides	1948	1951	1955	Binding of 50s ribosomal unit	Broad spectrum
Streptogramins	1953	1998	1956	Binding of 50s ribosomal unit	Gram positive
Glycopeptides	1953	1958	1960	Inhibition of cell wall biosynthesis	Gram positive
Quinolones	1961	1968	1968	Inhibition of DNA synthesis	Broad spectrum
Monocarboxylic acids (mupirocin)	1971	1985	1987	Binds to isoleucyl tRNA synthetase	Gram positive
Carbapenems	1976	1985	1985	Inhibition of cell wall biosynthesis	Broad spectrum
Oxazolidinones	1978	2000	2001	Binding of 50s ribosomal unit	Gram positive bacteria
Monobactams	1979	1986	1981	Inhibition of cell wall synthesis	Gram-negative
Lipopeptides	1986	2003	1987	Depolarization of cell membrane	Gram positive
Diarylquinolines	1997	2012	2006	Inhibition of ATP synthetase	Mycobacterium tuberculosis

thus the cost of analysis can be high. Although the cost is one factor that is used in any treatment regime, it may be more important to provide a rapid method for determining the resistant mechanism and thus formulate a treatment regime. Methods based on PCR are used for the detection of β-lactamases in clinical samples. However, with the spectrum of β-lactamases that have been identified, it would be a challenge to develop a ubiquitous primer that could be employed for detection of all β-lactamases. Methods based on spectrophotometric detection have been reported (Barlam and Neu, 1984). In these methods, a chromogenic cephalosporin is incubated with bacterial extract, and the absorbance is measured at 486 nm. The hydrolysis of the peptide bond results in a rise in absorbance. The main drawback of this assay is that it is time consuming and cannot be used for routine analysis.

15.2 Detection of β-lactamase Activity

One way to react to increasing antibiotic resistance is to use rapid methods for identification of microorganisms to aid not only prescription of the most cost-effective treatment strategy, but also transmission reduction. The decline in hospital-acquired incidence of methicillin-resistant *Staphylococcus aureus* (MRSA) is attributable to the implementation of suitable screening procedures (Huttner *et al.*, 2013). The rapid rise in use of mass spectrometry for bacterial identification and biotyping over the last few years, especially matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS), has placed pressure to broaden the applicability of the technique in clinical microbiology. One application which is increasingly being utilized is detection of antibiotic resistance, especially that towards the β-lactams.

β-lactams are the most commonly prescribed class of antibacterial drugs that show a broad spectrum of antibacterial activity. This class of drugs have a common structural feature, the β-lactam ring. Their mode of action is inhibition of bacterial cell wall biosynthesis. Their widespread use and abuse in both clinic and livestock has culminated in an increase in resistance to this class of antibacterial agents. The principal resistant mechanism to β-lactams is the hydrolysis of the β-lactam ring by β-lactamases. Some β-lactams undergo further degradation in which decarboxylation takes place. The resulting products are inactive and lower in molecular weight than the parent molecule (Figure 15.1).

Two methods have been reported for the classification of β-lactamases (Bush and Jacoby, 2010). One of these is the molecular classification system. This is based on the primary amino acid sequence of enzymes that have a serine residue in the active site

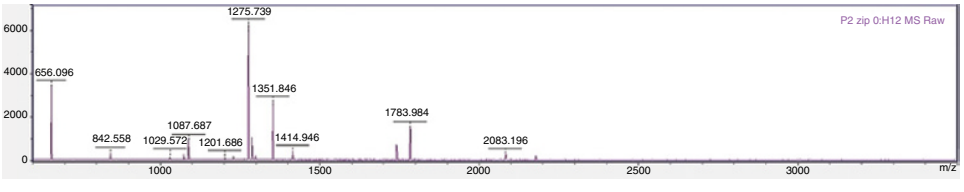


Figure 15.1 Peptide mass fingerprint of protein band corresponding to molecular weight of 29 kDa on SDS PAGE extracted from an ampicillin-resistant *E. coli*.

(classes A, C and D) which is involved in the hydrolysis of the β-lactams and the metallo-enzymes (class B) which require zinc ions for hydrolysis of β-lactams. The other method is based on functional classification which takes into account both the β-lactam and enzyme inhibitor profiles to place the enzymes into different groups. Cephalosporinases (group 1); broad-spectrum, inhibitor-resistant, and extended-spectrum β-lactamases and serine carbapenemases (group 2); and metallo-β-lactamase (group 3). Of these, carbapenemases are the most versatile, having a wide spectrum of activity; they hydrolyze most β-lactams and are resistant to most inhibitors. Resistance to carbapenems is now widespread. Detection of organisms producing the New Delhi metallo-β-lactamase-1 is now worldwide and has raised concerns. The most clinically common carbapenemases are KPC, VIM, IMP, NDM-1 and OXA 48 (Nordmann *et al.*, 2011). An assay that has been utilized clinically for detection of carbapenemase activity is the Hodge test (Girlich *et al.*, 2012), which measures growth of organisms in the presence of a carbapenem. It has very good sensitivity for detection of KPC and OXA-48 carbapenemases. However, its sensitivity for NDM-1 is lower. Furthermore, the specificity of this assay is low and the throughput is poor.

One approach that is becoming increasingly popular for the detection of β-lactamase activity and in particular carbapenemase is MALDI-TOF MS (Hrabák *et al.*, 2011; Lange *et al.*, 2013; Sauget *et al.,* 2014) (Chapter 10). The methods are based on the detection of bacteria-induced hydrolysis of β-lactam antibiotics. The method is relatively straightforward. A fresh bacterial culture is suspended in buffer and centrifuged, and the resulting pellet is incubated with a solution of the antibiotic for 1–3 h at 35 °C. The sample is then centrifuged or filtered combined with a matrix and then analyzed using MALDI-TOF MS. The sensitivity and resistance pattern are confirmed by analyzing the mass spectra for hydrolyzed and non-hydrolyzed forms of the β-lactam. For example, incubation of meropenem with an organism that is carbapenemase results in a spectrum with peaks attributable to molecular ions of meropenem $[M + H]^+$ at *m/z* 384.5 Da and its corresponding sodium adducts $[M+ Na]^+$ at *m/z* 406.5 Da and $[M + 2Na]^+$ at *m/z* 428.5 Da. Following incubation of meropenem with a carbapenemase-resistant strain, it has been shown that peaks assigned to the non-hydrolyzed form of the antibiotic decrease in signal intensity and can be detected (Sparbier *et al.*, 2012). However, no signals corresponding to hydrolysis products were detected. Recently, it has been shown that by including 0.01 % sodium dodecyl sulphate in the reaction buffer, degradation decarboxylated $[M\text{-}CO_2]^+$ at *m/z* 358 Da and its sodium adduct $[M\text{-}CO_2 + Na]^+$ at *m/z* 380.5 Da can be detected. To address any degradation due to spontaneous hydrolysis, the assay is normally performed with a negative control, and the organism is incubated with antibiotic in the presence of an inhibitor 3-aminophenylboronic acid. The method is suitable for detection of different classes of β-lactamases. The hydrolysis of the β-lactam ring gives rise to a new peak at +18 Da, which is a significant shift for mass spectrometric detection. In the case of some antibiotics, for example, piperacillin, decarboxylation of the hydrolyzed β-lactam gives a further mass shift of −44 Da. Additional peaks attributable to salt adducts of degradation products, usually sodium, may also be detected. One of the issues that has been highlighted with using MALDI-TOF MS for detection of antibiotic resistance is that both the intact antibiotic, degradation products and adducts are < 1000 Da and thus may be prone to interference from matrix ions (Hoff *et al.*, 2012). α-Cyano-4-hydroxycinnamic acid (CHCA) is a matrix that is normally used at mg/ml concentration in these types of analyses, and it has been

found that the dimer of this matrix with an *m/z* of 380 Da can interfere with the meropenem having an *m/z* of 384.5 Da. Also, detection of β-lactamase activity in the presence of other mechanisms of resistance such as efflux pumps, which have been observed in species of *Pseudomonas*, can be problematic. These prevent entry of antibiotic into the cell and thus interaction of the β-lactam with enzyme. One way to address this issue is to lyse the cells and incubate the extract with the antibiotic. A method reported by Hrabák *et al.* showed that addition of 0.1% sodium dodecyl sulphate to the incubation buffer increases the permeability of the outer membrane to antibiotics, thus allowing the interaction to take place.

MALDI-TOF MS is a valuable tool for detection of β-lactamases. The detection of β-lactamase-producing microorganisms using MALDI-TOF MS would benefit enormously from application software that would provide a means for automatic acquisition and interpretation of results. In 2014, Bruker Daltonics launched Microbial Biotyping Star BL®, which provides a rapid means of analyzing mass spectra of β-lactamase-producing microorganisms incubated with β-lactam antibiotics.

15.3 Other MALDI-TOF MS Methods

Resistance to aminoglycosides is facilitated by a range of mechanisms that include altered cell permeability and modification of the binding site of aminoglycoside to ribosome (Magnet and Blanchard, 2005). Gram-negative bacteria acquire resistance to most clinically useful aminoglycosides by methylation of 16S rRNA, which is accomplished by *S*-adenosyl-L-methionine (SAM)-dependent rRNA resistance methyltransferases (MTs) (Lioy *et al.*, 2014). This leads to a decrease or loss of affinity for antibiotic. Resistance to aminoglycoside antibiotics as a result of methylation of 16S rRNA has been observed in a wide range of gram-negative organisms (Liou *et al.*, 2006). Savic *et al.* (2009) purified methyl-free and in vivo methylated 16S rRNA from *E. coli* 30S ribosomal subunits; these were digested with specific RNase, and the digests were analyzed using MALDI-TOF MS. These authors observed differences in spectra for two sets of samples that were attributed to the non-methylated and methylated forms of RNA. Although this is a valuable diagnostic approach to detect methyltransferase resistance, the procedure may not be suitable for routine operation in a diagnostic laboratory.

MALDI-TOF MS has been utilized for the detection of β-lactamase itself from strains of ampicillin-resistant *E. coli* (Camara and Hays, 2007). These authors extracted the enzyme from broths that had and had not been supplemented with ampicillin. Analysis of the extracts using MALDI-TOF MS showed that the supplemented broth contained a peak at approximately 29 kDa. Separation of bacterial cell extracts using SDS-PAGE, followed by in-gel digestion with trypsin and analysis of the tryptic digests using liquid chromatography with mass spectrometric detection, confirmed a peak at ~29 kDa in the supplemented broth that can be attributed to β-lactamase. In a study carried out in our laboratory, *E. coli* was grown in broth supplemented with ampicillin. A sample of the broth was centrifuged, and the resulting pellet was treated with 70% formic acid and acetonitrile. The cell extract was separated using SDS PAGE electrophoresis. A band at around 29 kDa was detected. The band was excised and de-stained by repeated washing with mixture of acetonitrile and water (50:50, v/v). Following rehydration, the gel was

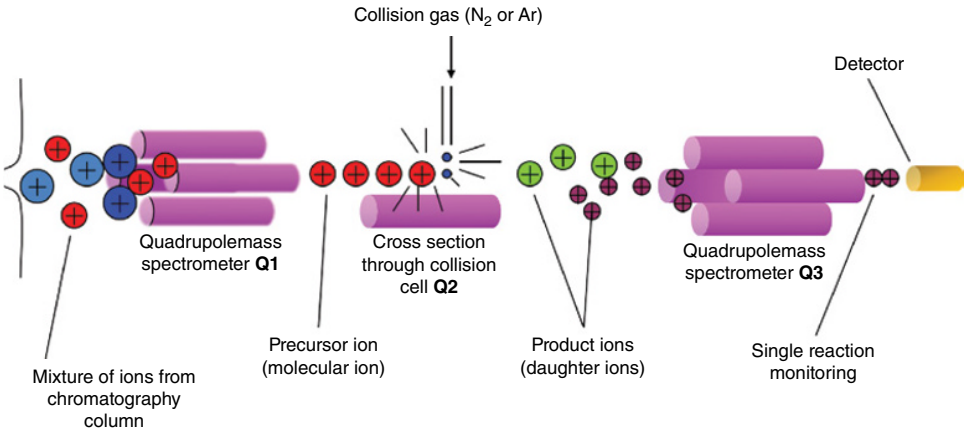


Figure 15.2 Schematic representation of selection and fragmentation of molecular ion, selection and detection of daughter ion with the use of triple quadrupole mass spectrometer.

treated with DTT, followed by iodoactemadie and then incubated with trypsin for 12 h. MALDI-TOF MS analysis of the tryptic digest and subsequent analysis of the peptide mass fingerprint (Figure 15.2) using Mascot revealed the protein to be β-lactamase. There is very little report on the use of this type of direct approach. It is low throughput and may not be sensitive enough to detect a low level of expression.

15.4 Liquid Chromatography Coupled with MS

The application of liquid chromatography (LC) with mass spectrometry (MS) for analysis of small to macromolecules has become a routine technique in many clinical laboratories. The combination of a reverse-phase liquid chromatography (RPLC) which offers high selectivity and efficiency with mass spectrometric detection which gives additional selectivity and structural information provides a powerful means of analysis. Sensitivity is another benefit of MS, and this allows analytes to be measured at picomoles to attomole levels. MS has also used in combination with other separation techniques like capillary electrophoresis (CE). Techniques like CE provide an alternative mechanism of separation and a significantly lower amount of sample that is required for analysis. In CE of small molecules, a high electric field is used to resolve analytes on the basis of differences in charge or hydrophobic properties. Although CE offers very high efficiencies, the low concentration sensitivity and issues with reproducibility of migration times have limited its use. When surveying the literature, it can be seen that the main method used for ionizing molecules in MS is electrospray ionization (ESI). This is a 'soft' ionization technique, which is preferred as there is minimal fragmentation of the molecule of interest. However, there are numerous other ionization methods that are used, which include atmospheric pressure chemical ionization (APCI) and atmospheric photoionization (APPI). Analysis of small molecules, peptides and proteins using MALDI results in analytes with a single charge. In the case of proteins and peptides, it imparts a +1 charge (sometimes +2 or +3, as well). At the same time, this simplifies and complicates downstream analysis. It is relatively straightforward to identify a protein with a +1

charge. However, this also makes it more difficult to analyze intact proteins, as their large size pushes their *m/z* values up, which gives both poor resolution and mass accuracy. Ions with +1 charge are difficult to fragment, says Kelleher, thus making MALDI a tougher ionization source for use in tandem MS applications such as post-translational modification analysis and peptide sequencing. In contrast, ESI produces a range of charged species for each molecule: +2, +3, +4 and so on. That large set of ions produced complicates mass analysis but makes it easier to perform tandem mass spectrometric analysis. Furthermore, because ESI is compatible with the chromatographic separations, it is often used for analysis of a wide range of compounds using ultrahigh pressure liquid chromatography (UPLC) to nano LC. Also, ESI can be used with a wide range of mass analysers. In addition, MALDI-TOF MS is not well suited for analysis of molecules in the lower mass range due to interference from fragment ions and matrix ions. Although MALDI-TOF MS has been used to detect carbapenemase activity in positive blood cultures, an enrichment step may be required to detect the enzyme in other samples such as swabs, urine cultures and wound specimens (Sparbier *et al.*, 2012). Furthermore, it cannot be used for quantitative analysis. In contrast, LC-MS remains the gold standard for both detection and quantitation of small molecules.

A study to evaluate MALDI-TOF MS and LC-MS for detection of carapenemase activity among SPM-1, GIM-1 and GES-5 producing *Pseudomonas aeruginosa*; OXA-143, IMP-10 and OXA-58 producing *Acinetobacter baumannii;* and KPC-2, NDM-1 ESBL/ΔOmppK36 producing *Klebsiella pneumonia* was reported by Carvalhaes *et al.* (2014). In their study, fresh bacterial cultures were incubated with two different concentrations of ertapenem in buffer (20 mM Tris-HCl, pH 6.8) for 2 and 4 h. Following centrifugation, an aliquot (1 µl) were analyzed using a Microflex LT instrument operated in linear positive ion mode and an LC-MS equipped with electrospray ionisation probe operated in positive ion mode. MALDI-TOF MS spectra of organisms that were sensitive to ertapenem showed no peak for molecular ion (*m/z* 475) and monosodium salt (*m/z* 497). The same interpretative criteria were used for LC-MS results. The authors demonstrated that class A (KPC-2, GES-5) and class B (SPM-1, IMP-1, IMP-10 and GIM-1) carbapenemase activity could be detected. A longer incubation period of 4 h was required for detection of class D carbapenemase activity. The results obtained using MALDI-TOF MS and LC-MS were in good agreement, and the former provided an easy, rapid and cheap method for detection of carbapenemase activity. One key advantage of using a single quadrupole instrument over MALDI-TOF MS is that it can be used to quantify the carbapenem and therefore the level of enzyme activity. In quadrupole mass analysers, an electric field is utilized to separate the ions according to their mass-to-charge ratio as they travel along a central axis of four parallel rods that have fixed and alternating voltage applied to them. The magnitude of this voltage can be set such that only ions of a certain *m/z* can pass through to the detector, while other ions are deflected into trajectories which cause them to collide with the rods and then pass out of the analyzer. This type of spectral analysis, called selected ion monitoring (SIM), gives excellent sensitivity. A quadrupole instrument can also be used to scan over a mass range. This is useful for both tuning the instrument and to record a total ion chromatogram. The sensitivity is significantly enhanced by monitoring the abundance of a single or few ions as a function of time. Inclusion of a stable isotope as an internal standard can be used to account for any errors in sample preparation and analysis and matrix effects.

A number of methods for detecting carbapenemase activity have been reported using high performance liquid chromatography (HPLC) in combination with tandem MS such as a triple quadrupole instrument. This is made up a linear series of three quadrupoles (Figure 15.3). In this type of analyzer, the molecular ion is selected for tandem MS in the first quadrupole in a similar way to SIM. The molecular ion is then activated in the collision cell by collision with a gas like nitrogen. This is the so-called collision-induced dissociation (CID), which results in a range of product ions. Of these one or more characteristic fragment ion(s) are selected to pass through the third quadrupole to the detector. This type of monitoring is called single reaction monitoring. There is no scanning, and thus high sensitivity is achieved. The quadrupole analyzers can be rapidly switched between several reactions, which can be fragments of a single molecular ion or products of several different compounds, to detect them with high sensitivity. This is known as multiple reaction monitoring (MRM). Like SIM, MRM is well suited to quantitative analysis of compounds. One major advantage of MRM over SIM is that the selectivity is enhanced, and this can lead to a significant improvement in sensitivity. Usually two fragment ions are selected for detection, which gives greater confidence in the results. Other combinations of mass analyzers such as quadrupole with TOF and ion trap with TOF are available, but these are more suited to qualitative analysis. Grundt and co-workers (2012) reported a rapid RPLC-MS/MS method for fast detection of ampicillin resistance in *E. coli*. Their method is based on incubating bacteria with ampicillin and quantifying ampicillin and ampicillin penicilloic acid within the supernatant following RPLC and CID. The transition used for ampicillin and ampicillin penicilloic acid were 350 to 160 *m/z* and 368 to 324 *m/z*, respectively. The HPLC-peak area ratio of ampicillin to ampicillin penicilloic acid was used to differentiate between ampicillin-resistant and susceptible strains. The authors reported that the resistance of *E. coli* to ampicillin can be determined within 90 min. This type of analysis can be utilized for detection of intact and hydrolysis product of any antibiotic providing that the molecular ion can be ionized. The rise in carbapenemase-producing bacteria poses a major threat in the management of antibiotic therapies for patients. Bacterial resistance to carbapenems is usually carried out using susceptibility and phenotypic methods like RAPIDEC®CARBANP test (Poirel and Nordmann, 2015) and Hodge test (Kim *et al.*, 2015). The RAPIDEC®CARBANP test is based on detecting a change in pH of the incubation medium and may be poor with respect to both sensitivity and specificity. Methods based on PCR (Naas *et al.*, 2013) are thought to be the gold standard for detection of carbapenemases. However, the range of genes coding for this enzyme means that a PCR method targeting each gene must be developed. Recently, LC-MS/MS methods have been reported for the detection of carbapenemase activity. Wang *et al.* (2013) used nano LC with a hybrid triple quadrupole/linear ion trap for detecting hydrolysis of meropenem by bacteria expressing carbapenemase activity. In their work, bacteria were grown in Luria Broth medium containing 50 $\mu g ml^{-1}$ meropenem. Following incubation at 37 °C for 2 h, the broth was centrifuged, and an aliquot of the supernatant was diluted, mixed with internal standard and analyzed using the nano-LC-MS/MS. Multiple fragment ions were chosen for single reaction monitoring (SRM), and these were selected on the basis of ion intensity, stability and matrix effects. For meropenem, precursor/product ion transitions were 402/122, 402/175, 402/216, 402/220 and 402/358, whereas the corresponding pairs in ^{18}O-labelled meropenem were 404/122, 404/175, 404/218, 404/222 and 404/360. The concentration of unlabelled meropenem hydrolyzed was

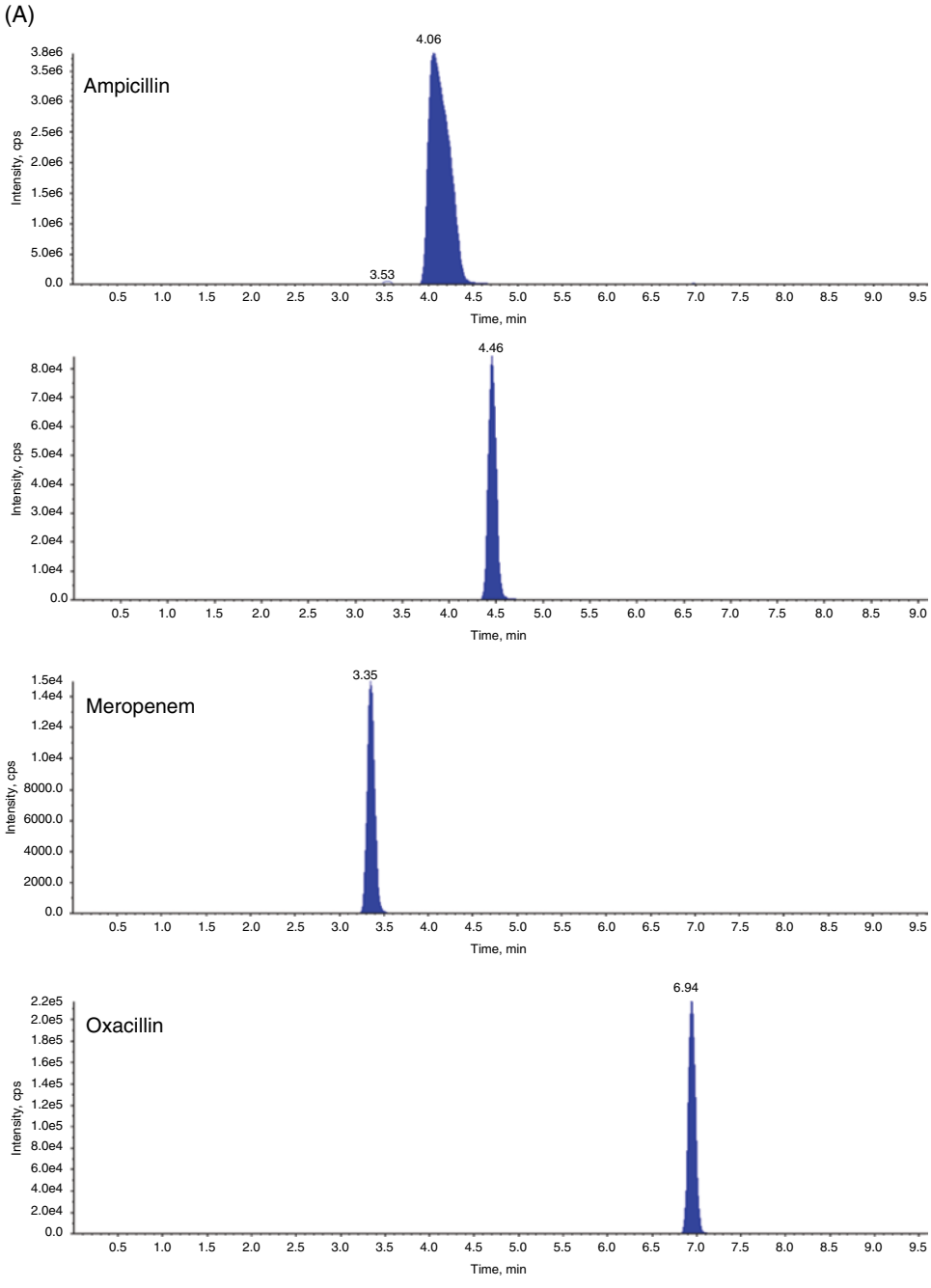


Figure 15.3 MRM chromatograms of mixture of ampicillin (10 mg/ml), cefotaxime (0.5 mg/ml), meropenem (0.5 mg/ml) and oxacillin (0.5 mg/ml) incubated at 37°C for 2 h (A) in phosphate buffer (B) after incubation with cell lysate of *P. aeruginosa*. The samples were diluted 1:22 times prior to injection. Chromatograms were acquired using an API3000 triple quadrupole mass spectrometer fitted with an electrospray ionization source and operated in positive ion mode.

(B)

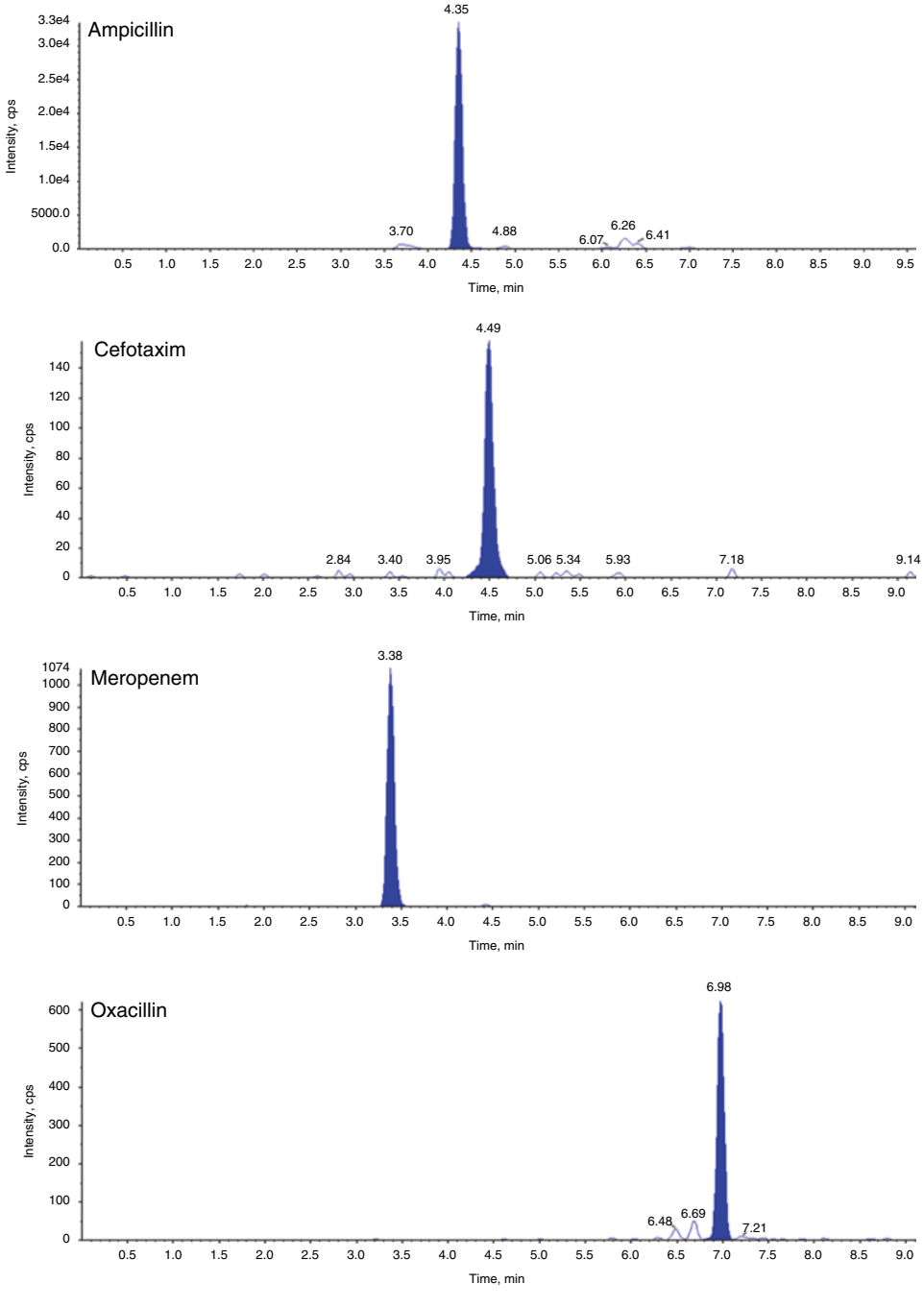


Figure 15.3 (Continued)

calculated using the peak area ratio of unlabelled versus the labelled internal standard. The concentration obtained from all five transitions was averaged. The authors showed that the concentration of meropenem after hydrolysis could be directly correlated with carbapenemase activity. Furthermore, the SRM/MRM capability of triple quadrupole instruments means that the method can be adopted to detect different β-lactamases in a single run. Thus, antibiotic resistance can be measured rapidly and with a high level of sensitivity. A UPLC-MS/MS method was reported by Carricajo *et al.* (2014) for the detection of carbapenemase activity in *Enterobacteriaceae*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii.* The method here was based on incubating bacterial suspension with 5 mgml^{-1} of meropenem or ertapenem for 3 to 4 h at 36 °C. Following centrifugation and addition of an internal standard (meropenem d6), the concentration of ertapenem and meropenem were measured using a tandem quadrupole instrument operated in positive ion mode. The transitions used for ertapenem, meropenem and the internal standard were 476 to 114 *m/z*, 383.9 to 254.1 *m/z* and 390.1 to 147.1 *m/z*, respectively. The rate of hydrolysis was determined by dividing the carbapenem concentration after incubation by the pre-incubation concentration of carbapenem. The study showed that hydrolysis of carbapenem occurred within 3 h for NDM-1 and KPC-producing enterobacteria, whereas a longer incubation period of 4 h was required for some CHDL-, VIM- and OXA-48-producing strains. The authors found that unlike molecular methods, the UPLC-MS/MS procedure did not allow identification of the carbapenemase enzyme and therefore failed to give molecular epidemiological information. The authors also concluded that initial cost of implementing a UPLC-MS/MS system is high in comparison to PCR. However, laboratories where this type of technology is already available can easily adopt this type of methodology. Furthermore, carbapenemase activity can be detected irrespective of the enzyme type. Kulkarnni *et al.* (2014) reported a rapid and sensitive LC-MS/MS method for detection of carbapenemase activity for *Enterobacteriaceae and* non-*Enterobacteriaceae*-expressing IMP, VIM, KPC, NDM and or OXA using ertapenem, imipenem and meropenem. In their assay, ertapenem or imipenem or meropenem was added to a final concentration of 4 μgml^{-1} to a bacterial suspension and the mixture incubated at 37 °C for 1 h. Following protein precipitation and centrifugation, an aliquot of the supernatant was diluted with deionized water for LC-MS/MS analysis in MRM mode. The MRM transitions used for detection of meropenem and imipenem were 384.1 to 68 *m/z* and 300.1 to 142.1 *m/z*, respectively. The primary hydrolysis products of meropenem and imipenem were detected using product/precursor ion transition of 402 to 358 *m/z* and 318.1 to 103 *m/z*, respectively. The MRM transitions for ertapenem and its hydrolysis product were not reported. A ratio of peak area of the intact drug to hydrolyzed product was used to determine carabapenemase activity. The authors found the hydrolysis ratio ranged from 0.00 to >100 for meropenem and from 0.02 to >100 for ertapenem and imipenem. The authors investigated the effects of matrix on ion suppression using a blank matrix. Although no matrix effects were observed using this approach, for the method to be more widely accepted it would be necessary to add a stable isotope of the carbapenem to every sample just prior to injection onto the HPLC column because ion suppression may vary from sample to sample. We recently developed a rapid LC-MS/MS platform to measure the hydrolyzing capability of microorganisms towards several β-lactams in a single analysis. In this assay, bacteria were cultured in Columbia blood agar and a 1.0 ml suspension prepared. An aliquot (100 μl) was used to determine the protein

concentration using the Bradford method. The remainder of the culture (100 µl) was incubated with 10 µl of a mixture of amoxicillin (0.25 mg/ml), cefotaxime (0.5 mg/ml), meropenem (0.5 mg/ml) and oxacillin (0.5 mg/ml) at 37 °C for 2 h. Two-hundred microlitres of cold acetonitrile was then added to precipitate proteins. The resulting precipitate was separated by centrifugation at 15,000 g for 5 min. The supernatant was carefully transferred to an HPLC vial and mixed with 300 µl of water and an aliquot was loaded onto a 5 × 2.1 mm i.d. C18 2.7 µm Ascentis Express column. The analytes were separated using a binary gradient elution profile composed of 'A'-0.1% formic acid in water and 'B'-0.1% formic acid acetonitrile. The elution profile was as follows: 0–10 min 5% to 75% B, down to 5% B 10–11 min and hold 11–15 min. Eluates were detected using an API3000 triple quadrupole mass spectrometer operated in positive ion mode.

The mass spectrometric parameters were optimized to generate the maximum level of the protonated form of β-lactams. The source-dependent parameters for the analytes consisted of collision gas, curtain gas, ion spray gas 1 and 2, ionspray voltage and the temperature of the heater gas, with optimum values of 4, 10, 60, 10, 4.5 kV and 450 °C, respectively. The analyte-dependent parameters were also tuned to obtain the maximum detector response. The ion spray voltage was set at 4.5 kV and the source temperature at 450 °C. The mass spectrometer was operated at unit mass resolution for both Q1 and Q3 in molecular reaction monitoring (MRM) mode. The precursor-to-product transitions of m/z 366 → 114 for amoxicillin, m/z 456 → 124 for cefotaxime, m/z 384 → 68 for meropenem and m/z 424 → 182 for oxacillin were monitored. Figure 15.3 shows MRM chromatograms of a control, mixture of β-lactams prior to incubation with bacterial culture and following 2 h incubation. The difference between the pre-incubation and post-incubation concentrations of β-lactam together with the protein concentration was used to express β-lactamase activity per minute per milligram of protein. The control served to account for any spontaneous hydrolysis that may occur.

The use of MALDI-TOF MS for detection of β-lactams and their hydrolysis products is unexpected given the interference that can occur from matrix ions. In comparison, LC with mass spectrometric detection is ideally suited for selective detection of small molecules like β-lactams in complex matrices. It has been reported that to achieve the level of sensitivity for carabapenemase activity achieved using LC with mass spectrometry with MALDI-TOF MS, a much higher bacterial inocula and β-lactam concentration in simple matrix is required. However, it has been shown that MALDI-TOF can be used effectively for detection of β-lactamase activity in isolated colonies. Detection of β-lactamases like carbapenemases using mass-spectrometry-based methods may not suitable for detection of other mechanisms used by bacteria for resistance to antibiotics, and it may be necessary to resort to traditional phenotypic antimicrobial susceptibility testing to account for all the different mechanisms of resistance.

15.5 Proteomics Approaches for Detection of Antibiotic Resistance

The MS approaches described thus far are based on detection of β-lactamase activity and therefore limited to detection of one type of antibiotic-resistant mechanism. A more global approach to studying antibiotic-resistant mechanisms in bacteria is use of comparative proteomics. This kind of strategy provides a means of understanding the

mechanisms of antibiotic resistance in more detail. Furthermore, it offers a means of identifying new mechanisms of resistance and may lead to uncovering secondary mutations that are associated with resistant phenotypes. In these approaches, bacteria are cultured under identical conditions, the cells harvested and then lysed and proteins are extracted and then separated using 2D-GE. Visual comparison of the gel is used to identify differences in protein profiles between the gels. Bands of interest are excised and subjected to reduction, alkylation followed by enzymatic digestion. The resulting peptides are extracted and analyzed using either MALDI-TOF MS or LC-MS/MS to identify the protein. The complexity of the protein profile can be reduced by using a combination of cellular pre-fractionation to determine cellular location and pathway predictions. This strategy was used to study penicillin resistance in *Streptococcus pneumonia* (Soualhine *et al.*, 2005). Both resistant and susceptible strains were studied. Following initial in vitro studies, strains that were penicillin resistant and susceptible were chosen. Comparative proteomics of these strains showed a number of differentially expressed proteins that included a PstS, which is a subunit of phosphate ABC transporter. The level of this protein was higher in resistant strains of *S. pneumonia*, and this was in parallel with RNA expression of the entire ABC transporter operon. Inactivation of the *pstS* gene led to greater susceptibility to penicillin in the wild-type strain.

In another study conducted by Hu *et al.* (2007), a reduction in susceptibility to ceftriaxone in *Salmonella enterica* serovar Typhimurium strain was associated with transposon insertion in the *yjeH* gene, a putative transporter gene. Proteomic analysis of outer membrane proteins of the mutant and the parent R200 showed that expression of porin OmpD was raised and putative outer membrane proteins STM1530, STM3031, heat shock protein MopA and a subunit of the proton pumping oxidoreductase NuoB was reduced. The authors suggested that these proteins are associated with ceftriaxone resistance, which is influenced by the change in expression of the putative transporter gene *yjeH*.

The occurrence of organisms that are resistant to third-generation β-lactams is becoming a major health issue. In relation to this, Liu *et al.* (2012) identified a *Stenotrophomonas maltophilia* DCPS-01 that was resistant to all β-lactams. Proteomic analysis of this microorganism demonstrated that in the presence of imipenem, DCPS-01 differentially expressed several proteins in comparison to their control strain, K279a. This included L1 MBL, stress proteins, elongation-related proteins and proteins that are associated with metabolism. Van Oudenhove *et al.* (2012) conducted a proteomic study in the same organism using 2D GE of cytoplasmic and membrane protein followed by Isobaric Tags for Relative and Absolute Quantitation ITRAQ® differential labelling and 2-D LC with tandem MS. The results showed that expression of 73 proteins was altered when the organism was challenged with imipenem. In addition to a rise in production of β-lactamase, an increase in other metabolic enzymes was also observed. The study showed the adaptation of the organism following exposure to imipenem. The inner membrane proteome of the carbapenemase strain of *Acinetobacter baumannii* was analyzed by Tiwari *et al.* (2012) using differential in-gel electrophoresis (DIGE) followed by LC-MS/MS analysis. The authors found 19 overexpressed and four down-regulated proteins in the resistant strain compared to the reference strain. Some of the up-regulated proteins were Amp-C and OXA-51, which metabolize carbapenems, and other proteins were enzymes involved in metabolism such as ATP synthase, malate dehydrogenase, which increases energy production required for survival.

Glycopeptide antibiotics like teicoplanin and vancomycin are widely used in treatment of infections caused by gram-positive bacteria like staphylococci. Vancomycin antibacterial mechanism of action is by way of preventing cross-linking of the peptidoglycan cell wall. Organisms that are resistant to vancomycin synthesize altered cell wall precursors which have a reduced affinity for this glycopeptide antibiotic. Pieper *et al.* (2006) carried out comparative proteomic studies of a subcellular fraction of the *S. aureus* isolate VP32. The authors used 2-DE and analyzed the separated proteins using MALDI-TOF/TOF MS and LC-MS/MS and found 65 proteins that were differentially expressed. Several enzymes involved in the biosynthesis of purines were up-regulated in VP32 relative to P100 and HIP5827. Furthermore, other proteins like peptidoglycan hydrolase and penicillin binding protein were significantly elevated in VP32. These proteins are involved in the biosynthesis of the cell wall peptidoglycan. The authors suggest that the altered level of these proteins may be responsible for a change in the cell wall turnover rate and an altered peptidoglycan structure. It has been shown that the vancomycin-intermediate *S. aureus* (VISA) is formed when this glycopeptide is used to treat infections caused by methicillin-resistant *S. aureus* (MRSA). To identify biomarkers of resistance to the glycopeptides, Drummelsmith *et al.* (2007) carried out comparative proteomics studies. The authors used high-resolution 2 DE with iTRAQ® mass tagging to identify differentially expressed proteins between clinical MRSA and VISA strains of the same multilocus sequence type (MLST). The results showed that 93 proteins were differentially expressed, and of these lytic transglycosylase SAV2095 was chosen for its higher abundance in all VISA strains relative to MRSA strains as well as several heterogeneous VISA strains that were found to be vancomycin sensitive using standard susceptibility testing methods. The occurrence of vancomycin resistance in enterococci has also been rising slowly and poses a serious health threat. Ramos *et al.* (2015) carried out proteomics studies to understand the mechanism of resistance to glycopeptide in *Enterococcus faecium* SU18 strain treated with and without vancomycin. The results showed that 14 proteins were differentially expressed in SU18: half were up-regulated and the other half down-regulated. The expression of proteins that are involved in vancomycin resistance, like VanA protein, VanA ligase, VanR and D-Ala-D-Ala dipeptidase, was raised whereas those involved in metabolism were down-regulated. The authors proposed that the level of SAV2095 could be used as a biomarker for rapid detection of VISA strains.

The antimicrobial daptomycin is a cyclic lipopeptide that is used clinically in treatment of infections caused by both susceptible and resistant strains of *S. aureus* that include bacteraemia, endocarditis and skin and soft tissue infections. Daptomycin works by inserting itself into the cell membrane and forming aggregates. This leads to damage of the membrane, which results in leakage of ions and depolarization. These events lead to inhibition of protein synthesis and result in bacterial cell death. Fischer *et al.* (2011) conducted comparative transcriptomics and proteomics studies of isogenic daptomycin-resistant (DAP^R) and daptomycin-sensitive strains (DAP^S) obtained from a patient with relapsing endocarditis during daptomycin treatment. The authors found that were significant differences in proteins expressed that included cell wall and biofilm forming proteins.

Tetracyclines are used for treatment of a wide variety of infections because they have a broad spectrum of activity and do not cause major side effects. The rise in identification of tetracycline-resistant organisms has resulted in a fall in clinical use of these antimicrobial agents. Several resistant mechanisms to tetracycline have been identified,

and these include efflux pumps, formation of ribosomal protection proteins and modification of the antibiotic to render it inactive (Speer *et al.*, 1992). Lee *et al.* (2014) carried out proteomic analysis using nano LC with a linear quadrupole ion trap Fourier transform mass spectrometer of *Acinetobacter baumannii* DU202 isolated from a sputum sample from a patient admitted to hospital. High level of expression of β-lactamases, a multidrug resistance efflux pump and resistance–nodulation–cell division (RND) multidrug efflux proteins was found in this organism. Peng *et al.* (2005) conducted a proteomic study of sarcosine-insoluble outer membrane fraction of *P. aeruginosa* resistant to ampicillin, kanamycin and tetracycline using 2-DE and MALDI-TOF MS. In addition to the known proteins associated with antibiotic resistance, the authors also found five new proteins. The authors concluded that the results of their work will provide a better understanding of antibiotic resistance between different species of bacteria.

A year later, Xu *et al.* (2006) carried out proteomics studies of the outer membrane proteins of ampicillin- and tetracycline-resistant strains of *E. coli* K-12 using 2-DE and MALDI-TOF MS. The organism was cultured in the presence of ampicillin or tetracycline at MIC and their respective controls. The authors found three proteins that are known to be associated with antibiotic-resistant (TolC, OmpC and YhiU) together with six new antibiotic-resistance-related proteins (FimD precursor, LamB, Tsx, YfiO, OmpW and NlpB). Comparative proteomics was undertaken of a *Coxiella burnetii* reference strain grown under tetracycline stress conditions and compared against a control investigated using fractional diagonal chromatography (COFRADIC) with MS. This approach was developed by Gevaert *et al.* (2005), who state that it overcomes the major limitation of peptide-centric proteomic approach, where an enormous number of peptides are analyzed using tandem MS. Furthermore, the limited capacity of current mass spectrometers is insufficient for a large number of peptides to be identified. In comparison, COFRADIC is a gel-free method that isolates a set of preselected peptides for LC-MS/MS analysis. Methods for isolating cysteinyl, methionyl, amino terminal and phosphorylated peptides have been reported. Using this strategy, Vranakis and coworkers (2012) in their work found five proteins were up-regulated that are involved in biosynthesis and metabolism. Of the proteins identified, 19 were down-regulated.

Aminoglycosides are potent broad-spectrum antibiotics that bring about their action by inhibiting protein synthesis. A number of different resistance mechanisms have been described for aminoglycosides that include efflux pumps and drug-modifying enzymes (Mingeot-Leclercq *et al.*, 1999). In a proteomic study by Sharma *et al.* (2010), whole cell extracts from *Mycobacterium tuberculosis* clinical isolates susceptible and resistant to streptomycin using 2DE and MALDI-TOF MS nine protein were overexpressed in SM-resistant isolates in comparison to sensitive isolates. These proteins were identified as DnaK, 60 kDa chaperonin2, malate dehydrogenase, oxidoreductase, electron transfer flavoprotein subunit alpha, antigen 84, 14 kDa antigen and two hypothetical proteins. Streptomycin is known to interact with amino acid residues in the active site of malate dehydrogenase, and this may be the reason why the level of this enzyme is altered.

Chloramphenicol is a broad-spectrum antibiotic that was first derived from the bacterium *Streptomyces venezuelae.* It is used to treat a range of infections that include meningitis, cholera and typhoid fever. Resistance to chloroamphenicol is due to reduced membrane permeability, mutation of the 50S ribosomal subunit and acetylation of the antibiotic by chloroamphenicol acetyltransferase. Biot *et al.* (2011) studied *Burkholderia thailandensis* using SDS-PAGE and nano LC with tandem MS. The authors showed that

in the chloroamphenicol-induced resistant strains, two different efflux pumps were overexpressed. These efflux pumps were able to expel several classes of antibiotics that included fluoroquinolones and tetracyclines. Resistance to fluoroquinolones has also been studies using proteomic-based approaches. This group belong to the quinolone family of antibiotics. These are broad-spectrum antibiotics that produce their antibacterial action by inhibiting enzymes involved in DNA replication, namely, DNA gyrase and DNA topoisomerase IV. Coldham *et al.* (2006) conducted a proteomic study of *Salmonella enterica* serovar Typhimurium grown in the presence of the fluoroquinolone antibiotics ciprofloxacin and enrofloxacin. Proteins were separated using 2D-GE, and following in-gel digests of protein spots, 2D-LC strong cation exchange with reverse phase was used to separate the tryptic digests. The authors showed that basal expression of the AcrAB/TolC efflux pump was raised in the MAR mutant compared with the untreated wild type. In addition, the expression of 43 other proteins was elevated, which was attributed to the physiological response to the fluoroquinolones.

Recently, Rees *et al.* (2015) reported a method that exploits the specificity and physiology of the staphylococci bacteriophage K to identify *Staphylococcus aureus* and determine its susceptibility to clindamycin and cefoxitin. The method utilized MS to monitor the replication of the bacteriophage prior to infecting samples thought to contain *S. aureus*. Peptides of a 51 kDal capsid protein of bacteriophage K were used as biomarkers to search for resistance in samples containing *S. aureus*, the rationale being that the bacteriophage K will amplify in the presence of a suitable host. If bacteriophage amplification is detected in samples containing the antibiotics clindamycin or cefoxitin, the sample is considered to be resistant to these antibiotics because this will only occur in a viable host. Consequently, the method allows the susceptibility to clindamycin and cefoxitin in *S. aureus* to be determined in a single protocol.

Proteomic approaches show that antibiotic resistance mechanisms are more complex than originally thought, and reliance solely on genome sequence data will not provide a definitive method for their detection. Proteomic signatures have revealed a significant number of expressed proteins that are involved in a variety of metabolic processes involved in antimicrobial resistance. However, proteogenomic approaches enable potential genes to be mapped initially, from which putative proteins may be detected by MS/MS-based methods. Chong *et al.* (2014) demonstrated such an approach for clinical isolates of *Clostridium difficile* and complex antimicrobial resistance profiles that can occur within single species (see also Chapter 14).

15.6 Conclusion

It is evident from Table 15.2 that both MALDI-TOF MS and LC-MS are comparable with respect to resolving power, mass accuracy and mass range. Note that the *m/z* range shown for MALDI-TOF is valid only for the linear mode and is restricted to around 100,000 or less for the reflectron mode. Although the resolution indicated for both techniques is similar, for MALDI-TOF MS a significantly higher resolution is obtained in reflectron mode. One of the key advantages of LC-MS over MALDI-TOF MS is that it can be used for quantitative analysis with a high level of sensitivity and reproducibility. Furthermore, analysis of small molecules, like antibiotics and their hydrolysis products, using MALDI-TOF MS can be prone to interference from matrix ions.

Table 15.2 Comparison of MALDI-TOF MS and LC-MS.

	MALDI-TOF	LC-MS
Resolution[a]	0.1–0.0002	1–0.0005
Resolving[a] **power**	15,000–2,500,000	7,500–2,500,000
Mass accuracy#	<0.6 to <100	<0.6 to 5
***m/z* range**	15–500,000	2–100,000
Quantitative	Semi-quantitative	Yes
Sensitivity	Good	nM to sub-fM
Reproducibility (from sample to sample)	Good	Excellent

[a] Information on diarylquinolines is based on technical specifications provided by manufacturers. If manufacturers have more than one instrument in a series, then only the information from the best-performing instrument has been used.

References

Barlam, T., and Neu, H. C. (1984). Pyridinium 2-azo-p-dimethylaniline chromophore, a chromogenic reagent for β-lactamase testing compared to nitrocefin. *European Journal of Clinical Microbiology* 3, 185–189.

Biot, F. V., Valade, E., Garnotel, E., Chevalier, J., Villard, C., Thibault, F. M., Vidal, D. R., and Pagès, J. M. (2011). Involvement of the efflux pumps in chloramphenicol selected strains of *Burkholderia thailandensis*: proteomic and mechanistic evidence. *PLoS ONE* 2011 Feb 9, 6(2), e16892.

Blair J. M., Richmond, G. E., and Piddock, L. J. (2014). Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. *Future Microbiol* 9(10), 1165–1177.

Bush, K. (1999). Beta-Lactamases of increasing clinical importance. *Curr Pharm Des* 1999 Nov, 5(11), 839–845.

Bush, K., and Jacoby, G. A. (2010). Updated functional classification of beta-lactamases. *Antimicrob Agents Chemother* 54(3), 969–977.

Camara, J. E., and Hays, F. A. (2007). Discrimination between wild-type and ampicillin-resistant *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight MS. *Anal Bioanal Chem* 389, 1633–1638.

Carvalhaes, C. G., Cayô, R., Visconde, M. F., Barone, T., Frigatto, E. A., Okamoto, D., Assis, D. M., Juliano, L., Machado, A. M., and Gales, A. C. (2014). Detection of carbapenemase activity directly from blood culture vials using MALDI-TOF MS: A quick answer for the right decision. *J Antimicrob Chemother* 69(8), 2132–2136.

Carricajo, A., Verhoeven, P. O., Guezzou, S., Fonsale, N., and Aubert, G. (2014). Detection of carbapenemase-producing bacteria by using an ultra-performance liquid chromatography-tandem mass spectrometry method. *Antimicrob Agents Chemother* 58(2), 1231–1234.

Chong, P. M., Lynch, T., McCorrister, S., Kibsey, P., Miller, M., Gravel, D., Westmacott, G. R., and Mulvey, M. R. (2014) Proteomic analysis of a NAP1 *Clostridium difficile* clinical isolate resistant to metronidazole. *PLoS ONE* 9(1), 1–10.

Coldham, N. G., Randall, L. P., Piddock, L. J., and Woodward, M. J. (2006). Effect of fluoroquinolone exposure to the proteome of *Salmonella enterica* serovvat Typhinimurium. *J Antimicrob Chemother* 58, 1145–1153.

Drummelsmith, J., Winstall, E., Bergeron, M. G., Poirier, G. G., and Ouellette, M. (2007). Comparative proteomics analyses reveal a potential biomarker for the detection of vancomycin-intermediate *Staphylococcus aureus* strains. *J Proteome Res* 6(12), 4690–4702.

Fischer, A., Yang, S. J., Bayer, A. S., Vaezzadeh, A. R., Herzig, S., Stenz, L., Girard, M., Sakoulas, G., Scherl, A., Yeaman, M. R., Proctor, R. A., Schrenzel, J., and François, P. (2011). Daptomycin resistance mechanisms in clinically derived *Staphylococcus aureus* strains assessed by a combined transcriptomics and proteomics approach. *J Antimicrob Chemother* 66(8), 1696–1711.

Gevaert, K., Van Damme, P., Martens, L., and Vandekerckhove, J. (2005). Diagonal reverse-phase chromatography applications in peptide-centric proteomics: ahead of catalogue-omics? *Anal Biochem* 345(1), 18–29.

Girlich, D., Poirel, L., and Nordmann, P. (2012). Value of the modified Hodge test for detection of emerging carbapenemases in Enterobacteriaceae. *J Clin Microbiol* 50(2), 477–479.

Grundt, A., Findeisen, P., Miethke, T., Jäger, E., Ahmad-Nejad, P., and Neumaier, M. (2012). Rapid detection of ampicillin resistance in Escherichia coli by quantitative mass spectrometry. *J Clin Microbiol* 50(5), 1727–1729.

Hoff, G. P., van Kampen, J. J. A., Meester R. J. W., van Belkum A., Goessens W. H. F., and Luider, T. M. (2012). Characterization of β-Lactamase enzyme activity in bacterial lysates using MALDI-mass spectrometry. *J Proteome Res* 11(1), 79–84.

Hrabák, J., Walková, R., Studentová, V., Chudácková, E., and Bergerová, T. (2011). Carbapenemase activity detection by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J Clin Microbiol* 49(9), 3222–3227.

Hu, W. S., Lin, Y. H., and Shih, C. C. (2007). A proteomic approach to study *Salmonella enterica* serovar Typhimurium putative transporter YjeH associated with ceftriaxone resistance. *Biochem Biophys Res Commun* 361(3), 694–699.

Huttner, A., Harbarth, S., Carlet, J., Cosgrove, S., Goossens, H., Holmes, A., Jarlier V., Voss, A., and Pittet, D. (2013). Antimicrobial resistance: a global view from the 2013 World Healthcare-Associated Infections Forum. *Antimicrob Resist Infect Control*.18; 2: 31.

Kim, H. K., Park, J. S., Sung, H., and Kim, M. N. (2015). Further modification of the modified Hodge test for detecting Metallo-β-Lactamase-Producing Carbapenem-Resistant Enterobacteriaceae. *Ann Lab Med* 35(3), 298–305.

Kulkarni, M. V., Zurita, A. N., Pyka, J. S., Murray, T. S., Hodsdon, M. E., and Peaper, D. R. (2014). Use of imipenem to detect KPC, NDM, OXA, IMP, and VIM carbapenemase activity from gram-negative rods in 75 minutes using liquid chromatography-tandem mass spectrometry. *J Clin Microbiol* 52(7), 2500–2505.

Lange, C., Schurbert, S., Jung, J., Kostrzewz, M., and Sparbier, K. (2013). Quantitative matrix –assisted laser desorption ionization-time of flight mass spectrometry for rapid resistance detection. *J Clin Microbiol* 52(12), 4155–4162.

Lee, S. Y., Yun, S. H., Lee, Y. G., Choi, C. W., Leem, S. H., Park, E. C., Kim, G. H., Lee, J. C., and Kim, S. I. (2014). Proteogenomic characterization of antimicrobial resistance in extensively drug-resistant *Acinetobacter baumannii* DU202. *J Antimicrob Chemother* 69(6), 1483–1491.

Liou, G. F., Yoshizawa, S., Courvalin, P., and Galimand, M. (2006). Aminoglycoside resistance by ArmA-mediated ribosomal 16S methylation in human bacterial pathogens. *J Mol Biol* 359, 358–364.

Lioy, V. S., Goussard, S., Guerineau, V., Yoon, E.-J., Courvalin, P., Galimand, M., and Grillot-Courvalin, C. (2014). Aminoglycoside resistance 16S rRNA methyltransferases block endogenous methylation, affect translation efficiency and fitness of the host. *RNA* 20(3), 382–391.

Liu, W., Zou, D., Wang, X., Li, X., Zhu, L., Yin, Z., Yang, Z., Wei, X., Han, L., Wang, Y., Shao, C., Wang, S., He, X., Liu, D., Liu, F., Wang, J., Huang, L., and Yuan, J. (2012). Proteomic analysis of clinical isolate of *Stenotrophomonas maltophilia* with blaNDM-1, blaL1 and blaL2 β-lactamase genes under imipenem treatment. *J Proteome Res* 11(8), 4024–4033.

Magnet, S., and Blanchard, J. S. (2005). Molecular insights into aminoglycoside action and resistance. *Chem Rev* 105, 477–498.

Mingeot-Leclercq, M. P., Glupczynski, Y., and Tulkens, P. M. (1999). Aminoglycosides:activity and resistance. *Antimicrob Agents Chemother* 43, 272–237.

Naas, T., Cotellon, G,, Ergani, A., and Nordmann, P. (2013). Real-time PCR for detection of blaOXA-48 genes from stools. *J Antimicrob Chemother* 68(1), 101–104.

Nordmann, P., Nass, T., and Poirel, L. (2011). Global Spread of Carbapenemase-producing Enterobacteriaceae. *Emerging Infect Dis* 17, 1791–1798.

Peng, X., Xu, C., Ren, H., Lin, X., Wu, L., and Wang, S. (2005). Proteomic analysis of the sarcosine-insoluble outer membrane fraction of *Pseudomonas aeruginosa* responding to ampicilin, kanamycin, and tetracycline resistance. *J Proteome Res* 4(6), 2257–2265.

Pieper, R., Gatlin-Bunai, C. L., Mongodin, E. F., Parmar, P. P., Huang, S. T., Clark, D. J., Fleischmann, R. D., Gill, S. R., and Peterson. S. N. (2006). Comparative proteomic analysis of *Staphylococcus aureus* strains with differences in resistance to the cell wall-targeting antibiotic vancomycin. *Proteomics* 6(15), 4246–4258.

Poirel, L., Nordmann, P. (2015). RAPIDEC® CARBA NP test for rapid detection of carbapenemase producers. *J Clin Microbiol* 53(9): 3003–3008.

Ramos, S., Chafsey, I., Silva, N., Hébraud, M., Santos, H., Capelo-Martinez, J. L., Poeta, P., and Igrejas, G. (2015). Effect of vancomycin on the proteome of the multiresistant *Enterococcus faecium SU18 strain*. *J Proteomics* 113, 378–387.

Rees, J. C., Pierce, C. L., Schieltz, D. M., and Barr, J. R. (2015). Simultaneous identification and susceptibility determination to multiple antibiotics of *Staphylococcus aureus* by bacteriophage amplification detection combined with mass spectrometry. *Anal Chem* 87(13), 6769–6777.

Sauget, M., Cabrolier, N., Manzoni, M., Bertrand, X., and Hocquet, D. (2014). Rapid, sensitive and specific detection of OXA-48-like-producing Enterobacteriaceae by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Microbiol Methods* 105, 88–91.

Savic, M., Lovric, J., Tomic, T. I., Vasiljevic, B., and Conn, G. L. (2009). Determination of the target nucleosides for members of two families of 16S rRNA methyltransferases that confer resistance to partially overlapping groups of aminoglycoside antibiotics. *Nucleic Acids Res* 37, 5420–5431.

Shakil, S., Khan, R., Zarrilli, R., and Khan, A. U. (2008). Aminoglycosides versus bacteria a description of the action, resistance mechanism, and nosocomial battleground. *J Biomed Sci* 15, 5–14.

Sharma, P., Kumar, B., Gupta, P., Singhal, N., Katoch, V. M., Venkatesan, K., and Bisht, D. (2010). Proteomics analysis of streptomycin resistant and sensitive clinical isolates of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 8, 59.

Smith, C. A., and Baker, E. N. (2002). Aminoglycoside antibiotic resistance by enzymatic deactivation. *Curr. Drug Targets Infect. Disord.* 2, 143–160.

Soualhine, H., Brochu, V., Ménard, F., Papadopoulou, B., Weiss, K., Bergeron, M. G., Légaré, D., Drummelsmith, J., and Ouellette, M. (2005). A proteomic analysis of penicillin resistance in *Streptococcus pneumoniae* reveals a novel role for PstS, a subunit of the phosphate ABC transporter. *Mol Microbiol* 58(5), 1430–1440.

Sparbier, K., Schubert, S., Weller, U., Boogen, C., and Kostrzewa, M. (2012). Matrix-assisted laser desorption ionisation- time of flight mass spectrometry-based functional assay for rapid detection of resistance against β-lactam antibiotics. *J Clin Microbiol* 50(3), 927–937.

Speer, B. S., Shoemaker, N. B., and Salyers, A. A. (1992). Bacterial resistance to tetracycline: mechanisms, transfer, and clinical significance. *Clin Microbiol Rev* 5(4), 387–399.

Sun, S., Selmer, M., and Andersson, D. I. (2014). Resistance to β-lactam antibiotics conferred by point mutations in penicillin-binding proteins PBP3, PBP4 and PBP6 in Salmonella enterica. *PLoS ONE* 2014 May 8, 9(5), e97202.

Sutterlin, H. A., Zhang, S., and Silhavy, T. J. (2014). Accumulation of phosphatidic acid increases vancomycin resistance in *Escherichia coli*. *J Bacteriol* 196(18), 3214–3220.

Tiwari, V., Vashistt, J., Kapil, A., and Moganty, R. R. (2012). Comparative proteomics of inner membrane fraction from carbapenem-resistant *Acinetobacter baumannii* with a reference strain. *PLoS ONE* 7(6), e39451.

Van Oudenhove, L., De Vriendt. K., Van Beeumen, J., Mercuri, P. S., and Devreese, B. (2012). Differential proteomic analysis of the response of *Stenotrophomonas maltophilia* to imipenem. *Appl Microbiol Biotechnol* 95(3), 717–733.

Vranakis, I., De Bock, P. J., Papadioti, A., Tselentis, Y., Gevaert, K., Tsiotis, G., and Psaroulaki, A. (2012). Quantitative proteome profiling of *C. burnetii* under tetracycline stress conditions. *PLoS ONE* 7(3), e33599.

Xu, C., Lin, X., Ren, H., Zhang, Y., Wang, S., and Peng, X. (2006). Analysis of outer membrane proteome of *Escherichia coli* related to resistance to ampicillin and tetracycline. *Proteomics* 6(2), 462–473.

Wang, M., Shen, Y., Turko, I. V., Nelson, D. C., and Li, S. (2013). Determining carbapenemase activity with ^{18}O labelling and targeted mass spectrometry. *Anal Chem* 85, 11014–11019.

World Health Organization (WHO) (2014). Antimicrobial Resistance a Global Report on Surveillance WHO, Geneva, Switzerland.

Zapun, A., Contreras-Martel, C., and Vernet, T., (2008). Penicillin-binding proteins and beta-lactam resistance. *FEMS Microbiol Rev* 32, 361–385.

16

Proteotyping: Tandem Mass Spectrometry Shotgun Proteomic Characterization and Typing of Pathogenic Microorganisms

Roger Karlsson,[1,2,] Lucia Gonzales-Siles,[3] Fredrik Boulund,[4] Åsa Lindgren,[1,3] Liselott Svensson-Stadler,[1,3] Anders Karlsson,[2] Erik Kristiansson[4] and Edward R.B. Moore[1,3]*

[1] *Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg, Sweden*
[2] *Nanoxis Consulting AB, Gothenburg, Sweden*
[3] *Department of Infectious Diseases, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*
[4] *Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden*

16.1 Introduction

Rapid and accurate species identification and diagnosis of infectious diseases are of great importance in clinical settings, especially for patients suffering from infections that could lead to life-threatening situations. There is also an increasing need for diagnostic methods that provide information about the phenotypes associated with antibiotic resistance and virulence. However, the specific traits of the infectious pathogen, including antibiotic resistance and virulence, are often associated with strain-level differences rather than species-level differences, and methods with higher resolution capable of strain-level typing are therefore increasingly necessary. The terms *microbial typing* and *microbial type* are used in microbiology in different contexts, referring to various characteristic delineations, for example, biotype, phenotype, genotype, sequence-type, serotype, pathotype, and ecotype (van Belkum, Durand *et al.*, 2013). In the broadest sense, the typing of microorganisms refers simply to the categorization of individual organisms, based upon specified criteria or methods of analyses. However, in practice, the terms also are often used to designate subspecies-level designations (Figure 16.1).

The implementation of matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometry (MS) protocols for the identification of species and diagnostics of infectious diseases has been a great success. However, recognized limitations of MALDI-TOF MS identification of microorganisms encourage further development of the implemented protocols as well as development of approaches based on tandem MS (LC-MS/MS). The ever-evolving analytical instruments for MS together with advances in bioinformatics and constantly expanding genomic and proteomic

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

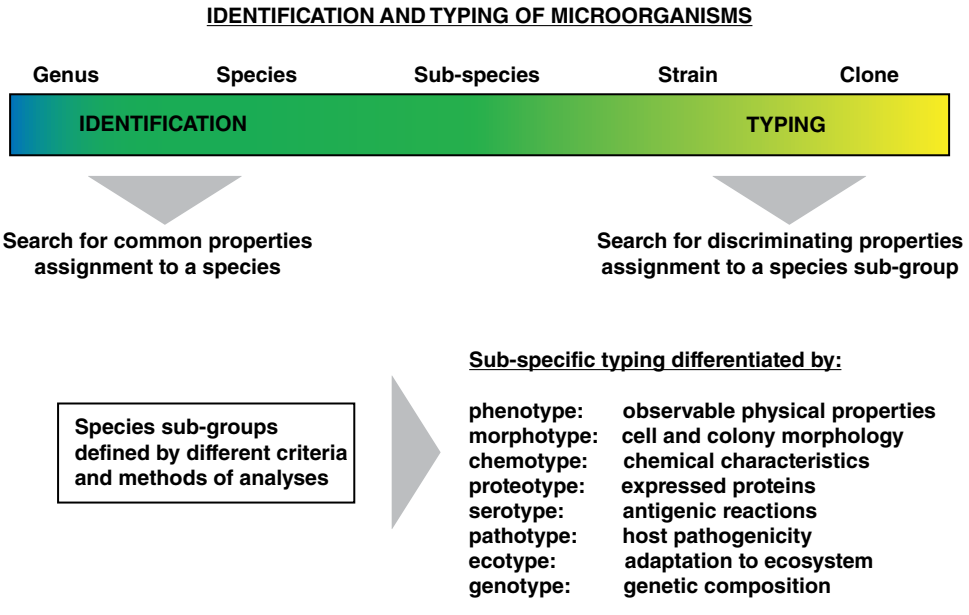


Figure 16.1 Illustration showing the distinction between microbial "identification" and "typing", that is, what levels of resolution are needed for which levels of taxonomy. Also shown are different commonly used terms for different subgroup types.

databases will generate corresponding progress in proteotyping, that is, the use of proteomics for microbial analysis and characterization.

16.2 MS and Proteomics

Microbial proteomics-based analyses comprise the study of the expression of genes and the structure and function of the resulting cellular proteins. The developments of proteomics have relied on recent advances in MS instrumentation and MS-based analyses. The earliest attempts to apply MS for analyzing biological samples were limited to relatively small-molecular-weight compounds (Anhalt, 1975). The development of soft ionization techniques, such as matrix-assisted laser desorption/ionization (MALDI) (Hillenkamp, 2000) and electrospray ionization (ESI) (Whitehouse, Dreyer *et al.*, 1985; Fenn, Mann *et al.*, 1989) were major advances, preserving the integrities of larger-molecular-weight compounds such as proteins and protein complexes (Dreisewerd, 2003). Briefly, in the case of MALDI MS, analyses are performed by placing samples onto a target plate with a matrix solution. A laser pulse is 'shot' against the target spot, whereby the sample analytes and matrix molecules are ionized and desorbed from the target plate, and the charged analytes are accelerated into a mass analyzer. In the case of ESI, a solution is eluted continuously from a capillary tip, and a voltage is applied between the tip of the outlet and the MS analyzer inlet, creating an aerosol of droplets. By using a heated capillary or a heated sheath gas, the droplets evaporate, and the charged analytes enter the mass analyzer (Banks, 2014). Four types of mass analyzers

are commonly used for proteomics applications: quadrupole ion trap (QIT); linear ion trap (LIT) or linear trap-quadropole (LTQ), TOF or Fourier transform ion cyclotron resonance (FTICR) (Yates, Ruse *et al.*, 2009). MALDI is commonly used as an ionization source, together with the so-called scanning MS analyzers, such as the TOF (i.e. measuring the flight time of the analytes, which corresponds to their masses and charges), with the analysis commonly done in a pulsed mode. ESI is used in a continuous mode of analysis, with the analytes continuously eluted from a sample separation step, such as liquid chromatography (LC), prior to entering the MS instrument (Fenn, Mann *et al.*, 1989; Banks, 2014). In this setup, ESI is usually combined with ion beam MS analyzers, such as the quadrupole (which separates ions according to the stability of the mass/charge ratio), and the trapping MS analyzers, such as the ion trap, Orbitrap and ion cyclotron resonance (ICR) analyzer (which separates ions according to their m/z resonance frequency) (Yates, Ruse, *et al.*, 2009).

Hybrid mass spectrometers combining different mass analyzers have been created to make use of the different strengths and performances of the individual analyzers, including triple quadrupole (TQ), quadrupole-LIT (Q-LIT), quadrupole-TOF (Q-TOF) and linear ion trap–FTICR (LTQ-FTICR) (Yates, Ruse *et al.*, 2009); currently, one of the most popular types of mass spectrometers is the Orbitrap (Makarov, 1999; Hu, Noll *et al.*, 2005), and recently, a hybrid instrument, the LTQ-Orbitrap, demonstrating high resolution, high mass accuracy and good dynamic range, became available commercially. Table 16.1 shows an overview of the different mass analyzers and their analytical performances.

Mass spectrometers imparting high resolution, high mass accuracies, high sensitivities, large dynamic ranges and high scan rates are necessary for MS-based studies of the proteome of a microorganism, such as a bacterium (Dworzanski, Deshpande *et al.*, 2006; Armengaud, 2013; Karlsson, Gonzales-Siles *et al.*, 2015). The proteome is the compilation of the entire set of proteins produced by a given microbial cell, an

Table 16.1 Overview of different mass analyzers and their analytical performance.

Instrument	Resolution	Mass accuracy (ppm)	Sensitivity	Dynamic range	Scan rate (spectra/time unit)
LIT (LTQ)	2,000	100	Femtomole	1×10^4	fast
TQ (TSQ)	2,000	100	Attomole	1×10^6	moderate
Q-LIT	2,000	100	Attomole	1×10^6	moderate, fast
Q-TOF/IT-TOF	10,000	2–5	Attomole	$I \times 10^6$	moderate/fast
LTQ-FTICR/Q-FTICR	500,000	<2	Femtomole	1×10^4	slow/slow
LTQ-Orbitrap	100,000	2	Femtomole	1×10^4	moderate

The resolution, mass accuracy, sensitivity, dynamic range and scan rates are indicated for commonly used instruments such as the linear ion trap – LIT (LTQ), triple quadrupole – TQ (TSQ), quadrupole-linear ion trap – Q-LIT, quadrupole-time-of-flight – Q-TOF, ion trap-time-of-flight – IT-TOF, Fourier transform ion cyclotron resonance – FTICR and Orbitraps (Adapted from Karlsson, R., L. Gonzales-Siles, F. Boulund, L. Svensson-Stadler, S. Skovbjerg, A. Karlsson, M. Davidson, S. Hulth, E. Kristiansson and E. R. Moore (2015). "Proteotyping: Proteomic characterization, classification and identification of microorganisms – A prospectus." *Syst Appl Microbiol* **38**(4): 246–257).

organism or a biological system of different organisms. High scan rates and high sensitivities enable the detection of more peptides per unit time during the front-end separation (LC), features that are critical for successful analyses of complex samples, including whole-cell lysates or clinical samples containing mixtures of target molecules from different microbial species. High resolution and mass accuracy enable accurate identifications of peptides, which are essential features for performing proteomic analyses of closely related species or strains. A key improvement of MS instruments has been in the increased dynamic range, that is, the range of sample concentrations that a mass analyzer can handle at each time point (Armengaud, 2013). If the range of concentrations of peptides or proteins entering the mass analyzer at the same time is very large, only the most abundant ones will be detected. Injecting more or less sample will not solve the problem. However, modification of the sample preparation strategies might be implemented in order to reduce the sample complexity prior to analysis (Armengaud, 2013). The issue of poor dynamic range is important to keep in mind when performing metaproteomics (Chao and Hansmeier, 2012), wherein the abundance of individual members will vary within the microbial population structure. However, the dynamic range of MS instruments is also important for proteomic analyses when highly abundant proteins, such as ribosomal proteins, might mask proteins of lower abundance, such as outer-membrane proteins or virulence factors.

16.3 MALDI TOF MS

Since its invention in the late 1980s, MALDI-TOF MS has been used for a wide range of analytical applications in the life sciences (Hillenkamp F., 2000). Whole-cell MALDI-TOF MS, generating proteomic mass patterns (i.e. mass spectra profiles; Figure 16.2) for the identification of microorganisms, has been a major breakthrough for microbial characterization and identification (Kallow, Erhard *et al.*, 2010; Welker and Moore, 2011), particularly in clinical laboratories, where the combination of relatively simple sample preparations, rapid analysis times and low cost are essential for processing large numbers of samples (Bizzini, Jaton *et al.*, 2011). MALDI-TOF MS typically targets ribosomal proteins, other housekeeping proteins and other structural proteins that are

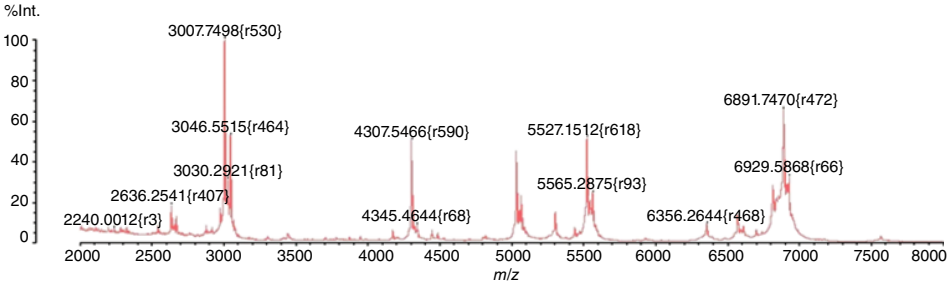


Figure 16.2 An example of a typical spectra obtained by MALDI-TOF-MS analysis of *Staphylococcus aureus*. The intensities of the peaks are normalized according to the highest peak (100%), and the mass-to-charge ratio (m/z) is indicated at the major peaks.

abundant in the cell, relatively independent of growth state or external stimuli and are usually in the mass range of 2,000 to 20,000 Da (Welker, 2011). The level of resolution to be expected from whole-cell MALDI-TOF MS analyses of microorganisms is generally limited to species-level differentiation and identification as shown in Figure 16.3 for *Staphylococcus* spp. This is due to the limitations of MS technology, where a limited number of data points (mass spectra peaks) are derived from the preferential ionization and detection of abundant proteins for any sample analyzed. However, MALDI-TOF MS has the potential for uses other than species identification, which is demonstrated in Figure 16.4. Using the standard protocol, whole-cell MALDI-TOF MS could discriminate between different pulse-field gel electrophoresis (PFGE) subtypes of *S. aureus*, indicating a capacity for strain differentiation using MALDI-TOF MS.

The use of MALDI-TOF MS for applications other than species-level identification, such as subspecies-level differentiation for epidemiological typing or detection of particular functional activities (e.g. antibiotic resistance, virulence), generally requires specialized cell preparatory protocols, before MS analyses, optimized to enhance the levels of resolution that can be attained. The optimizations needed to perform such analyses may include the use of particular matrix solutions, adjusting the mass range and laser power and exploiting more extensive sample preparations, for example, cell fractionation. Additionally, as demonstrated by Månsson *et al.* (Månsson, Resman *et al.*, 2015) in a study differentiating capsulated and non-capsulated *H. influenza*, the particular classification algorithm used for defining subgroups may be significant. A major limitation of MALDI-TOF MS, in the way that it is used today, is the same as for all methods dependent on database searches and comparisons, that is, the reference database. No matter how extensively the database develops, it will never be complete. This may result in some cases of non-identification or, even worse, misidentification of microorganisms. The establishment of species 'super-spectra', comprising compilations of the spectra of representative strains of a given species, provides reliable, comprehensive references for the comparison of spectra from individual isolates and strains (Welker and Moore, 2011).

Protocols required for subspecies-level differentiation and typing are dependent on the features that are being investigated and will need to be optimized for each particular taxon. One example is the analysis of serotypes of the food-borne pathogen, *Listeria monocytogenes*, for which a whole-cell MALDI-TOF MS method was developed and applied to track the origins of *L. monocytogenes* strains, grouping them according to their serotypes (Jadhav, Gulati *et al.*, 2015). The results obtained with MALDI-TOF MS corresponded well to the commonly used PFGE typing of bacterial serotypes.

Attempts to find specific biomarkers for strains of methicillin-resistant *Staphylococcus aureus* (MRSA) have been made, for example, using the presence of a mass peak at m/z 2415 as an indicator of MRSA strains in the CC5 MLST clonal complex. However, strains lacking this peak could be an MRSA or MSSA (methicillin-sensitive *Staphylococcus aureus*) strain (Josten, Dischinger *et al.*, 2014). Additionally, biomarkers for *S. aureus* lineages have been proposed (Josten, Reif *et al.*, 2013), although the specificities of these markers have been questioned (Lasch, Fleige *et al.*, 2014). Methicillin sensitivity in *S. aureus* has been analyzed by MALDI-TOF-MS, using the weak cation exchange ProteinChip Array (CM10) (designated SELDI-TOF-MS) (Shah, Rajakaruna *et al.*, 2011). Strain profile data generated were analyzed, using artificial neutral networks (ANNs), and seven key ions were found that were predictive of MRSA and MSSA. The

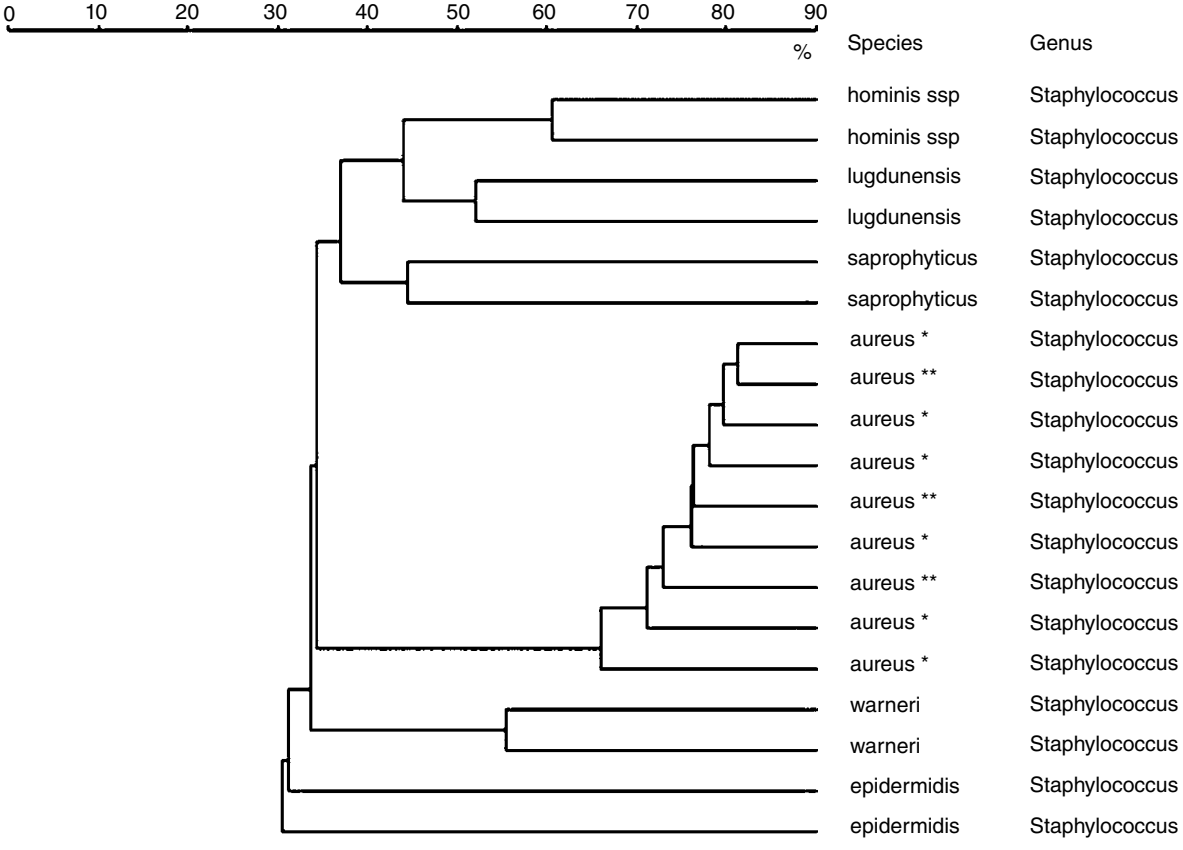


Figure 16.3 A dendrogram derived from MALDI-TOF MS spectra showing the relative similarities of *Staphylococcus* species. This demonstrates the ability to differentiate and identify *S. aureus* strains within the genus *Staphylococcus*, that is, species-level identification, and the inability, using standard MALDI-TOF MS protocols, to differentiate MRSA (shown by *) from MSSA strains (shown by **).

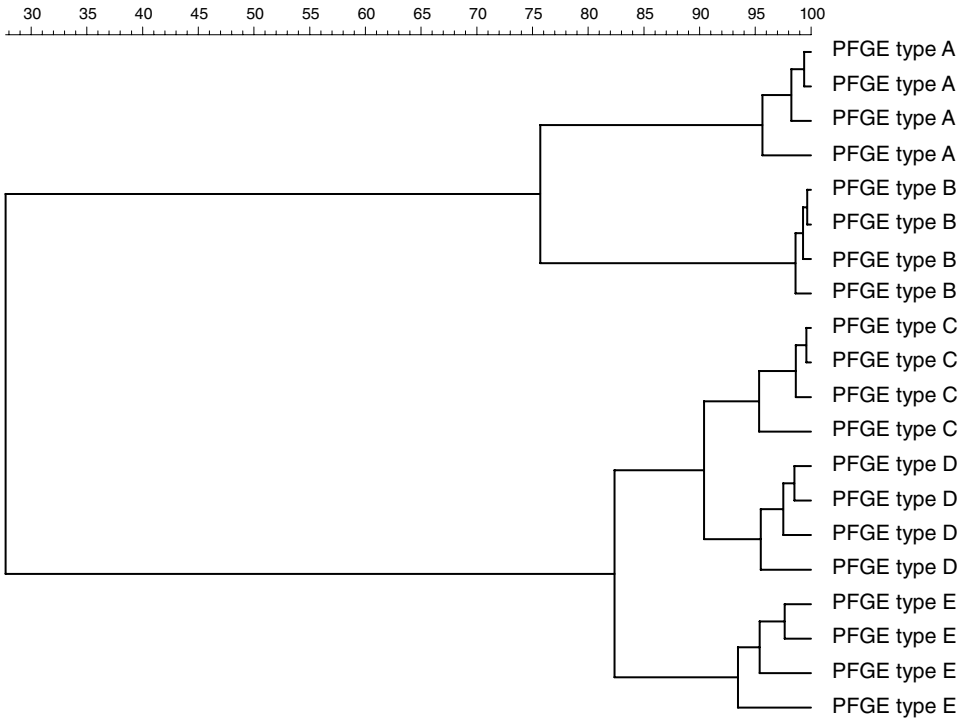


Figure 16.4 A dendrogram derived from MALDI-TOF MS spectra of the relative similarities of *S. aureus* strains. The subtypes identified by pulsed-field gel eletrophoresis (PFGE) are discriminated in the separate subclusters, indicating a high correlation between the different approaches. This shows that MALDI-TOF MS protocols have potential for strain-level differentiation in *S. aureus*.

vast majority of all strains were identified correctly, demonstrating the potential for subspecies-level characterizations. Expanding the mass range for the detection of m/z peaks outside the standard 2–20 kDa range that is used for accredited clinical diagnostic analyses is one approach to increase the resolution of MALDI-TOF MS. This approach has been applied successfully in the case of epidemiological typing of *Clostridium difficile* (Rizzardi and Akerlund, 2015), wherein the discriminatory peaks most useful for delineating species ribotypes were detected in the 30–50 kDa region of the spectra. It should be noted that this analytical modification was developed with whole-cell MALDI-TOF MS, and no extensive preparation protocol was used. The strains defined by the so-called high-molecular-weight (HMW)-typing corresponded well to the delineation of *C. difficile* strains by the accepted standard for *C. difficile* strain typing, that is, hybridization probe-based ribotyping, although not all HMW types corresponded to individual ribotypes. However, HMW-MALDI-TOF typing offers a rapid and cost-effective way to screen an outbreak, wherein turnover times of analyses are critical. In addition to *C. difficile* typing, expanding the mass range was also shown to be an effective method to detect the loss of the OmpK36 porin in *Klebsiella* spp. (Cai, Hu *et al.*, 2012). In that study, the outer-membrane proteins (OMPs) were extracted prior to MALDI-TOF MS analysis and then analyzed in the 15–80 kDa range. The results corresponded well to the results obtained by the commonly used SDS-PAGE method (Cai, Hu *et al.*, 2012).

Another approach to use MALDI-TOF MS for subspecies-level analyses has been applied by Durighello *et al.* (Durighello, Bellanger *et al.*, 2014), combining whole-cell MALDI-TOF MS analysis with nano-LC-MS/MS, to define the biomarkers for strain-typing *Francisella* spp. As some *Francisella* spp. are highly virulent, a rapid method for distinguishing between virulent and non-virulent strains was required. In the study, whole-cell MALDI-TOF MS was performed, identifying the 100 most intense m/z peaks, which were converted into theoretical masses. Nano-LC-MS/MS was used to analyze the 3–20 kDa proteome of the *Francisella* spp., and the theoretical masses obtained with MALDI-TOF MS were compared with the proteome data, detecting three specific biomarkers. The different patterns of these three biomarkers could be recognized using whole-cell MALDI-TOF MS, and were applied for defining specific strains of the *Francisella* spp.

The most successful approach for subspecies-level analyses may be the identification and exploitation of specific biomarkers or sets of biomarkers for differentiating between different strains. However, biomarkers presumed to be specific for given subsets of strains often lose their specificities when more strains are introduced into the analyses (Spinali, van Belkum *et al.*, 2015).

MALDI-TOF MS may lack the resolution that can be obtained by other molecular methods for microbial subspecies-level analyses. In most cases, MALDI-TOF MS is limited in its ability to perform such analyses. Metabolic features, including virulence and antibiotic resistance factors, usually cannot be detected using standard protocols. Furthermore, MALDI-TOF-MS identification requires, in most cases, cultivation and isolation of the microorganisms prior to analyses and will not be applicable to direct analyses of clinical samples or mixtures of bacteria, even though attempts are being made to develop protocols for such analyses. However, MALDI-TOF MS may prove to serve important roles as a screening tool, for example, for rapid epidemiological typing, to track outbreaks of pathogens, especially in local settings.

16.4 Tandem MS Shotgun Proteomic Analyses

MS-based shotgun proteomics (McDonald and Yates, 2003) analyses offer more detailed and comprehensive analyses of microorganisms. Two approaches may be applied: the so-called top-down and bottom-up proteomics (Meyer, Papasotiriou *et al.*, 2011). In short, the top-down approach relies on the analysis of intact proteins, whereas the bottom-up approaches include digestion (enzymatic or chemical) of the intact proteins into peptides, prior to the MS analysis (Figure 16.5).

16.5 Top-Down Proteomics

When performing top-down proteomics, intact proteins are partitioned by a separation technique (typically, LC), ionized to a highly charged state (~ + 15) and then analyzed by the mass analyzer (McLafferty, Breuker *et al.*, 2007; Gault, Malosse *et al.*, 2014). Following mass measurements, the intact proteins are fragmented, for example, by using collision-induced dissociation (CID), higher-energy collisional dissociation (HCD) or electron transfer dissociation (ETD), and the fragments are measured to

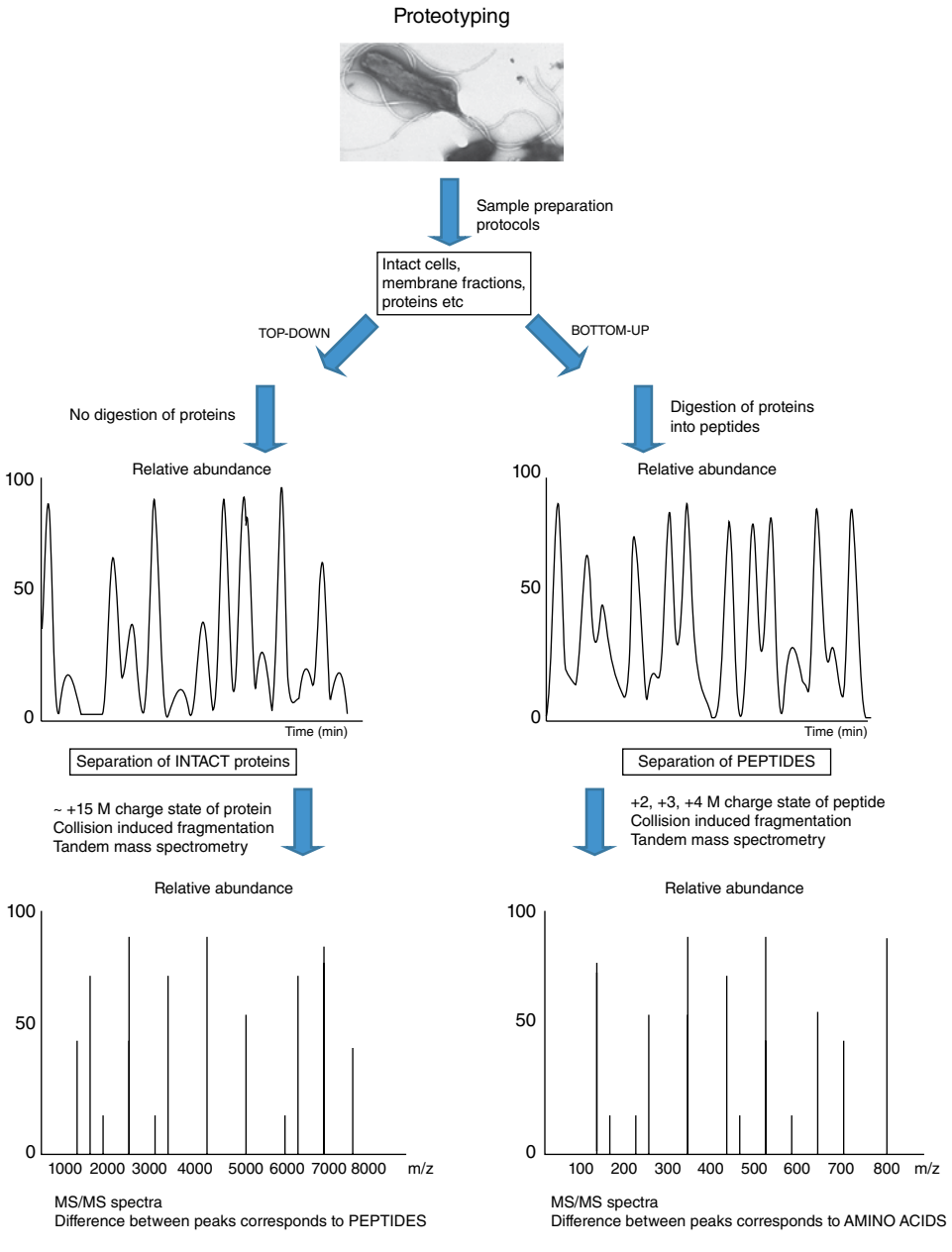


Figure 16.5 Top-down and bottom-up approaches employed for proteotyping. In top-down approaches (left part of the figure), the intact proteins are separated by liquid chromatography without prior digestion. Note the highly charged states of the intact proteins after ionization. Here, the generated tandem MS spectra reflect the mass differences of intact peptides. In bottom-up approaches (right part of the figure), the proteins are first digested into peptides (for example, using a digestive enzyme, such as trypsin), whereby the peptides subsequently are separated by LC and ionized prior to mass analysis. Here, the tandem MS spectra reflect the mass differences of amino acids.

create a mass spectrum. Tandem MS/MS is employed to infer the amino acid sequences of proteins, wherein the differences in the peaks correspond to peptides (McLafferty, Breuker *et al.*, 2007). The MS/MS spectra are matched against reference spectra in databases containing the spectra of known proteins. Intact proteins of masses greater than 6 kDa often carry a large amount of charge (protons); to be able to observe the molecular masses with LITs (up to 2000 Da), they must be analyzed in high-charged states. In order to perform top-down proteomics, high-resolution mass spectrometers (including FTICR and Orbitrap) must be used, because the mass analyzers must resolve similar mass peaks of the highly charged states, enabling discrimination of isotopic peaks and also the fragment ions (Wynne, Edwards *et al.*, 2010). Top-down proteomics has been used for the identification of microorganisms (Fagerquist, Bates *et al.*, 2006; see Chapters 18 and 20), including species lacking a reference genome sequence (Wynne, Edwards *et al.*, 2010). By matching the MS/MS spectra to known proteins in the databases and simultaneously characterizing the mass differences between the known proteins in the database and unknown proteins of the analyzed sample, it was possible to place the bacterium in the correct phylogenetic position (Wynne, Edwards *et al.*, 2010). Additionally, one of the major benefits of top-down proteomics is the ability to observe different protein isoforms, or proteoforms (www.topdownproteomics.org), together with their post-translational modifications (PTMs) (Williams, Monday *et al.*, 2005). For example, in a study of *Neisseria meningitidis*, top-down proteomics was able to reveal four different proteoforms of a particular type of pili, based on different glycosylation patterns (Gault, Malosse *et al.*, 2014). Further benefits of the top-down approaches include high sequence coverage of proteins and the possibility of discovering protein PTMs (Williams, Monday *et al.*, 2005), as well as possible protein–protein interactions. As the intact proteins are analyzed, the quantification of individual proteins is also more accurate, compared to bottom-up approaches (Swaney, Wenger *et al.*, 2010; Tran, Zamdborg *et al.*, 2011). Drawbacks include limitations in the ion charge state and sensitivity issues (Yates, Ruse *et al.*, 2009). There are also potential problems concerning ionization, because the ionization efficiency can differ substantially between proteins and because the diversity of proteins is much more varied compared to the shorter peptide species. Another major drawback is the more difficult front-end separation of intact proteins compared to the separation of peptides (Yates, Ruse *et al.*, 2009), which can reduce considerably the number of biomarkers that are elucidated and the throughput of the number of samples. Previously, top-down proteomics approaches have predominantly been used for analyzing individual proteins or simple protein mixtures. However, recent developments in MS instrumentation have improved the success of top-down approaches for proteomics applications (Macek, Waanders *et al.*, 2006; Chi, Bai *et al.*, 2007). For example, multidimensional separation using strong cation exchange chromatography in combination with reversed phase has been demonstrated to routinely detect as many as 4000 proteoforms from the proteome of *E. coli* (Whitelegge, 2013).

16.6 Bottom-Up Proteomics

Bottom-up proteomics refers to the analyses of peptides derived from proteins. Two protocols may be applied, following 'sort-then-break' and 'break-then-sort' strategies. In the sort-then-break approach, proteins are first separated, using electrophoresis or

LC, and then digested into peptides, which are analyzed using LC-MS (Henzel, Billeci *et al.*, 1993; Ogorzalek Loo, Hayes *et al.*, 2005). In the break-then-sort approach, peptides are generated from the digestions of protein mixtures and then sorted. Success in a bottom-up proteomics approach, particularly considering the break-then-sort approach, is dependent on the efficacy of separation of the peptides prior to MS analyses. In order to recover as many peptides as possible for MS analysis, several different types of peptide separation protocols can be used, including reversed-phase chromatography, ion-exchange chromatography, size exclusion, isoelectric focusing or combinations of these protocols (Issaq, Chan *et al.*, 2005; Fournier, Gilmore *et al.*, 2007).

The break-then-sort bottom-up strategy, in combination with a peptide separation protocol such as LC, coupled to an MS analyzer, is commonly referred to as 'shotgun' proteomics (McDonald and Yates, 2003). Using LC-MS/MS, the procedure for proteomics analyses of microorganisms is to determine as many mass spectra as possible for a sample (i.e. a pure culture microbial strain or a sample comprising a mixture of strains). Automated database search software enables rapid, high-throughput analyses of complex samples. The search software compares the mass spectra from each individual peptide, generated from the enzymatic digestion of proteins, in sizes between 5 and 30 amino acids long (Banks, 2014), to reference mass spectra having similar precursor ion mass and charge ratios.

Bottom-up proteomics relies on the use of proteolytic digestion, most commonly using the endoproteinase trypsin. Trypsin catalyzes the hydrolysis of peptide bonds, predominantly at the carboxyl termini of arginine and lysine residues (Olsen, Ong *et al.*, 2004), creating positively charged peptides with a low charge state and a high mass-to-charge ratio (m/z) that generate easily interpretable mass spectra (Thelen and Miernyk, 2012). Trypsin digestion of a cell lysate will yield a complex mixture of peptides, from low to high concentrations, to be analyzed. This exerts a demand on the dynamic range of the MS instrument when analyzing complex mixtures. This issue of dynamic range has always been a key issue for expanding the range of the peptides that can be detected; progress in this area is ever evolving, and today´s MS instruments are performing consistently better (Armengaud, 2013). As mentioned previously, closely related strains display only subtle differences in their genomes and, consequently, in their proteomes. When bottom-up approaches are employed, the digestion step is influenced by the quality of the sample or protein; poor digestion may lead to low sequence coverage of proteins, which will lead to difficulties when trying to discriminate similar proteins and, consequently, closely related microorganisms. A bottom-up approach has the advantage of higher throughput, compared to top-down approaches, and features high sensitivities and the ability to analyze highly complex samples. The drawbacks include the loss of some information of the intact protein, that is, information about any PTMs.

In order to reduce sample complexity, thus enabling detection of more peptides, different sample preparation strategies, including cell fractionation protocols, may be employed (Backert, Kwok *et al.*, 2005; Lopez-Campistrous, Semchuk *et al.*, 2005; Dumas, Desvaux *et al.*, 2009; Solis and Cordwell, 2011; Karlsson, Davidson *et al.,* 2012; Hebraud, 2014; Olaya-Abril, Jimenez-Munguia *et al.*, 2014). A reduced sample complexity will result in improved separation and ionization efficiencies, which will enhance the detection of low-abundance peptides. Sample preparation strategies may include the targeting of a specific sub-proteome, such as the cell wall (Solis and Cordwell, 2011). As microorganisms interact with their surroundings, particularly in the cases of pathogen–host

interactions, the surface-exposed proteins, including OMPs, are the first to experience changes in their surrounding environment. As such, OMPs are most likely to be the first to respond to adaptive changes by the microorganism in order to survive and thrive in new environmental conditions (Lin, Huang *et al.*, 2002). Corroborating this hypothesis is evidence that a microorganism displays various degrees of allele heterogeneity, wherein the genes for OMPs are observed to exhibit the highest levels of sequence variation (Yahara, Kawai *et al.*, 2012). Other genes, comprising the 'core' genome, are conserved among all strains of a microbial taxon, with lower levels of sequence variation. Consequently, this has led to strategies of targeting certain parts of the proteome of a microorganism by studying and analyzing the parts that are most discriminatory between strains, including the OMPs. For example, it has been shown that analyses of OMP extracts from *Yersenia pestis* and *Escherichia coli* offer improved strain differentiations, compared to whole-cell lysates, when performing proteomics-based characterizations of strains of these species (Jabbour, Wade *et al.*, 2010). Methods have, therefore, been developed for cell-wall isolations and digestions, cell-surface 'shaving', using protease treatments of intact microorganisms, or cell-surface labelling protocols, followed by isolations and digestions (Solis and Cordwell, 2011). Importantly, advances in bioinformatics have further lead to better predictions for the localization of surface-exposed proteins. Surface-shaving approaches or 'surfaceomics' have been developed as well, and one of the strategies employed is the lipid-based protein immobilization (LPI) approach, wherein intact microorganisms are immobilized, on the surfaces inside a FlowCell™ and exposed to protease digestions (Chooneea, Karlsson *et al.*, 2010; Karlsson, Chooneea *et al.*, 2010; Karlsson, Davidson *et al.*, 2012). This approach was used for analyzing *H. pylori* strains; intact cells of H. pylori strains J99 and ATCC 26695 were digested in an LPI FlowCell™ to produce peptides for LC-MS/MS analysis. Following matching of the tandem mass spectra against databases containing whole-genome sequences of the bacteria, strain-unique peptides were identified, and ranking was performed for discriminating and identifying the strains (Karlsson, Davidson *et al.*, 2012). Figure 16.6 below shows examples of unique peptides originating from the protein thioredoxin reductase and how the peptides match sequences of thioredoxin reductase for the strains *H. pylori* ATCC 26695 and *H. pylori* J99.

16.7 Proteotyping

Proteotyping is the comprehensive characterization, classification and identification of microorganisms, using high-resolution MS and proteomic analysis. The term *proteotyping* has been used for the analysis of the expressed genes of the influenza virus (Ha and Downard, 2011), but recently the term has been expanded to include the shotgun proteomic analyses of microorganisms (Karlsson, Davidson *et al.,* 2012; Penzlin, Lindner *et al.,* 2014; Karlsson, Gonzales-Siles *et al.,* 2015). The principle underlying proteotyping relies on the variation of amino acid sequences of proteins, as the corresponding genes vary through mutation events. The analyses of the expressed genes and the resulting proteins of a microorganism can thus be viewed as a 'snapshot' of the encoded part of its genome. Theoretically, as long as the differences in target proteins are detectable, proteomics approaches should be able to be applied to the analyses of microbial features and, thus, utilized for the characterization, classification and identification of microorganisms.

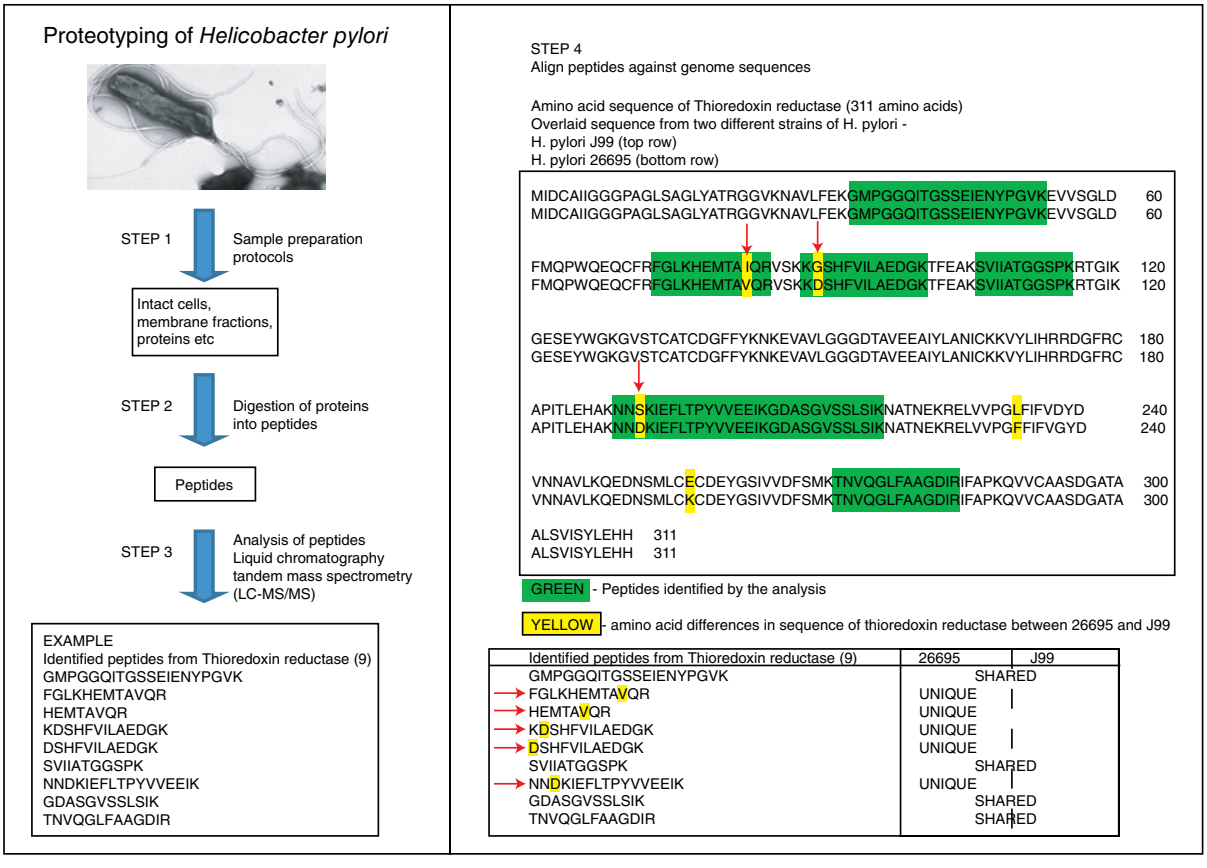


Figure 16.6 Workflow of proteotyping. The *Helicobacter pylori* strain ATCC 26695 was analyzed with the bottom-up proteotyping approach. Step 1 shows sample preparation protocols involving cell fractionation protocols, or methods for keeping cells intact are employed. Step 2 is the digestion of the sample proteins into peptides. Step 3 represents the analysis of the generated peptides, using LC-MS/MS. An example of nine peptides coming from the protein thioredoxin reductase of *H. pylori* is shown. In Step 4, the peptides are aligned against available genomes in databases, here exemplified by two sequences for the thioredoxin reductase, one from the strain *H. pylori* ATCC 26695 and one from the strain *H. pylori* J99. The green-shaded areas represent the identified peptides. The yellow-marked amino acids show where the amino acid sequences differ between the two strains. The red arrows show where the identified peptides contain these amino acid differences, thus enabling identification of the correct strain, as these are unique for the particular analyzed strain, *H. pylori* ATCC 26695.

Proteotyping, as an analytical method, is intimately correlated with genotypic or genomic data and offers an alternative approach for a holistic characterization of microorganisms. In particular, proteotyping offers the advantage of elucidating expressed taxonomic biomarkers for antibiotic resistance and virulence, as well as catabolic and anabolic pathways of cell metabolism, with implications for medical, biotechnological and environmental applications. Apart from these aspects, proteotyping also provides the means to obtain insight into microbial activities, that is, from isolated individual organisms and from more complex samples comprising multiple microorganisms (i.e. metaproteomics) (Chao and Hansmeier, 2012).

When using proteotyping for characterizing microorganisms, it should be recognized that species that are not highly similar are easily discernible using genotypic or phenotypic tools, whereas microorganisms that are phylogenetically closely related, that is, closely related species or strains of a microbial species, present particular requirements for the methodologies used. This is because closely related microbes display only subtle differences in their genomes and, consequently, in their proteomes. Therefore, these differences may be identified only by analyzing a comprehensive fraction of the proteome expressed by the microorganism. Furthermore, the overall proteomic expression of a microbial genome will depend on factors such as age, growth and environmental conditions. Therefore, proteomic expression may vary depending on the state of the microorganism as well as responses to different types and degrees of various stresses. This motivates the ultimate aim of detecting and characterizing all proteins potentially expressed by microorganisms to obtain high discriminatory levels of analysis (Meyer, Papasotiriou *et al.*, 2011). One important step towards achieving this goal is to analyze microorganisms under different nutrient and cultivation conditions and different stresses (Lopez-Campistrous, Semchuk *et al.*, 2005; Pieper, Huang *et al.*, 2009). The halophile *Halobacterium salinarum* was subjected to approximately 30 growth conditions, and each of the different conditions was analyzed for genomic expression (Klein, Aivaliotis *et al.*, 2007). Approximately 35% of the theoretical proteome could be detected through this approach. The importance of using non-standard growth conditions was also demonstrated with *Ruegeria pomeroyi* under varying conditions, thus reaching identifications of approximately 50% of the theoretical proteome (Christie-Oleza, Fernandez *et al.*, 2012; Armengaud, 2013). This type of work will lead to new types of guidelines on how to process microorganisms to achieve the most comprehensive coverage of the proteome and the highest discriminatory level of proteotyping through, for example, various sample preparations and growth conditions. And, as the breadth and depth of genome sequence data for microbial species increases and annotation of determined genome sequences improves, the proteome coverage will likely be seen to improve dramatically.

Bioinformatics is vital for the analysis of the data generated by shotgun proteomics. In proteotyping, the overall aim of the bioinformatics analysis is to determine the microbial composition in a sample by accurately estimating the relative abundance of the present microorganisms. A general approach to this problem is assigning the measured peptides to their correct affiliation in the taxonomic tree and, based on their distribution, deducing the taxonomic composition of the sample. However, the peptides generated by shotgun proteomics are short and only cover a part of the proteome, and only a few will thus have the ability to distinguish between closely related species. A central step is therefore to identify which of the peptides that are 'discriminative', that is,

find the subset of peptides that provide information on the organisms present in the sample. A common way to identify the discriminative peptides is to align each individual peptide to a comprehensive database containing reference proteomes. The alignment can be made either directly to selected reference proteins or to all genes in the reference genomes by translating the genes in the genomes into protein sequences. Typically, peptides that uniquely match a reference at the given taxonomic level (strain, species, genus, etc.) are then classified as discriminative. The taxonomic composition in a sample can then be statistically estimated from the relative abundance of discriminate peptides, and for samples containing multiple microbial organisms, each type will have its own set of discriminate peptides, which can be used to calculate the complete taxonomic composition.

The complete bioinformatics workflow necessary for proteotyping is described in Figure 16.7. The workflow takes as input a collection of mass spectra. The four main computational steps – identifying the peptides, map them to the reference genomes, provide filtering of non-informative matches and identifying which peptides are discriminatory – are marked by a dashed box. The end result produced by the workflow is an estimate of the taxonomic identity of the microorganism or mixture of microorganisms in the sample. In the following sections, the workflow will be described in more detail, and the challenges at each stage and the considerations that must be made when dealing with them will be discussed.

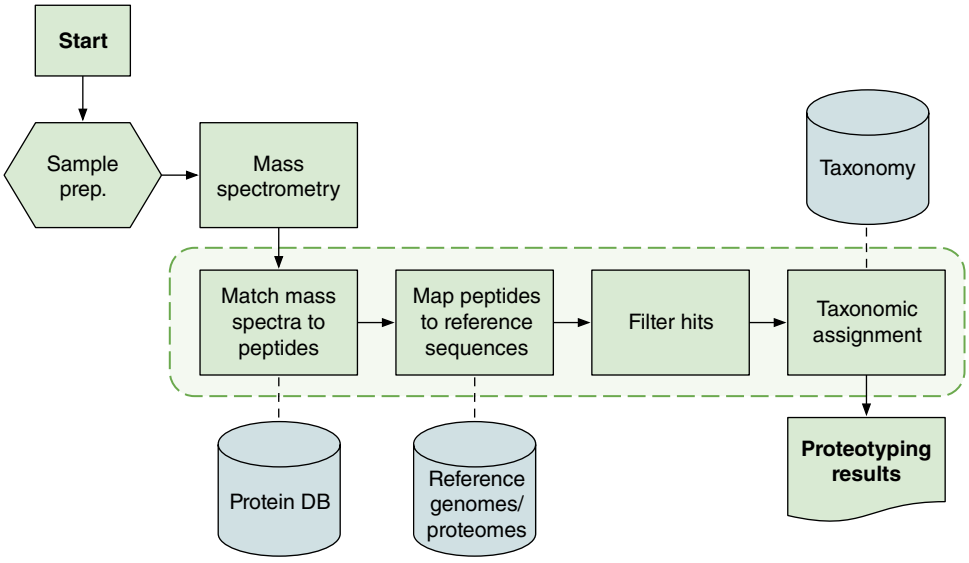


Figure 16.7 A general overview of the proteotyping workflow. The process starts in the top-left corner with sample preparation and finishes with the proteotyping results in the lower-right corner. Each box represents a specific experimental or analytical stage of the workflow. The three cylinders correspond to reference databases that are essential for the analysis, including a comprehensive protein database, a set of reference genomes and their association with the taxonomic tree. Four fundamental bioinformatics stages are highlighted by the dashed rectangle in the centre of the graph. The final proteotyping result is an estimate of the taxonomic composition of the sample.

16.8 Matching MS Spectra to Peptides

In the first step after MS, the spectra must be converted to peptide sequence information for further 'downstream' analysis. This is typically done by matching the generated spectra to a database of reference peptide sequences. Several implementations of different algorithmic approaches are available, both commercial (e.g. Mascot (Perkins, Pappin *et al.*, 1999), SEQUEST (Eng, McCormack *et al.*, 1994)) and free, open-source (e.g. X!Tandem (Craig and Beavis, 2004), MyriMatch (Tabb, Fernando *et al.*, 2007)). These types of algorithms are often referred to as MS 'search engines', because they use MS data to search for suitable matches to peptide sequence databases. In essence, most algorithms operate by comparing the molecular weights of the peptides recorded in the mass spectra to the expected molecular weights of peptides in a reference database. An early and well-known example of such an algorithm for probabilistic molecular weight search is MOWSE (Pappin, Hojrup *et al.*, 1993), which formed the basis of the widely used commercial search engine Mascot (Perkins, Pappin *et al.*, 1999). It should be noted that the matching of spectra is a computationally costly process. The accuracy of the matching is also highly dependent on the size of the reference database. To enable matching to databases containing as many proteins as possible, optimized implementation using multithreading and/or the ability to be distributed over computer clusters and clouds are often required.

Another approach to produce peptide sequences from spectra attempts to derive the peptide de novo from the spectral patterns themselves, without the use of a reference database. Examples of such algorithms include commercial alternatives (e.g. PEAKS (Ma, Zhang *et al.*, 2003)) and free, open-source implementations, for example, Lutefisk (Taylor and Johnson, 1997), PepNovo (Frank and Pevzner, 2005), NovoHMM (Fischer, Roth *et al.*, 2005)). Accuracies of de novo approaches are generally not comparable with the previously mentioned reference-based methods (Hernandez, Muller *et al.*, 2006). However, recent advances have seen an increase in the number of tools that combine the results produced by the two different approaches, for example, PEAKS-DB (Zhang, Xin *et al.*, 2012) to achieve higher accuracy.

Most methods for proteotyping directly utilize the results from a database search algorithm to provide the information used to estimate the taxonomic composition of samples (Dworzanski, Snyder *et al.*, 2004; Jabbour, Deshpande *et al.*, 2010; Tracz, McCorrister *et al.*, 2013; Penzlin, Lindner *et al.*, 2014). In doing so, such methods rely on the assumption that the peptides identified by the mass spectrum search engine are also suitable for calculating an accurate estimate of the taxonomic affiliation of the peptide. However, just as in the case of de novo peptide sequencing, where it is impossible to deduce the most likely affiliation of the peptides without additional database searches, there are arguments to separate the matching of the spectra to peptide sequences from the assignment of their taxonomic origin. The matching of the spectra is highly dependent on a large reference database containing as many protein variants as possible. If the number of different proteins and protein variants available in the reference database is too limited, a substantial proportion of the spectra from the sample will be satisfactorily match and will thus be discarded. Using too limited a reference database may also introduce a bias towards proteins that are highly conserved between organisms and thus lower the potential for discriminating between microbial types (Dotsch, Klawonn *et al.*, 2010). This can negatively affect the downstream analysis. When the taxonomic assignment is

removed from the matching of spectra, the reference protein database can be optimized on the basis of comprehensiveness. The accuracy of the matching of spectra can therefore be substantially increased by including proteins from sources with less reliable taxonomic information, such as open reading frames predicted from genomes with draft assemblies and even metagenomes (Markowitz, Chen *et al.*, 2015). Another important advantage is that the assignment of taxonomic affiliation, as a separate step, can be done directly to references organisms. By using translated alignments, each peptide can be aligned to the entire genomes, thereby making this process almost completely annotation agnostic. This makes the entire analysis less susceptible to errors introduced by errors in the annotation, such as missed and falsely predicted gene models.

16.9 Mapping Peptides to Reference Sequences

The second step of the workflow aims to identify the taxonomic affiliation of the peptide sequences by comparing them to the genome of a set of reference organisms. In order to capture all biological variability in the sample, this step needs to be performed with sensitive peptide-to-nucleotide aligners. Indeed, a single amino acid substitution in one peptide is enough to make it discriminative between two closely related species or strains. Mapping the peptides to reference sequences in nucleotide format requires an alignment algorithm capable of performing translated mapping. In translated mapping, the reference genome sequences are translated into all six possible reading frames, and the alignment is made between the peptides from the sample and the resulting amino acid sequences from the reference genomes database. Depending on the desired quality of alignment, several different algorithms can be considered. Examples of two commonly used algorithms that can perform translated mapping are TBLASTN (Altschul, Madden *et al.*, 1997) or The BLAST-Like Alignment Tool (BLAT) (Kent, 2002). BLAT is particularly well suited to perform this type of alignment because it is optimized to map large sets of DNA or protein sequences into a reference database containing (relatively) few but long sequences, for example, bacterial chromosomes. It should be noted that it is also possible to apply peptide alignment algorithms that lacks translated mapping functionality via an alternative approach. Reference genomes can, a priori, be translated into amino acid sequences in all six possible reading frames and compared to the peptides using a standard protein alignment algorithm (e.g. BLASTP).

It is crucial that the reference database used at this stage exclusively contains sequences of the highest possible quality. In particular, each reference genome sequence has to have a correct taxonomic assignment, preferably down to the strain level. Furthermore, gaps, low-quality regions, and sequencing errors can result in missed or, in the worst case, wrongly aligned peptide fragments. This can, in turn, introduce substantial levels of noise into the data, which may lead to erroneous results. The proteotyping analysis workflow is also sensitive to the presence of sequence regions that are horizontally transferred between species, such as conjugative elements (e.g. plasmids), phage sequences, and transposons. These elements often show a high conservation and may appear in evolutionary distant species. Peptide fragments mapping to genes in such regions may, therefore, appear to be erroneously unique. Because high plasticity is a natural property of many bacterial genomes, the reference sequences must be carefully curated to minimize the presence of such sequences.

The alignment of peptides against the reference database will result in zero, one or several matches for each peptide. Because these alignments differ in quality and relevance, it is important to apply suitable filters to minimize the risk of including incorrect or suboptimal matches in the downstream analysis. In particular, peptide fragments matching several different reference sequences need to be examined further before they can be assigned a taxonomic position in the next stage of the workflow. Filtering matches is essential to maintain high accuracy in the taxonomic assignment. The filtering process can been divided into two consecutive parts, here termed *absolute* and *relative* filtering. The absolute filtering step aims to remove alignments that have a too poor informational value, that is, matches that are too distantly related to the sequence that the peptide fragment originates from. It uses a cut-off criterion and removes any matches that do not fulfil the criterion. Peptide alignments passing this cut-off are thus deemed to be of sufficiently high quality to be used for the typing. This cut-off is typically set to an absolute, fixed value for all fragments, regardless of their length and sequence characteristics. The cut-off is often based on one or multiple features, for example, sequence similarity over the aligned region, the overall peptide fragment length and the fraction of peptide fragment covered by the alignment. Note that the absolute filtering step discards peptides without any matches passing the absolute filtering criteria. Such fragments are thus removed and not passed on to the subsequent steps of the analysis.

The relative filtering step aims to further narrow down the set of matches for each peptide fragment. In particular, the majority of the peptides have typically two or more similar matches to reference sequences that pass the absolute filtering. For each peptide, this filtering step compares all the matches relative its best matches, and those that are deemed to be of substantially lower quality are removed. Thus, the relative filtering removes non-informative matching with the aim of increasing the number of discriminative peptides. Typically, the relative filtering step is realized by examining the quality of the alignment, and the matches with a sequence similarity lower than a fixed percentage unit below the best match are removed. This will thus remove matches to closely related reference sequences that have high enough scores to pass the absolute filtering step, although, if included, would negatively impact the ability to correctly assign the peptide fragment to a taxonomic unit. A suitable value for the relative filtering parameter is, in our experience, somewhere in the range of 2%–5% sequence similarity (compared to the best hit) depending on the quality of the data and the species studied. Note that the relative filtering step will never completely remove a peptide from further analysis, as all comparisons in this step are made relative to the best match in the set of matches for each peptide, and one match will still remain for each peptide.

16.10 Taxonomic Assignment of Protein Sequences

After applying the absolute and relative filtering step and thereby removing incorrect and non-informative matches, the remaining peptide fragments are used to infer the taxonomic composition of the sample. This is done by assigning the peptides to positions in the taxonomic tree on the basis of their matches to the reference sequences. Because each peptide typically matches multiple reference genomes annotated at varying taxonomic levels, this processes is not straightforward and needs to be done with respect to the structure of the taxonomic tree. For example, some genomes have a clear annotation at the strain level, whereas others may have been associated only with the

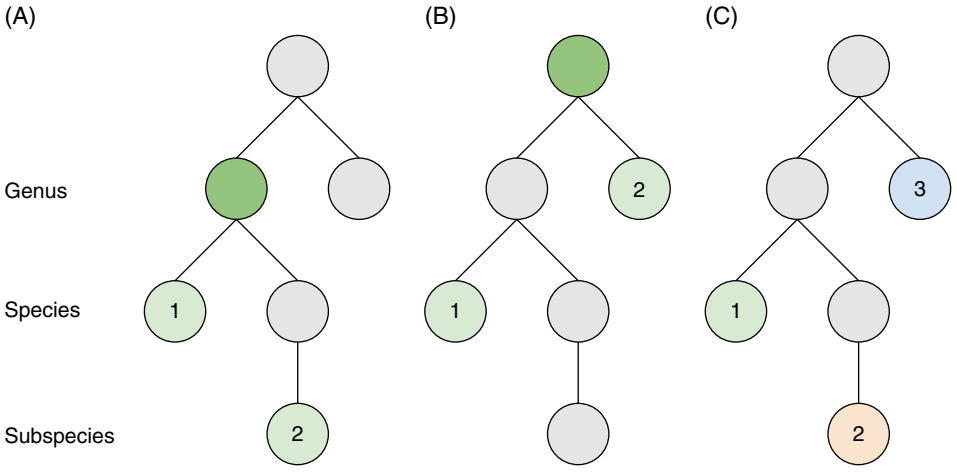


Figure 16.8 The lowest common ancestor (LCA) algorithm. The figure shows three examples (A, B, C) of peptides with matches to reference genomes at different levels in the taxonomic tree. (A) A single peptide has matched nodes 1 and 2. These nodes have a lowest common ancestor on the genus level, which means that this peptide is discriminative on the genus level, but not on any lower level. (B) A single peptide has matched nodes 1 and 2, leading to a lowest common ancestor above the genus level. Such a peptide contains very little useful information for proteotyping on species and subspecies levels. (C) Three different peptide have matched nodes 1, 2 and 3. There is no lowest common ancestor; each peptide is discriminative on its own at the level it matched (species, subspecies and genus, respectively).

species or even the genus level. These discrepancies can be overcome by applying the lowest common ancestor (LCA) algorithm (Figure 16.8) (Aho, Hopcroft *et al.*, 1973). The LCA algorithm maximizes the information captured for each peptide and then infers its most likely affiliation in the taxonomic tree. This is also the taxonomic level where the peptide is discriminative, that is, where it affiliates uniquely with one or several reference genomes. Figure 16.8 shows three examples of how peptide matches can be used to infer the taxonomic assignment of a peptide with matches to reference genomes located at different levels in the taxonomic tree.

After the peptides have been assigned to the taxonomic tree, the taxonomic composition can be derived. This is typically done by counting the discriminative peptides found at each node at a given level in the taxonomic tree (e.g. species level). The relative abundance, calculated by relating the number of discriminative peptides at each node to the total number of discriminative peptides, provides an estimate of the composition of organisms in the sample. It should be emphasized that this approach can be applied to pure cultures as well as complex mixtures of microorganisms commonly encountered in clinical samples.

16.11 Challenges Assigning Fragments to Lower Taxonomic Levels

Typing microorganisms at lower taxonomic levels (e.g. species and subspecies level) presents a number of challenges. At the lower taxonomic levels, the discrepancies between proteomes from different microbial organisms will decrease and, as a result,

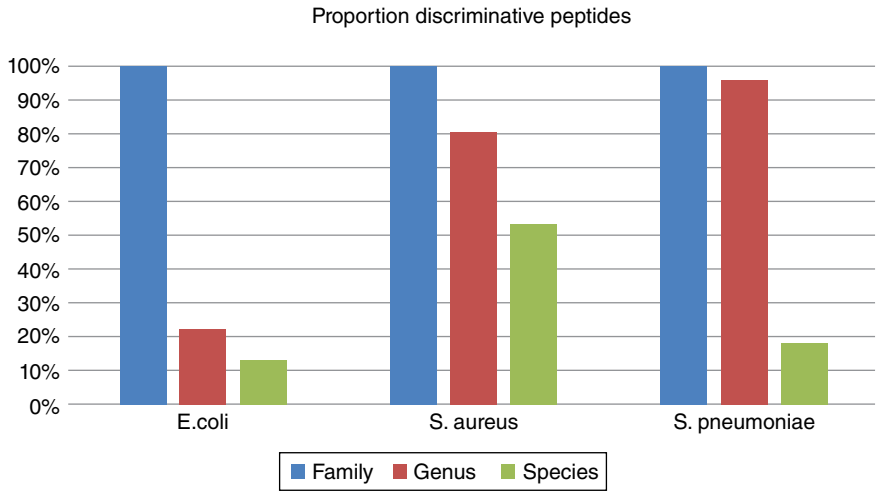


Figure 16.9 The number of discriminative peptides decreases rapidly at lower taxonomic levels. This figure shows the number of discriminative fragments and the family, genus and species levels for the bacteria, *E. coli, S. aureus* and *S. pneumoniae*. The drop in discriminative fragments for each taxonomic level differs between species and is dependent on the comprehensiveness of the reference databases and the similarity of the proteome of the closest relative. The numbers were calculated on the basis of an implementation of the workflow using X!! Tandem (Bjornson *et al.*, 2008) together with a reference databases composed of NCBI GenBank non-redundant proteins (Benson *et al.*, 1999) and the Human Microbiome Project (HMP) reference proteomes (Turnbaugh *et al.*, 2007). Alignment of peptides were done using BLAT (Kent, 2002) against a manually curated version of NCBI RefSeq (Pruitt, Tatusova, and Maglott, 2007) and HMP reference genomes (Turnbaugh *et al.*, 2007).

there will be fewer discriminative peptides that can be used in the proteotyping. For example, the proportion of discriminative peptides drops rapidly from family to genus and then even further at the species level. This is demonstrated in Figure 16.9, which shows these effects for samples of pure cultures from three common pathogens: *E. coli*, *S. aureus* and *S. pneumoniae*. The number of discriminative peptides drops substantially for *E. coli* and *S. pneumoniae* and, at the species level, only 10%–20% of the peptides were discriminative at the family level. The reduction of peptides for *S. aureus* is small, where more than 50% of the peptides that are discriminative on the family level are also discriminative on the species level. Another example is the subspecies analysis of *H. pylori*. In this study, two *H.pylori* strains (i.e. J99 and ATCC 26695) were mixed in various ratios, ranging from 1:9 to 9:1 (Karlsson, Davidson *et al.*, 2012). Prototyping identified fewer than 100 discriminative peptides at this taxonomic level. This corresponded to only ~10% of the total identified peptides, but it was enough to accurately estimate the relative abundance of the strains in the mixtures (Figure 16.10). The rapid drop of discriminative fragments at the lower taxonomic level will, in many cases, negatively affect the accuracy in the estimation of the species composition and thereby reduce the overall performance. Proteotyping at the strain and species level is therefore requires higher number of discriminative peptides. This further emphasizes the need for comprehensive reference databases for matching of spectra and assigning the taxonomic affiliations.

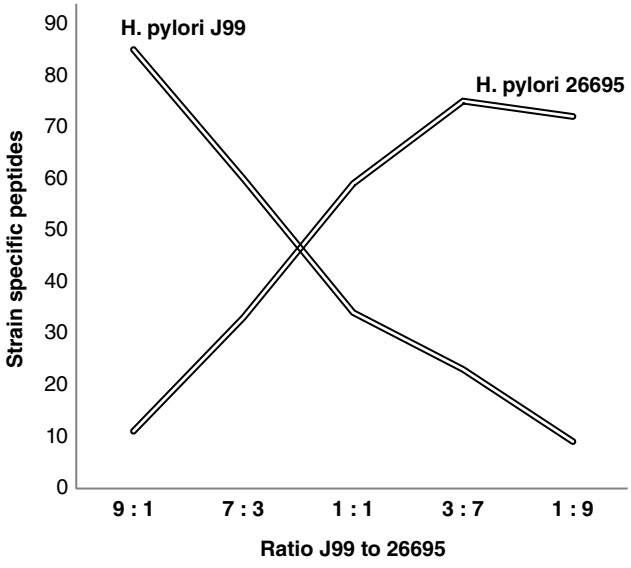


Figure 16.10 Numbers of strain-specific peptides from different mixtures of two *H. pylori* strains (J99 and ATCC 26695). The x-axis shows the how the strains were mixed, ranking from 9:1 to 1:9. Proteotyping analysis revealed that <100 peptides were discriminative, which corresponds to ~10% of the total identified peptides (Adapted from Karlsson, R., L. Gonzales-Siles, F. Boulund, L. Svensson-Stadler, S. Skovbjerg, A. Karlsson, M. Davidson, S. Hulth, E. Kristiansson and E. R. Moore (2015). "Proteotyping: Proteomic characterization, classification and identification of microorganisms – A prospectus." *Syst Appl Microbiol* **38**(4): 246–257).

16.12 Proteotyping for Diagnosing Infectious Diseases

Fast and efficient diagnostics is vital for the efficient treatment of patients with severe infections, such as sepsis, pneumonia and acute meningitis. Immunocompromised patients are susceptible to infections from commensals and other opportunistic bacteria, which makes reliable species identification crucial. These patients are dependent on diagnostics tools that can discriminate between harmless colonizers and the disease-causing microorganisms. The rapidly growing prevalence of infections caused by bacteria that resist antibiotic treatment has increased the need for faster and more reliable determination of antibiotic resistance profiles (Organization, 2014). DNA-based approaches, such as targeted PCR and whole-genome sequencing, can be used to efficiently detect the presence of genes or genetic alterations linked to virulence and decreased susceptibility to antibiotics. Proteotyping provides the means to identify and quantify the actual expression patterns of proteins and their associated pathways, which provides a more accurate picture of infectious agents and their pathogenic potential. Cheaper methods that are both less labour intensive and less time consuming are, however, needed before high-resolution proteomic characterization and discrimination of infectious microorganisms can be adapted in routine clinical settings.

An important consideration for any new method used in the clinical diagnostics of infectious diseases is the potential for analyzing clinical samples directly, that is, without

any cultivation steps, which may require 24 h to several weeks, depending on the relevant infectious microorganisms. One advantage of shotgun proteotyping is the possibility of analyzing the generated peptides directly from clinical samples, without culturing, targeting biomarkers that characterize infectious microbial species. However, the effective detection and identification of infectious bacteria in clinical samples requires, besides sensitivity for detecting typically low numbers of bacterial cells, reductions in the amount of interfering human proteins that are in the samples. Using proteotyping workflows (Karlsson, Davidson *et al.*, 2012), the correct infective microbial species of positive samples (as confirmed by accredited methods, including MALDI-TOF-MS) could be detected and identified by direct analysis of nasopharyngeal samples (Figure 16.11).

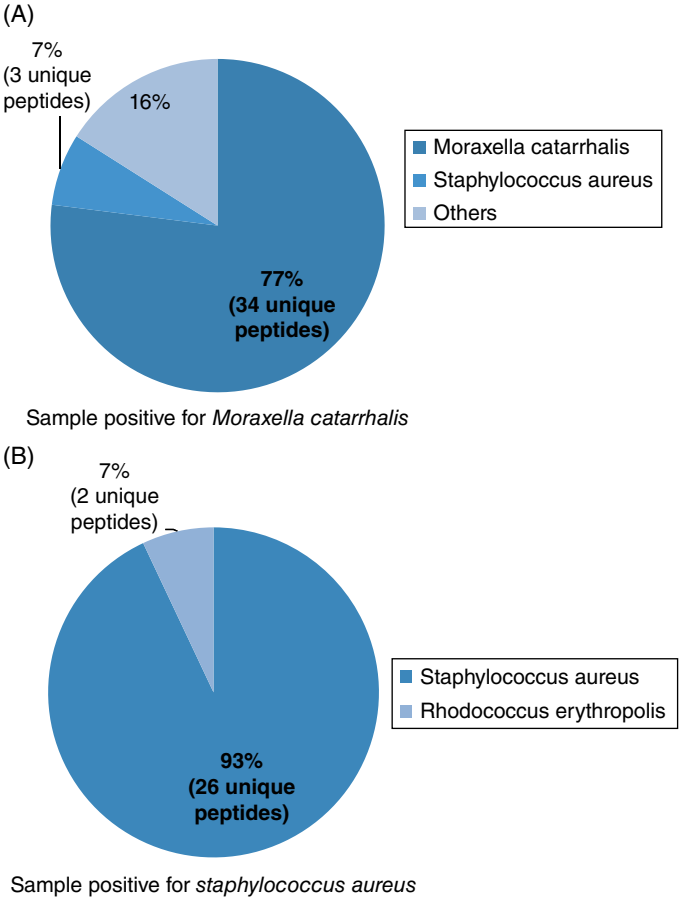


Figure 16.11 Proteotyping protocol implemented directly on nasopharyngeal samples. (A) A clinical sample identified as positive for *Moraxella catarrhalis*, by culture and MALDI-TOF-MS identification analysis. The proteotyping workflow found 34 unique peptides for *M. catarrhalis* and showed that this was the dominant species in the sample. (B) Sample identified as positive for *Staphylococcus aureus* by culture and MALDI-TOF-MS identification analysis. The proteotyping workflow found 26 unique peptides for *S. aureus* as the dominant species in the sample.

16.13 Outlook

Microbiology has evolved from the days when the characterizations of microbial isolates were limited to descriptions of morphologic and metabolic features. Phylogeny has been adopted as the basis of microbial taxonomy, and genome sequences are recognized to be the ultimate references for determining phylogeny (Wayne, Brenner *et al.*, 1987). In turn, phylogeny and microbial taxonomy provide the basis for the diagnostics of infectious diseases. Evolutionary genetics provides the rationale for adopting DNA sequence data for analyzing microbial diversity (Palys, Nakamura *et al.*, 1997). Microbial sequence types evolve into bacterial lineages in different ecosystems, wherein sequence divergence between the lineages will be greater than that within them, resulting in monophyletic 'ecotypes' (Cohan, 2002). Although the 'units' of microbial systematics are most commonly the genus and species designations, it could be argued that subspecies-level delineations may sometimes provide the more practical perspective, with respect to applications in the treatment of disease, the study of ecology and biotechnological exploitation (Moore, Mihaylova *et al.*, 2010). It is often at the strain level or, perhaps, more precisely, at the ecotype level, that applications involving microbiological processes are realized. Thus, the ability to detect and characterize microorganisms at the subspecies, as well as at the species level, according to comprehensive expression profiles, will become increasingly relevant for microbial systematics. The comprehensive characterization of microorganisms for classification and identification historically has required a range of methodologies to cover the diversity of newly discovered microorganisms and their increasingly complex taxonomy (Figure 16.12).

In recent years, the development of new methodologies for the systematic analyses of microorganisms has been changing towards methods that provide a holistic organismal 'blueprint', rather than applying combinations of different methods that target selected microbial features with limited resolution (Figure 16.12).

Newly developed genomics-based approaches for microbial characterization, including next generation sequencing (NGS) and approaches involving mass-spectrometry-based proteotyping (exploiting indirect measures of genome sequence data) enable microbial characterization across the entire taxonomic range of microorganisms. NGS technologies offer superior sequencing depth, and the entire genome of a bacterium can today be fully characterized and assembled rapidly and at low cost. Thus, whole-genome sequencing has been suggested as a new and versatile tool to be adapted for the routine diagnostics of bacteria and other microorganisms. However, DNA-based technologies measure only the presence of gene sequences and their variation, representing the potential of genome expression. Phenotypes associated with the over- and under-expression of chromosomal genes may be difficult to detect. For example, altered expression of ampC, porins and efflux pumps are commonly associated with resistance phenotypes in, for example, *E. coli*, *A. baumannii* and *P. aeruginosa* (Strateva and Yordanov, 2009; Rumbo, Gato *et al.*, 2013). Although genomics-based approaches offer the potential for detecting the relevant genes encoding for increased virulence risk and decreased antibiotic susceptibility, proteotyping offers the means to determine clinical patterns that are actually expressed, which is relevant for the optimal treatment of infectious disease. Shotgun proteotyping of microorganisms has huge potential, given a sufficient number of determined peptide fragments, to achieve comprehensive, high-resolution, species- and subspecies-level identifications. The use of discriminative peptide fragments creates

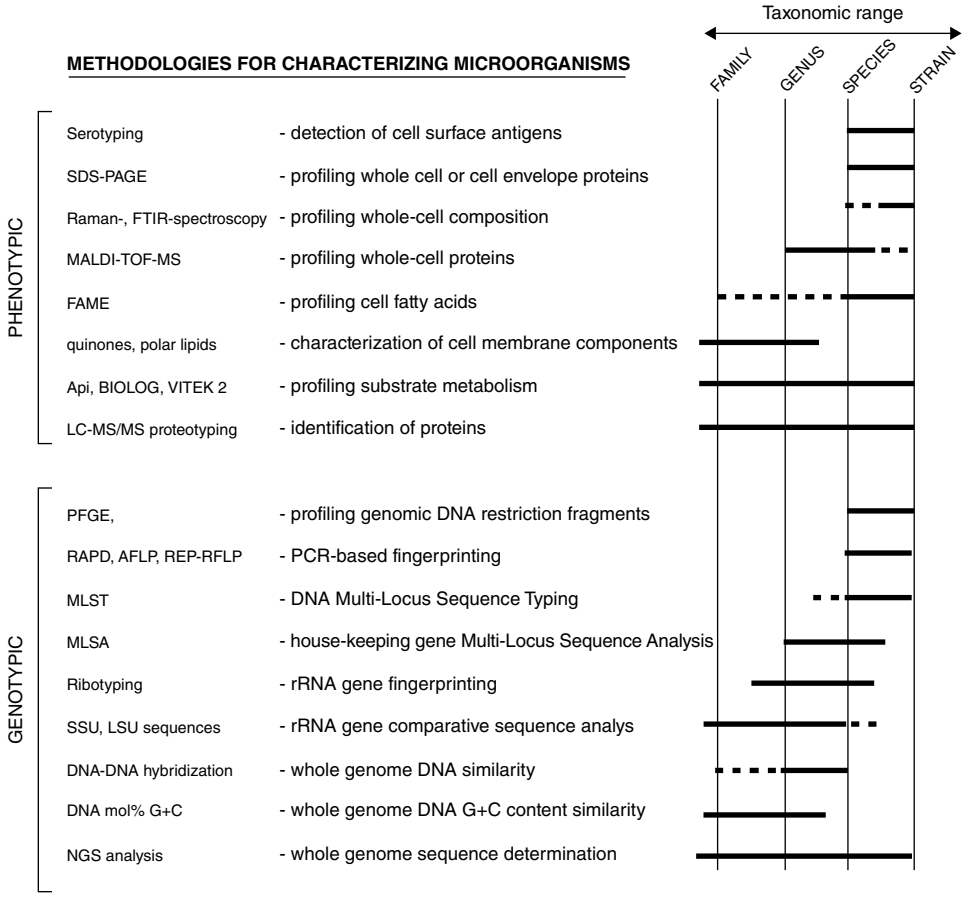


Figure 16.12 Methodologies for phenotypic and genotypic characterizations. The taxonomic resolution for the different methodologies are indicated by a solid black line; the dotted lines indicate applications to limited extents. Some methodologies, such as LC-MS/MS proteotyping, are applicable for resolving microorganisms over the entire taxonomic range indicated, that is, from the family to strain levels. (Adapted from Welker, M. and E. R. Moore (2011). "Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology." *Syst Appl Microbiol* **34**(1): 2–11).

an analysis method that adapts to the sample, requires no a priori information of suitable target genes and can be applied directly to clinical samples without prior cultivation. In one analysis, proteotyping can perform microbial species detection and identification and simultaneous characterization of metabolic and pathogenic features (Charretier, Dauwalder *et al.*, 2015; Charretier, Kohler *et al.*, 2015).

There are, however, a number of challenges related to the analysis of generated proteomic data. The random sampling of expressed proteins, in combination with limited genome sequence depth, often results in low degrees of overlap of detected peptides in replicate samples and low reproducibility. The most serious issue concerning the use of proteotyping is that it is still highly dependent on comprehensive sets of high-quality reference whole-genome sequences. Indeed, erroneous or missing reference data may

result in false matching and misidentification of peptides from the generated spectra or peptides with incorrect taxonomic affiliations. Such errors will introduce unwanted variation, that is, 'noise', which can result in substantial reductions in overall performance. The current reference databases, such as NCBI Genomes and PATRIC (Fricke and Rasko, 2014), are heavily biased towards known pathogens, whereas many closely related commensal species are underrepresented. Such biases have the potential to induce confusion for the reliable detection and identification of important pathogens. Ultimately, reference genomes from all bacteria typically encountered in clinical samples are necessary to ensure high sensitivity and specificity. Expanding the current reference databases by targeted sequencing of underrepresented species and strains is, therefore, an important task. Fortunately, the introduction of new NGS technologies is resulting in rapid decreases in the cost of sequencing of bacterial genomes; the genome reference databases are, therefore, expected to grow dramatically in the immediate future, greatly benefiting proteotyping applications and making the technique applicable to a larger set of clinically relevant species.

In addition to their comprehensiveness, the reference genome sequence databases need to contain data and metadata of as high quality as possible. Microorganisms with low-coverage draft genomes and incomplete assemblies can fail to include important coding sequence regions, thereby reducing the number of identified peptides. Proteotyping is also highly reliant on the existing taxonomic annotation, and errors in the metadata can propagate into the analysis and result in misclassifications. A related problem in correct microbial systematics and diagnostics is the relevance of mobile genetic elements (MGEs), such as plasmids, phages, conjugative elements and transposons, which may be transferred horizontally between evolutionarily distant organisms (Juhas, 2015; Soucy, Huang *et al.*, 2015). MGEs are common in the genomes currently present in the reference databases. Thus, peptides from expressed genes on MGEs potentially can result in decreases in the total number of discriminative peptides and, in the worst case, can result in false classifications and reduced overall performance. The public repositories are known to contain an unsatisfactory level of errors; to reach the full potential of proteotyping, novel databases that apply stricter quality assessment criteria may need to be developed, such that high-quality genome sequences of an effective number of strains from a comprehensive range of relevant species are included.

In order to bypass the issues with database-dependent peptide matching, it has been shown that strain-level bacterial differentiation is possible, using tandem mass spectra (raw data), instead of matching the spectra to existing databases, for analyzing *E. coli* isolates (Shao, Ma *et al.*, 2015).

16.14 Conclusion

Proteotyping provides important advantages, with the potential to be applied as a cultivation-independent method that simultaneously can perform microbial species detections and identifications, as well as characterization of metabolic and pathogenic features. It is foreseen that a major driver for the development and use of tandem MS and proteotyping in clinical settings will be the rapidly growing databases of whole-genome reference sequences, which will refine microbial phylogeny and provide a foundation for proteomics-based identification.

Acknowledgments

The authors acknowledge the European Commission and the project Tailored-Treatment (project number HEALTH-F3-602860-2013); andthe Swedish regional funding agency projects: Västra Götaland ALF-LUA (project numbers ALFGBG-210591; ALFGBG 437221; and VGFOUREG-232981) for support in the development and application of new MS-proteomics protocols to the analyses of microorganisms. The Proteomics Core Facility at The Sahlgrenska Academy of the University of Gothenburg (http://proteomics.cf.gu.se/Proteomics) is acknowledged for providing expertise and MS-proteomics analyses of microorganisms.

References

Makarov, A. (1999). Mass spectrometer. U.S. **5,886,346 1999.**

Aho, A. V., Hopcroft, J. E., & Ullman, J. D (1973). On finding lowest common ancestors in trees. *In Proceedings of the Fifth Annual ACM Symposium on Theory of Computing*: 253–265.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs." *Nucleic Acids Res* **25**(17): 3389–3402.

Anhalt j., F. C. (1975). "Identification of bacteria using mass spectrometry." *Anal. Chem.*, **47**(2): 219–225.

Armengaud, J. (2013). "Microbiology and proteomics, getting the best of both worlds!" *Environ Microbiol* **15**(1): 12–23.

Backert, S., T. Kwok, M. Schmid, M. Selbach, S. Moese, R. M. Peek, Jr., W. Konig, T. F. Meyer and P. R. Jungblut (2005). "Subproteomes of soluble and structure-bound Helicobacter pylori proteins analyzed by two-dimensional gel electrophoresis and mass spectrometry." *Proteomics* **5**(5): 1331–1345.

Banks, C. A. S., M.,Lakshminarasimhan and M. P. Washburn (2014). *Shotgun Proteomics*. John Wiley & Sons: Chichester, UK.

Bizzini, A., K. Jaton, D. Romo, J. Bille, G. Prod'hom and G. Greub (2011). "Matrix-assisted laser desorption ionization-time of flight mass spectrometry as an alternative to 16S rRNA gene sequencing for identification of difficult-to-identify bacterial strains." *J Clin Microbiol* **49**(2): 693–696.

Cai, J. C., Y. Y. Hu, R. Zhang, H. W. Zhou and G. X. Chen (2012). "Detection of OmpK36 porin loss in *Klebsiella* spp. by matrix-assisted laser desorption ionization-time of flight mass spectrometry." *J Clin Microbiol* **50**(6): 2179–2182.

Chao, T. C. and N. Hansmeier (2012). "The current state of microbial proteomics: Where we are and where we want to go." *Proteomics* **12**(4–5): 638–650.

Charretier, Y., O. Dauwalder, C. Franceschi, E. Degout-Charmette, G. Zambardi, T. Cecchini, C. Bardet, X. Lacoux, P. Dufour, L. Veron, H. Rostaing, V. Lanet, T. Fortin, C. Beaulieu, N. Perrot, D. Dechaume, S. Pons, V. Girard, A. Salvador, G. Durand, F. Mallard, A. Theretz, P. Broyer, S. Chatellier, G. Gervasi, M. Van Nuenen, C. Ann Roitsch, A. Van Belkum, J. Lemoine, F. Vandenesch and J. P. Charrier (2015). "Rapid bacterial identification, resistance, virulence and type profiling using selected reaction monitoring mass spectrometry." *Sci Rep* **5**: 13944.

Charretier, Y., T. Kohler, T. Cecchini, C. Bardet, A. Cherkaoui, C. Llanes, P. Bogaerts, S. Chatellier, J. P. Charrier and J. Schrenzel (2015). "Label-free SRM-based relative quantification of antibiotic resistance mechanisms in *Pseudomonas aeruginosa* clinical isolates." *Front Microbiol* **6**: 81.

Chi, A., D. L. Bai, L. Y. Geer, J. Shabanowitz and D. F. Hunt (2007). "Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry." *Int J Mass Spectrom* **259**(1–3): 197–203.

Chooneea, D., R. Karlsson, V. Encheva, C. Arnold, H. Appleton and H. Shah (2010). "Elucidation of the outer membrane proteome of *Salmonella enterica* serovar Typhimurium utilising a lipid-based protein immobilization technique." *BMC Microbiol* **10**: 44.

Christie-Oleza, J. A., B. Fernandez, B. Nogales, R. Bosch and J. Armengaud (2012). "Proteomic insights into the lifestyle of an environmentally relevant marine bacterium." *ISME J* **6**(1): 124–135.

Cohan, F. M. (2002). "What are bacterial species?" *Annu Rev Microbiol* **56**: 457–487.

Craig, R. and R. C. Beavis (2004). "TANDEM: Matching proteins with tandem mass spectra." *Bioinformatics* **20**(9): 1466–1467.

Dotsch, A., F. Klawonn, M. Jarek, M. Scharfe, H. Blocker and S. Haussler (2010). "Evolutionary conservation of essential and highly expressed genes in *Pseudomonas aeruginosa*." *BMC Genomics* **11**: 234.

Dreisewerd, K. (2003). "The desorption process in MALDI." *Chem Rev* **103**(2): 395–426.

Dumas, E., M. Desvaux, C. Chambon and M. Hebraud (2009). "Insight into the core and variant exoproteomes of Listeria monocytogenes species by comparative subproteomic analysis." *Proteomics* **9**(11): 3136–3155.

Durighello, E., L. Bellanger, E. Ezan and J. Armengaud (2014). "Proteogenomic biomarkers for identification of *Francisella* species and subspecies by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry." *Anal Chem* **86**(19): 9394–9398.

Dworzanski, J. P., S. V. Deshpande, R. Chen, R. E. Jabbour, A. P. Snyder, C. H. Wick and L. Li (2006). "Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification." *J Proteome Res* **5**(1): 76–87.

Dworzanski, J. P., A. P. Snyder, R. Chen, H. Zhang, D. Wishart and L. Li (2004). "Identification of bacteria using tandem mass spectrometry combined with a proteome database and statistical scoring." *Anal Chem* **76**(8): 2355–2366.

Eng, J. K., A. L. McCormack and J. R. Yates (1994). "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database." *J Am Soc Mass Spectrom* **5**(11): 976–989.

Fagerquist, C. K., A. H. Bates, S. Heath, B. C. King, B. R. Garbus, L. A. Harden and W. G. Miller (2006). "Sub-speciating *Campylobacter jejuni* by proteomic analysis of its protein biomarkers and their post-translational modifications." *J Proteome Res* **5**(10): 2527–2538.

Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse (1989). "Electrospray ionization for mass spectrometry of large biomolecules." *Science* **246**(4926): 64–71.

Fischer, B., V. Roth, F. Roos, J. Grossmann, S. Baginsky, P. Widmayer, W. Gruissem and J. M. Buhmann (2005). "NovoHMM: A hidden Markov model for de novo peptide sequencing." *Anal Chem* **77**(22): 7265–7273.

Fournier, M. L., J. M. Gilmore, S. A. Martin-Brown and M. P. Washburn (2007). "Multidimensional separations-based shotgun proteomics." *Chem Rev* **107**(8): 3654–3686.

Frank, A. and P. Pevzner (2005). "PepNovo: De novo peptide sequencing via probabilistic network modeling." *Anal Chem* **77**(4): 964–973.

Fricke, W. F. and D. A. Rasko (2014). "Bacterial genome sequencing in the clinic: Bioinformatic challenges and solutions." *Nat Rev Genet* **15**(1): 49–55.

Gault, J., C. Malosse, S. Machata, C. Millien, I. Podglajen, M. C. Ploy, C. E. Costello, G. Dumenil and J. Chamot-Rooke (2014). "Complete posttranslational modification mapping of pathogenic *Neisseria meningitidis* pilins requires top-down mass spectrometry." *Proteomics* **14**(10): 1141–1151.

Ha, J. W. and K. M. Downard (2011). "Evolution of H5N1 influenza virus through proteotyping of hemagglutinin with high resolution mass spectrometry." *Analyst* **136**(16): 3259–3267.

Hebraud, M. (2014). "Analysis of *Listeria monocytogenes* subproteomes." *Methods Mol Biol* **1157**: 109–128.

Henzel, W. J., T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley and C. Watanabe (1993). "Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases." *Proc Natl Acad Sci U S A* **90**(11): 5011–5015.

Hernandez, P., M. Muller and R. D. Appel (2006). "Automated protein identification by tandem mass spectrometry: Issues and strategies." *Mass Spectrom Rev* **25**(2): 235–254.

Hillenkamp F., K. M. (2000). "Matrix-assisted laser desorption/ionisation, an experience." *Int J Mass Spectrom* **200**: 71–77.

Hu, Q., R. J. Noll, H. Li, A. Makarov, M. Hardman and R. Graham Cooks (2005). "The Orbitrap: A new mass spectrometer." *J Mass Spectrom* **40**(4): 430–443.

Issaq, H. J., K. C. Chan, G. M. Janini, T. P. Conrads and T. D. Veenstra (2005). "Multidimensional separation of peptides for effective proteomic analysis." *J Chromatogr B Analyt Technol Biomed Life Sci* **817**(1): 35–47.

Jabbour, R. E., S. V. Deshpande, M. M. Wade, M. F. Stanford, C. H. Wick, A. W. Zulich, E. W. Skowronski and A. P. Snyder (2010). "Double-blind characterization of non-genome-sequenced bacteria by mass spectrometry-based proteomics." *Appl Environ Microbiol* **76**(11): 3637–3644.

Jabbour, R. E., M. M. Wade, S. V. Deshpande, M. F. Stanford, C. H. Wick, A. W. Zulich and A. P. Snyder (2010). "Identification of *Yersinia pestis* and *Escherichia coli* strains by whole cell and outer membrane protein extracts with mass spectrometry-based proteomics." *J Proteome Res* **9**(7): 3647–3655.

Jadhav, S., V. Gulati, E. M. Fox, A. Karpe, D. J. Beale, D. Sevior, M. Bhave and E. A. Palombo (2015). "Rapid identification and source-tracking of *Listeria monocytogenes* using MALDI-TOF mass spectrometry." *Int J Food Microbiol* **202**: 1–9.

Josten, M., J. Dischinger, C. Szekat, M. Reif, N. Al-Sabti, H. G. Sahl, M. Parcina, I. Bekeredjian-Ding and G. Bierbaum (2014). "Identification of agr-positive methicillin-resistant *Staphylococcus aureus* harbouring the class A mec complex by MALDI-TOF mass spectrometry." *Int J Med Microbiol* **304**(8): 1018–1023.

Josten, M., M. Reif, C. Szekat, N. Al-Sabti, T. Roemer, K. Sparbier, M. Kostrzewa, H. Rohde, H. G. Sahl and G. Bierbaum (2013). "Analysis of the matrix-assisted laser desorption ionization-time of flight mass spectrum of *Staphylococcus aureus* identifies mutations that allow differentiation of the main clonal lineages." *J Clin Microbiol* **51**(6): 1809–1817.

Juhas, M. (2015). "Horizontal gene transfer in human pathogens." *Crit Rev Microbiol* **41**(1): 101–108.

Kallow, W., M. Erhard, H. N. Shah, E. Raptakis and M. Welker (2010). MALDI-TOF MS for microbial identification: Years of experimental development to an established protocol. H. N. Shah and S. E. Gharbia (eds.), *Mass Spectrometry for Microbial Proteomics*. John Wiley & Sons: Chichester, UK, **12**: 255–276.

Karlsson, R., D. Chooneea, E. Carlsohn, V. Encheva and H. N. Shah (2010). Characterization of bacterial membrane proteins using a novel combination of a lipid based protein immobilization technique with mass spectrometry. H. N. Shah and S. E. Gharbia (eds.), *Mass Spectrometry for Microbial Proteomics*. John Wiley & Sons: Chichester, UK, 157–174.

Karlsson, R., M. Davidson, L. Svensson-Stadler, A. Karlsson, K. Olesen, E. Carlsohn and E. R. Moore (2012). "Strain-level typing and identification of bacteria using mass spectrometry-based proteomics." *J Proteome Res* **11**(5): 2710–2720.

Karlsson, R., L. Gonzales-Siles, F. Boulund, L. Svensson-Stadler, S. Skovbjerg, A. Karlsson, M. Davidson, S. Hulth, E. Kristiansson and E. R. Moore (2015). "Proteotyping: Proteomic characterization, classification and identification of microorganisms – a prospectus." *Syst Appl Microbiol* **38**(4): 246–257.

Kent, W. J. (2002). "BLAT – the BLAST-like alignment tool." *Genome Res* **12**(4): 656–664.

Klein, C., M. Aivaliotis, J. V. Olsen, M. Falb, H. Besir, B. Scheffer, B. Bisle, A. Tebbe, K. Konstantinidis, F. Siedler, F. Pfeiffer, M. Mann and D. Oesterhelt (2007). "The low molecular weight proteome of *Halobacterium salinarum*." *J Proteome Res* **6**(4): 1510–1518.

Lasch, P., C. Fleige, M. Stammler, F. Layer, U. Nubel, W. Witte and G. Werner (2014). "Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of *Enterococcus faecium* and *Staphylococcus aureus* isolates." *J Microbiol Methods* **100**: 58–69.

Lin, J., S. Huang and Q. Zhang (2002). "Outer membrane proteins: Key players for bacterial adaptation in host niches." *Microbes Infect* **4**(3): 325–331.

Lopez-Campistrous, A., P. Semchuk, L. Burke, T. Palmer-Stone, S. J. Brokx, G. Broderick, D. Bottorff, S. Bolch, J. H. Weiner and M. J. Ellison (2005). "Localization, annotation, and comparison of the *Escherichia coli* K-12 proteome under two states of growth." *Mol Cell Proteomics* **4**(8): 1205–1209.

Ma, B., K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby and G. Lajoie (2003). "PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry." *Rapid Commun Mass Spectrom* **17**(20): 2337–2342.

Macek, B., L. F. Waanders, J. V. Olsen and M. Mann (2006). "Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer." *Mol Cell Proteomics* **5**(5): 949–958.

Månsson, V., F. Resman, M. Kostrzewa, B. Nilson and K. Riesbeck (2015). "Identification of *Haemophilus influenzae* Type b Isolates by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry." *J Clin Microbiol* **53**(7): 2215–2224.

Markowitz, V. M., I. A. Chen, K. Chu, A. Pati, N. N. Ivanova and N. C. Kyrpides (2015). "Ten years of maintaining and expanding a microbial genome and metagenome analysis system." *Trends Microbiol* **23**(11): 730–741.

McDonald, W. H. and J. R. Yates, 3rd (2003). "Shotgun proteomics: Integrating technologies to answer biological questions." *Curr Opin Mol Ther* **5**(3): 302–309.

McLafferty, F. W., K. Breuker, M. Jin, X. Han, G. Infusini, H. Jiang, X. Kong and T. P. Begley (2007). "Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics." *FEBS J* **274**(24): 6256–6268.

Meyer, B., D. G. Papasotiriou and M. Karas (2011). "100% protein sequence coverage: A modern form of surrealism in proteomics." *Amino Acids* **41**(2): 291–310.

Moore, E. R., S. A. Mihaylova, P. Vandamme, M. I. Krichevsky and L. Dijkshoorn (2010). "Microbial systematics and taxonomy: Relevance for a microbial commons." *Res Microbiol* **161**(6): 430–438.

Ogorzalek Loo, R. R., R. Hayes, Y. Yang, F. Hung, P. Ramachandran, N. Kim, R. Gunsalus and J. A. Loo (2005). "Topdown, bottom-up, and side-to-side proteomics with virtual 2-D gels." *Int J Mass Spectrom* **240**: 317–325.

Olaya-Abril, A., I. Jimenez-Munguia, L. Gomez-Gascon and M. J. Rodriguez-Ortega (2014). "Surfomics: Shaving live organisms for a fast proteomic identification of surface proteins." *J Proteomics* **97**: 164–176.

Olsen, J. V., S. E. Ong and M. Mann (2004). "Trypsin cleaves exclusively C-terminal to arginine and lysine residues." *Mol Cell Proteomics* **3**(6): 608–614.

Organization, W. H. (2014). *Antimicrobial Resistance: Global Report on Surveillance*. WHO Press. Geneva, Switzerland.

Palys, T., L. K. Nakamura and F. M. Cohan (1997). "Discovery and classification of ecological diversity in the bacterial world: The role of DNA sequence data." *Int J Syst Bacteriol* **47**(4): 1145–1156.

Pappin, D. J., P. Hojrup and A. J. Bleasby (1993). "Rapid identification of proteins by peptide-mass fingerprinting." *Curr Biol* **3**(6): 327–332.

Penzlin, A., M. S. Lindner, J. Doellinger, P. W. Dabrowski, A. Nitsche and B. Y. Renard (2014). "Pipasic: Similarity and expression correction for strain-level identification and quantification in metaproteomics." *Bioinformatics* **30**(12): i149–156.

Perkins, D. N., D. J. Pappin, D. M. Creasy and J. S. Cottrell (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." *Electrophoresis* **20**(18): 3551–3567.

Pieper, R., S. T. Huang, J. M. Robinson, D. J. Clark, H. Alami, P. P. Parmar, R. D. Perry, R. D. Fleischmann and S. N. Peterson (2009). "Temperature and growth phase influence the outer-membrane proteome and the expression of a type VI secretion system in Yersinia pestis." *Microbiology* **155**(Pt 2): 498–512.

Rizzardi, K. and T. Akerlund (2015). "High Molecular Weight Typing with MALDI-TOF MS – A Novel Method for Rapid Typing of Clostridium difficile." *PLoS ONE* **10**(4): e0122457.

Rumbo, C., E. Gato, M. Lopez, C. Ruiz de Alegria, F. Fernandez-Cuenca, L. Martinez-Martinez, J. Vila, J. Pachon, J. M. Cisneros, J. Rodriguez-Bano, A. Pascual, G. Bou, M. Tomas, I. Spanish Group of Nosocomial, A. Mechanisms of, A. Resistance to, M. Spanish Society of Clinical, D. Infectious and D. Spanish Network for Research in Infectious (2013). "Contribution of efflux pumps, porins, and beta-lactamases to multidrug resistance in clinical isolates of Acinetobacter baumannii." *Antimicrob Agents Chemother* **57**(11): 5247–5257.

Shah, H. N., L. Rajakaruna, G. Ball, R. Misra, A. Al-Shahib, M. Fang and S. E. Gharbia (2011). "Tracing the transition of methicillin resistance in sub-populations of *Staphylococcus aureus*, using SELDI-TOF Mass Spectrometry and Artificial Neural Network Analysis." *Syst Appl Microbiol* **34**(1): 81–86.

Shao, J., S. Ma, D. Zheng, W. Chen and Y. Luo (2015). "Simultaneous determination of four active components in rat plasma by ultra-high performance liquid chromatography tandem-mass spectrometry/mass spectrometry and its application to a pharmacokinetic study after oral administration of Callicarpa nudiflora extract." *Pharmacogn Mag* **11**(43): 509–517.

Solis, N. and S. J. Cordwell (2011). "Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins." *Proteomics* **11**(15): 3169–3189.
Soucy, S. M., J. Huang and J. P. Gogarten (2015). "Horizontal gene transfer: Building the web of life." *Nat Rev Genet* **16**(8): 472–482.
Spinali, S., A. van Belkum, R. V. Goering, V. Girard, M. Welker, M. Van Nuenen, D. H. Pincus, M. Arsac and G. Durand (2015). "Microbial typing by matrix-assisted laser desorption ionization-time of flight mass spectrometry: Do we need guidance for data interpretation?" *J Clin Microbiol* **53**(3): 760–765.
Strateva, T. and D. Yordanov (2009). "*Pseudomonas aeruginosa* – a phenomenon of bacterial resistance." *J Med Microbiol* **58**(Pt 9): 1133–1148.
Swaney, D. L., C. D. Wenger and J. J. Coon (2010). "Value of using multiple proteases for large-scale mass spectrometry-based proteomics." *J Proteome Res* **9**(3): 1323–1329.
Tabb, D. L., C. G. Fernando and M. C. Chambers (2007). "MyriMatch: Highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis." *J Proteome Res* **6**(2): 654–661.
Taylor, J. A. and R. S. Johnson (1997). "Sequence database searches via de novo peptide sequencing by tandem mass spectrometry." *Rapid Commun Mass Spectrom* **11**(9): 1067–1075.
Thelen, J. J. and J. A. Miernyk (2012). "The proteomic future: Where mass spectrometry should be taking us." *Biochem J* **444**(2): 169–181.
Tracz, D. M., S. J. McCorrister, P. M. Chong, D. M. Lee, C. R. Corbett and G. R. Westmacott (2013). "A simple shotgun proteomics method for rapid bacterial identification." *J Microbiol Methods* **94**(1): 54–57.
Tran, J. C., L. Zamdborg, D. R. Ahlf, J. E. Lee, A. D. Catherman, K. R. Durbin, J. D. Tipton, A. Vellaichamy, J. F. Kellie, M. Li, C. Wu, S. M. Sweet, B. P. Early, N. Siuti, R. D. LeDuc, P. D. Compton, P. M. Thomas and N. L. Kelleher (2011). "Mapping intact protein isoforms in discovery mode using top-down proteomics." *Nature* **480**(7376): 254–258.
van Belkum, A., G. Durand, M. Peyret, S. Chatellier, G. Zambardi, J. Schrenzel, D. Shortridge, A. Engelhardt and W. M. Dunne, Jr. (2013). "Rapid clinical bacteriology and its future impact." *Ann Lab Med* **33**(1): 14–27.
Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E., Starr, M. P. and Trüper, H. G. (1987). "Report of the ad hoc committee on reconciliation of approaches to bacterial systematics." *Int. J. System. Bacteriol* **37**: 463–464.
Welker, M. (2011). "Proteomics for routine identification of microorganisms." *Proteomics* **11**(15): 3143–3153.
Welker, M. and E. R. Moore (2011). "Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology." *Syst Appl Microbiol* **34**(1): 2–11.
Whitehouse, C. M., R. N. Dreyer, M. Yamashita and J. B. Fenn (1985). "Electrospray interface for liquid chromatographs and mass spectrometers." *Anal Chem* **57**(3): 675–679.
Whitelegge, J. (2013). "Intact protein mass spectrometry and top-down proteomics." *Expert Rev Proteomics* **10**(2): 127–129.
Williams, T. L., S. R. Monday, P. C. Feng and S. M. Musser (2005). "Identifying new PCR targets for pathogenic bacteria using top-down LC/MS protein discovery." *J Biomol Tech* **16**(2): 134–142.

Wynne, C., N. J. Edwards and C. Fenselau (2010). "Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC-MS/MS on an Orbitrap." *Proteomics* **10**(20): 3631–3643.

Yahara, K., M. Kawai, Y. Furuta, N. Takahashi, N. Handa, T. Tsuru, K. Oshima, M. Yoshida, T. Azuma, M. Hattori, I. Uchiyama and I. Kobayashi (2012). "Genome-wide survey of mutual homologous recombination in a highly sexual bacterial species." *Genome Biol Evol* **4**(5): 628–640.

Yates, J. R., C. I. Ruse and A. Nakorchevsky (2009). "Proteomics by mass spectrometry: Approaches, advances, and applications." *Annu Rev Biomed Eng* **11**: 49–79.

Zhang, J., L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie and B. Ma (2012). "PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification." *Mol Cell Proteomics* **11**(4): M111 010587.

17

Proteogenomics of *Pseudomonas aeruginosa* in Cystic Fibrosis Infections

Liang Yang[1] *and Song Lin Chua*[2]

[1] *Singapore Centre for Environmental Life Sciences Engineering (SCELSE), Nanyang Technological University; and School of Biological Sciences, Nanyang Technological University, Singapore*
[2] *Singapore Centre for Environmental Life Sciences Engineering (SCELSE), Nanyang Technological University; and Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore*

17.1 Introduction: *Pseudomonas aeruginosa* as a Clinically Important Pathogen

P. aeruginosa is a gram-negative gamma-proteobacterium which is found in varying environmental conditions, ranging from soil, marine habitats, plants and animals. It has a relatively large genome size of 6.3 Mb, with a significantly high proportion of genes dedicated to regulatory genes that allow it to survival in changing environments [1].

However, its metabolic versatility enables it to become an opportunistic pathogen to infect human tissues [2]. *P. aeruginosa* can attach and thrive on most surfaces, which enables it to inhabit medical instruments such as catheters and surgical tools, resulting in a prevalent rate of nosocomial infections in hospital settings [3, 4].

P. aeruginosa has been known to cause acute and chronic infections in humans. Acute infections are usually inflammatory and septic when *P. aeruginosa* infects damaged tissues, especially in burn wounds [5], cornea infections [6], septicaemia [7] and meningitis [8]. On the other hand, chronic *P. aeruginosa* infections tend to affect patients with compromised immunity [9] and cystic fibrosis (CF) patients [10, 11].

Moreover, *P. aeruginosa* contains multi-drug-resistant efflux pumps and beta-lactamase, which accounts for its intrinsic resistance to most antibiotics. Hence, *P. aeruginosa* was declared one of the six 'top priority dangerous, drug-resistant microbes' by the Infectious Diseases Society of America [12].

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

17.2 CF and Pathophysiology

CF is a genetic disorder which is autosomal recessive, caused by mutations in the cystic fibrosis transmembrane conductor regulator (CFTR) gene [13]. This is especially common in the Caucasian population, affecting 1 in 2500 [14]. It is characterized by abnormal flow of chloride ions and water across cell membranes, resulting in the production of thick and sticky mucus [15,16]. This is also clinically relevant in the lungs as mucus clogs up the airway to the lungs and reduces lung function, often leading to premature death [17].

17.3 CF Infections

The accumulation of mucus and poor clearance can result in the retention of inhaled bacteria. Infection and inflammation of the airways are highly common in CF patients, thereby exacerbating breathing difficulties and eventual lung failure. Hence, the lifespan of an average CF patient undergoing the right treatment is around 40 years [18].

Together with *Staphylococcus aureus*, *Burkholderia cepacia* and *Haemophilus influenza*, *P. aeruginosa* is one of the most common pathogens in CF patients (Figure 17.1) [19]. However, while the rest are predominant in early-stage CF infections, *P. aeruginosa* tends to become the major pathogen in late-stage CF infections [20,21]. The capacity of *P. aeruginosa* to cause long-term chronic infections can be attributed to its ability to form robust biofilms in the CF lung [3,22]. This raises the question of how *P. aeruginosa* adapts to the host environment during the CF infections. An overview of the adaptation and pathogenesis of *P. aeruginosa* in the CF lung will be explained in the following sections.

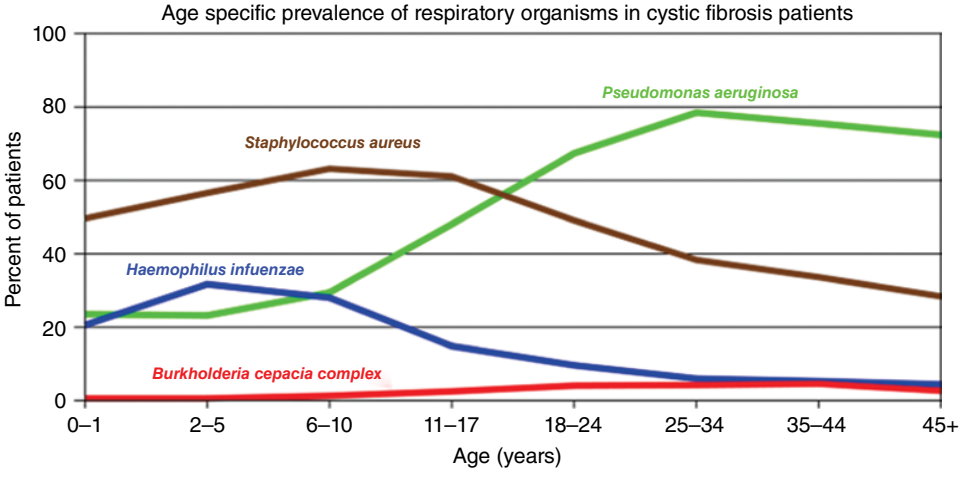


Figure 17.1 Cystic fibrosis lung infections caused by different species of pathogens. Whereas *S. aureus* is predominant in the early stages of infections, *P. aeruginosa* is the most dominant pathogen in late stages of infections. Figure adapted from the 2009 Patient Registry Report, issued by the Cystic Fibrosis Foundation, with permission [23].

17.4 Biofilm Formation in *P. aeruginosa*

Bacterial cells are able to form surface-attached biofilms consisting of microcolonies and their secreted extracellular polymeric substances (EPSs). EPSs mainly include exopolysaccharide, adhesion proteins and extracellular DNA (eDNA), which serves as a physical shield to protect cells in biofilms. The development of biofilm lifestyle is crucial in the adaptation of *P. aeruginosa* in the CF lung infections [24]. Whereas *P. aeruginosa* commonly produces Pel and Psl exopolysaccharides during biofilm formation in early acute infections, alginate plays a more significant role in the biofilm formation in chronic persistent infections in the CF lung. Mucoid clinical isolates from the sputum of most CF patients over-produce alginate, which acts as a virulence factor selected for long-term inhabitation of the CF lung [25].

There are several conditions that are proposed to promote the evolution of mucoidy and biofilm formation of *P. aeruginosa* in CF lung infections, as stated below:

1) Oxidative stress: The production of reactive oxygen species (ROS) and reactive nitrogen intermediates (RNI) by the host immune system, such as polymorphonuclear leukocytes (PMNs) [26], causes high levels of oxidative stress that the bacteria cannot withstand, resulting in damage to the DNA and proteins [27]. The DNA damage can easily result in higher rates of genome mutation and select for variants that can survive better in this environment. Alginate produced by the mucoid strains was shown to chelate ROS, further protecting the bacterial cells from oxidative damage [28].
2) Immune system: Alginate can protect bacterial cells from phagocytosis by phagocytes such as macrophages and neutrophils by enveloping the bacterial cells [29]. It can also inhibit the activation of complements [30]. In addition, the *P. aeruginosa* biofilms are able to activate a rhamnolipid 'shield' that is highly cytotoxic to the immune cells [31].
3) Antibiotic treatment: Antibiotic resistance and tolerance are common as CF patients are often treated with large amounts of antibiotics throughout their lives [32]. *P. aeruginosa* evolves and adapts to various classes of antibiotics such as antimicrobial peptides [33], aminoglycosides [34] and beta-lactams [35]. EPS (e.g. alginate) from the biofilms acts as a physical barrier that prevents the effective dose of antibiotics from reaching the bacterial cells [36,37]. The high phenotypic diversity of a biofilm also ensured the presence of dormant persisters with low metabolic activity while maintaining a high level of antibiotic tolerance [38].

Conversion to mucoidy occurs when the *mucA* gene, which encodes a negative regulator of alginate synthesis, is mutated [39]. As an anti-sigma factor, MucA is able to inhibit AlgU, a sigma factor required for the expression of the *algD* operon that synthesizes alginate [40]. AlgU is also instrumental in controlling the expression of virulence and motility genes that are important in biofilm formation [41]. Mutation of the *mucA* gene is highly common in CF clinical isolates, with up to 80% of patients possessing strains of such traits [42].

Hence, with EPS acting as a barrier, biofilms are much more tolerant to antimicrobial agents and host immune systems as compared to planktonic cells [43]. The poor antibiotic penetration [44] and presence of slow growing cells [45] and phenotypic varied cells [46] all contribute to the positive selection of biofilm phenotypes during long-term survival of *P. aeruginosa* in the CF lung.

17.5 Virulence of *P. aeruginosa*

Other than forming biofilms to withstand stressful host conditions, *P. aeruginosa* also produces a plethora of virulence factors that target the host cells. *P. aeruginosa*'s virulence depends on various secreted and cell-surface-associated factors and can be utilized differently in acute and chronic infections. The virulence factors are controlled by several regulatory systems, including the quorum sensing (QS) systems and two-component sensor kinases [47].

QS is an important cell-to-cell communication mechanism used by most bacterial species to respond to population density and synchronize their group behaviour [48]. This is mediated by autoinducers (QS signal molecules) produced by bacteria and accumulated as the population density increases [48]. Upon reaching the threshold, autoinducers bind to their specific receptors, allowing the coordination of expression of a large set of genes at the transcriptional level [49]. There are three QS systems in *P. aeruginosa*: *las*, *rhl* and *pqs* (Figure 17.2). Both *las* and *rhl* systems are N-acyl homoserine lactone [50] based, whereas the *pqs* system is 2-alkyl-4-quinolone based [51]. The *las* system positively regulates downstream *rhl* and *pqs* systems [50,52,53]. However, the *rhl* system regulates the *pqs* system negatively [53]. The *pqs* system also regulates both the *las* and *rhl* QS systems [54].

During acute infections, QS activates the expression of virulence factors, allowing the *P. aeruginosa* cells to overwhelm the host defences. The *las* QS-regulated elastase is a peptidase that can cause proteolytic damage to host tissues and disrupt tight junctions between cells [55,56]. The *rhl* QS-regulated rhamnolipids are highly cytotoxic to host cells because they disrupt the cell membrane and cause cytosol leakage [31]. The *pqs* QS-regulated pyocyanin can generate free radicals that damage host cells, allowing the bacteria to persist in the CF airways [57–61].

QS systems regulate motility to *P. aeruginosa*, allowing it to cause acute infections [62]. *P. aeruginosa* possesses a single polar flagellum for swimming and swarming motility, and type IV pili for twitching motility [63,64]. The polymeric flagellum is made up of monomers of flagellin (FliC), and there are many gene products involved in flagellar formation and function [65]. Flagella can confer resistance to antimicrobial proteins such as pulmonary surfactant protein A in lungs [66]. Flagellin is pro-inflammatory, and thus can stimulate inflammation and mucus secretion [67], while impairment of mucus secretion in CF patients results in retention of bacteria and further infection [68].

The QS systems are also involved in the progression to chronic infections in CF airways, with QS molecules being detected in the sputum [24]. QS regulates the formation of biofilms and secretion of virulence factors involved in chronic infections, such as pyoverdine and pyochelin, which are important siderophores in iron uptake [69,70], and the type VI secretion system (T6SS) [71]. The upregulation of T6SS (*hcp* operon) in biofilms enables *P. aeruginosa* cells to kill other bacterial species and host cells via the secretion of Tse1-3 and Tse2 [72,73].

The two-component sensor kinases control the expression of virulence factors in acute and chronic infections post-transcriptionally via the alteration of mRNA [74]. Examples of such kinases are the GacS, RetS (regulator of exopolysaccharides and T3SS) [75] and LadS (loss adherence sensor) [76], with the GacS/GacA two-component system (TCS) being central to the pathway.

The GacS/GacA TCS is instrumental in controlling T3SS and T6SS expression, and exopolysaccharide production [75,77–79]. It represses T3SS and motility, but upregulates T6SS and polysaccharide genes expression, to induce biofilm formation [75]. When activated, the GacA induced the expression of RsmY and RsmZ sRNAs [80]. In turn, RsmY and RsmZ sequester the RsmA protein at high affinity [81], allowing the induction of T6SS and EPS, and repression of T3SS and motility. These genes are also QS-related as they positively induced the production of QS autoinducers, pyocyanin, cyanide and lipase [82–84].

In the ever-changing CF lung environment, *P. aeruginosa* has to respond by altering its gene expression and downstream protein expression. Hence, it is imperative to study the global gene and protein expressions, which can reveal a 'glimpse' of the response at a specific time and condition. This can help to determine the specific genes/proteins important to the chronic adaptation and persistence. Genomics and proteomics are now widely used by researchers to elucidate the global gene/protein expression and adaptive evolution of *P. aeruginosa* in the CF lung.

17.6 Genomics to Study Bacterial Pathogenesis

Technological advances in whole genome sequencing in terms of depth and resolution greatly facilitate the research of bacterial evolution and virulence, which further provides information for identifying new drug targets, modifying clinical practices and health policies. Past approaches were merely based on sequencing small fragments such as loci of the bacterial genome and could only tell part of a story on virulence [85]. However, high-throughput genome sequencing and population genomics are gaining traction in recent years to study population evolution and dynamics, due to higher sensitivity, real-time performance and speed. This enables the genomes to be compared at the single-nucleotide level.

Bacterial genome sequence reads and assemblies cannot be easily generated with the use of various sequencing platforms, such as Illumina, 454 and IonTorrent [86–88]. This allowed the lining of overlapping short sequences together into contiguous sequences, named contigs [89]. The assembled contigs were ordered against a reference genome of a bacterial species, so that the draft genes can be annotated to various functional classes [90]. Finally, pairwise comparative genome analysis can be employed to compare sequences between genomes, for identification of genes that could be important in virulence or antibiotic resistance [91].

Genomics studies conducted on different bacterial species include the comparison between wild-type, mutant and clinical isolates [92]; the study of evolution and spread of antibiotics resistance [93]; and the investigation of epidemics [94,95].

By sequencing entire populations across their genome, the following can be investigated:

1) Evolution within host: To study the evolution dynamics via genetic variation and mutations in the pathogenic bacteria. It also shows how determining factors in the host drive bacteria evolution
2) Transmission of disease: To understand the spread of pathogens in an epidemic and the virulence factors important in transmission
3) Population dynamics: To study the changes in the genomes of all microorganisms in the infection site in real time.

17.7 Proteomics to Study Bacterial Pathogenesis

Although genome analysis can provide valuable insights into gene expressions and mutations, proteomics can provide information on protein expression, dynamics and abundance on a global level. Proteomics is important because genomics does not give a clear indication of whether the mRNA encoded by a gene can be translated into a functional protein. For instance, high mRNA levels may not necessarily correspond to high protein production due to post-transcriptional modification by small non-coding RNAs. Genes subjected to alternative splicing and other modifications further contribute to the complexity of proteomic analysis.

Early studies of proteomics often rely on two-dimensional gel electrophoresis (2-DE) to separate proteins into individual spots and their subsequent identification by mass spectrometry [96]. This is usually limited by the depth of coverage and abundance of proteins. However, with the advent of shotgun mass-spectrometry-based proteomics, important techniques such as matrix- assisted laser desorption/ionization time-of-flight (MALDI-TOF) [97] and isobaric tag for relative and absolute quantification (iTRAQ) [98] have been developed to increase sensitivity, resolution, robustness and data processing.

Hence, increased quality of data from proteomics can be used to integrate genomic analysis via bioinformatics, as genome databases can be used as a reference for protein functions and characteristics. Such complementation of both technologies is important in the field of systems biology, acting as a useful tool for data-driven hypothesis. Up to now, many bacterial proteomes have been analyzed [99–101]. However, proteomics would not be possible without the complete mapping of the bacterial genome, which provided the basis of possible gene products (proteins). The availability of the relatively large but complete genome sequences of *P. aeruginosa* allows large-scale proteomic studies to be carried out to study its wide variety of virulence factors and adaptability to different host environments.

Genomics analysis indicated that many open reading frames (ORFs) and hypothetical genes had yet to be characterized in *P. aeruginosa* [102]. Proteomics can be used to confirm that the existence of those proteins is no longer hypothetical [103]. Proteins with hypothetical functions can also be predicted to be associated with certain subcellular locations, pathways and function classes on the basis of their homology of gene sequences from other species, thus providing clues about their true functions [103].

Proteomics has the additional advantage over genomics in its ability to characterize proteins from various localized factions, such as secreted proteins and membrane proteins [104]. These two classes of proteins are important in bacterial pathogenesis, as they are mostly involved in export of virulence factors and EPS, adherence and motility. Proteomics can also be used to study the turnover and degradation rate of proteins. A high turnover rate of a protein might suggest it can only function for a short period, even if it is highly expressed.

In the study of how *P. aeruginosa* adapts to its host environment or become resistant to antibiotics, proteomics can usually be used to show response to changes in environment, such as antibiotic treatment and survival in the host. Proteins that respond and change in abundance can be identified for downstream molecular and

functional studies. Over-expression or knockouts of genes can then be used concurrently with functional assays such as motility, cytotoxicity and toxicity studies for validation of proteomic findings.

However, there are certain limitations in proteomics which make the complementary support of genomics or transcriptomics necessary. Information on non-coding genes or small RNAs (sRNAs) that do not encode for any protein but with important physiological functions will be lost by proteomics. For instance, information on RsmY and RsmZ sRNAs, which are important in biofilm formation and dispersal [105], cannot be captured by proteomics. Hence, combining genomics or transcriptomics can provide hints to changes in downstream proteins in proteomics.

Unlike genes, which can be amplified by polymerase chain reaction (PCR), the absence of any amplification technique will require larger samples of proteins to be extracted. Fortunately, with the success of MS techniques, lower concentrations of proteins were required, and low copy proteins could be detected [106,107].

In the following text, we will discuss how both genomic and proteomic approaches can be utilized in studying the pathogenesis of *P. aeruginosa* in CF infections.

17.8 Genomics of *P. aeruginosa* in CF Infections

The first complete genome sequence of *P. aeruginosa* is the genome of the PAO1 strain, a wound isolate [1], which now serves as the major reference strain for most studies on *P. aeruginosa* [103]. On top of its core genome, which is highly conserved, *P. aeruginosa* also has accessory genomes of various magnitudes which are not conserved across strains [108]. The accessory genome consists of plasmids and genomic islands inserted into the chromosome at various locations by horizontal gene transfer (HGT). These accessory genomes are also termed as regions of high plasticity as they are unique to individual strains and contain at least four contiguous ORFs [108]. Ranges of accessory genomes could vary from PAO1 that merely contains small inserts of maximum 14 kbp [108] to the LESB58 clinical isolate that possesses multiple large genomic islands and prophages [109].

Hence, it is interesting to note that the *P. aeruginosa* retains its core genome and customizes its accessory genome to adapt to specific environments. This is unlike other species of pathogenic bacteria that usually undergo genomic reduction to fit its habitat [108].Other bacteria such as gamma-proteobacteria usually acquire each genomic island by HGT [110]. The addition of new genetic sequences of each strain will increase the pool of genes for *P. aeruginosa*. The total gene pool represents the pangenome of *P. aeruginosa*, which consists of the core genome that is indispensable in all *P. aeruginosa* strains and 'dispensable' ones that are found in two or more strains and 'unique' ones that can only be found in one strain [111]. This implies that *P. aeruginosa* has a large access to the gene pools of different species of bacteria, allowing it to adapt to varying environments.

Other than acquisition of new genomic islands, there are other factors, such as inversions [112,113], deletions [114] of genes that affect the plasticity of the genome. Single-nucleotide substitutions can occur [115] and are classified into two types: synonymous (replacement of a nucleotide does not change the amino acid being coded) or non-synonymous (replacement of one nucleotide changes the amino acid being translated and potentially changes the protein function). One study compared the

genomes of same *P. aeruginosa* PA14 clonal complex from different habitats showed that most single-nucleotide polymorphisms (SNPs) did not occur randomly and were found in the accessory genome instead of in the core genome [115].

17.9 Interclonal Genome Diversity

Genome sequencing of clinical isolates of different clonal complexes had shown that their genetic variation ranged approximately 0.5% in their core genomes [116]. A clonal complex refers to a group of isolates that are different from one another at one locus out of seven being tested by multi-locus sequence typing (MLST) analysis [117]. This indicated that the core genomes are highly conserved with low sequence diversity. However, as the accessory genome is highly variable, it accounted for the majority of the variations in the genome. The genomic islands that undergo most variations are involved in the biosynthesis of pyoverdine, an iron siderophore, and flagellar motility [116]. This confers specific phenotypes that offer survival advantage in certain environments. For example, mutation in the *rpoN* gene is often found during different clonal lineages of *P. aeruginosa* around the world [118,119], which results in deficiency of flagellum expression and escape of *P. aeruginosa* from phagocytosis [120].

Although it is well established that there is high interclonal diversity in *P. aeruginosa*, a study of 240 strains obtained from different environmental habitats and patients pointed to the fact that most strains belonged a few dominant clones, implying the prevalence of those clones in the environment and disease [121]. The *P. aeruginosa* PA14 clone is found to be the most widespread linage in both environment and disease settings, with a high expression level of ExoU, a T3SS-secreted cytotoxin [121,122]. Other common clonal lineages include Clone C [123] and M [124]. Although each clone has a certain 'preference' for different accessory segments, there are certain segments that are more likely to be varied, resulting in interclonal diversity.

17.10 Intraclonal Genome Diversity

Intraclonal genome diversities of strains within the same clonal complex and/or isolated within the same patient are also studied by using comparative genomic approaches. For instance, the intraclonal genomes between isolates from the *P. aeruginosa* clonal complex PA14 over 20 years were compared by parallel sequencing in the Hanover CF clinic [125]. It was observed that the PA14 in one patient had diversified into three different clades in the first five years of infection in the patient, but only one of them continued to accumulate changes in genome sequence such as deletions over time [125]. This implied that high diversity can still occur in the same clone from the same patient.

On the other hand, comparison between isolates from two CF patients infected with *P. aeruginosa* clone TB in the same local outbreak showed subtle differences in the genome, with one isolate possessing an additional genomic island from other bacterial species, thus providing the evidence for microevolution [126]. Moreover, transcriptomic and metabolomic profiling proved a huge difference between both isolates in various phenotypes important in the response to host environment [126]. Surprisingly, it was found that one isolate was more susceptible to immune killing than the other.

A high level of within-the-population diversity can also be observed in evolution experiments, regardless of the ancestor, environment and time span, with stable phenotypes emerging over time [127]. Hence, the ability of the species to diversify is an innate ability to promote survival in the harsh and ever-changing environments.

17.11 Clonal Spread of *P. aeruginosa* in CF Patients

With extensive microevolution and genetic adaptations, isolates that are highly successful in the survival and colonization in the CF lungs, are eventually selected. Interestingly, clones that are highly transmissible among CF patients with long-term chronic infections and aggressive within the lungs appeared [128–131]. Although it was previously thought that most CF patients got *P. aeruginosa* infections from the environment [132–134], genomic sequencing of different clinical isolates had instead suggested the existence of epidemics. One important example of clinical relevance is the Liverpool epidemic strain (LES) [135], which spread to many countries in Europe as well as the United States and Canada [136].

The *P. aeruginosa* LES strains were highly resistant to most antibiotics, further implicating treatment selects for dominant clones [137]. Comparative transcriptomic analysis of LES isolates revealed mutations in efflux pump genes and upregulation of AmpC beta-lactamase [138].

In accordance with the fact that patients infected with LES strains were found to have poor prognosis with a higher risk of death and lung transplantation, it was shown that LES isolates were more virulent than the non-epidemic isolates [139]. Although the LES clone transmitted across patient populations, intraclonal diversity still existed between isolates of same patients, as they could exhibit varied virulence to host cells [138,140]. Some of the LES isolates produced higher levels of QS-regulated virulence factors, such as pyocyanin and LasA protease, which were extremely toxic to host immune and epithelial cells of the airway [141]. Although LES was usually found to be non-motile [138], motile variants could still be isolated [142].

A recent comprehensive comparative genomics study of various LES strains from different patients of different geographic origin had shown high variability in the prophage and genomic islands in accessory genome among isolates [142]. With most LES isolates being non-mucoid and virulent [143], the LES clonal complex is indeed unique from the isolates that follow the traditional route of becoming mucoid and avirulent in chronic infections. Hence, one has to be careful in studying genomics from clinical isolates, with variations between and within clonal complexes possibly obfuscating data interpretation and delaying suitable treatments.

17.12 Parallel Evolution

Parallel evolution involves the independent evolution of replicate lineages to possess similar traits when grown in similar environments [144]. During the chronic CF lung infections, even with high interclonal and intraclonal genome diversities, the evolution of *P. aeruginosa* usually follows a predictable and distinct pattern. Via genotyping, the initial stages of CF lung infections were characterized by recurrent acute

infections which can be treated easily with antibiotics. However, microevolution of *P. aeruginosa* resulted in a chronic infection state, featuring mucoidy and antibiotic resistance. It is interesting to note that similar changes in gene expression can be observed in the transition from early clinical isolates to isolates from chronic infections [119,145].To study the differences in genotype of *P. aeruginosa*, SNP genotyping with AT biochips (Clondiag Chip Technologies, Germany), based on selected genes from highly conserved regions, and genome sequencing were often used. Other methods are described in Chapter 21C.

17.13 Mutations in Early-Stage CF *P. aeruginosa* Isolates

Despite having parallel evolution in the genetic changes among clinical isolates, high levels of phenotypic variation of *P. aeruginosa* still exist in each patient's sputum, as each isolate still possess different colony morphologies [146]. There are various theories explaining such genetic diversity in *P. aeruginosa* CF populations. Firstly, 'cheaters' that do not produce 'public goods' are often evolved and selected from the *P. aeruginosa* populations. Public goods can include QS signal molecules and siderophores [147,148]. Social cheaters with loss-of-function mutations in QS regulator *lasR* and production of QS autoinducers were found to be prevalent in acute and chronic infections, allowing them to take advantage of the rest of the group [149,150].

Hypermutator strains also contribute to the high phenotypic diversity of *P. aeruginosa* CF isolates, due to their enhanced and spontaneous mutation rates [151,152]. Mutations in DNA mismatch repair and error prevention genes such as *mutS*, *mutL* and *mutY* often contribute to the appearance of hypermutators, which accelerate evolution and shape a distinct evolutionary path in CF infections [153,154]. Hence, despite the capabilities of genomic sequencing to identify parallel evolution in phenotypes among clinical isolates, it is emphasized that one must take into consideration of the high levels of genotypic and phenotypic diversity in the same patient. It is essential to sample multiple isolates from the sample time point of a patient when studying *P. aeruginosa* CF adaptation.

Although it is known that chronic *P. aeruginosa* CF isolates possessed parallel evolved phenotypes, the evolutionary path taken by *P. aeruginosa* from early to late infections were not well understood. Tracking of genomic evolution of isolates over time from the same patients serves as the key approach to elucidate which genes will be subjected to mutational changes first under the CF environment. Smith *et al.* had tracked the evolution of a *P. aeruginosa* clonal complex in a single patient over eight years, sampling and sequencing the isolates longitudinally [118]. This allowed the identification of mutations of genes accumulated in the late isolate, as compared to the early isolate.

One of the earliest and most common mutations of *P. aeruginosa* CF isolates occurs in the LasR, the important transcriptional regulator of QS, which can even be found in clinical isolates from two-year-old CF patients [155]. This is also confirmed with phenotypic analyses showing that these isolates are deficient in QS autoinducer production [156]. Other than being important in social cheating, loss of function in LasR confers a growth advantage in the presence of specific amino acids, especially phenylalanine, as CbrB, a catabolism regulator, is upregulated in *lasR* mutant [157]. The CbrA/CbrB is a TCS that balances carbon and nitrogen metabolism, and prevent excess ammonia

build-up from nitrogen assimilation [158]. A similar observation is also reported showing that *P. aeruginosa lasR* mutant grows better than wild type in rich medium at stationary phase [159], implying the catabolic advantage of *lasR* mutants in enhancing viability. When an experimental *P. aeruginosa* PA14 strain was grown in CF sputum, amino acid synthesis, degradation and transport genes were induced [160]. Hence, it is highly likely that amino acid metabolism can drive *lasR* mutations.

Furthermore, *P. aeruginosa lasR* mutant downregulates the *pqs* QS system, which is required for autolysis and DNA release [161]. The sputum from CF patients was found to contain high concentrations of amino acids, notably phenylalanine and leucine [162], which was reported to activate *pqs*QS [163] and might cause autolysis of a substantial amount of cells with functional *LasR* but not in cells with *lasR* mutations. Other studies suggest that LasR-regulated virulence factors such as elastase are selected against by the immune system in the CF lung, resulting in attenuated virulence in LasR mutants [164].

Mutations in the alternative sigma factor RpoN and PvdS also happen in early-stage CF *P. aeruginosa* isolates, and it is often these mutations are maintained during long-term CF adaptation [118,165]. Mutations in the *rpoN* gene downregulate *P. aeruginosa* motility [166] and have pleiotropic effects on its central metabolism and QS [167]. RpoN mutants are deficient in surface pili, flagella, and nonpilus adhesins and are thus resistant to ingestion by both macrophages and neutrophils [168]. Mutations in PvdS as well as other genes involved in *P. aeruginosa* iron siderophore production are highly selected in early-stage CF isolates [119]. A recent study showed that mutation in *P. aeruginosa* siderophore-mediated iron uptake can maintain its mucoid phenotype due to alginate overproduction [169].

17.14 Mutations in Late-Stage CF *P. aeruginosa* Isolates

Other mutations tend to appear in late-stage *P. aeruginosa* CF infections. The multi-drug efflux-associated *mexZ* gene was found to be a common target for mutagenesis [118,170]. MexZ is the negative regulator of MexX and MexY, which together with OprM porin form the MexXY-OprM multi-drug efflux porin [171]. The mutation of *mexZ* de-inhibits both *mexX* and *mexY* genes, leading to their increased expressions and antibiotic resistance [172]. Other patients had also been found to possess isolates with increased expression of MexX and MexY instead [173]. As CF patients are frequently treated with antibiotics, such as tobramycin, the selection pressure for antibiotic-resistant genes can be considered to be far greater than other genes [173].

Another gene with high mutation frequency from late-stage CF *P. aeruginosa* isolates is *mucA*, with up to 60% of 38 isolates from 26 CF patients possessing *mucA* mutations [170]. Genetic sequencing of the *mucA* gene in most clinical isolates found that the loss of function is a result of premature termination of the mucA coding sequence [39]. Hence, *mucA* is a highly preferential gene for conversion to mucoidy in the CF lung, providing a protective barrier against immune clearance in the lungs. Other than alginate production, additional downstream genes had been identified to possess promoter sites for potential AlgU binding [174]. For instance, lipoproteins (IptA and IptB) which can activate inflammatory pathways in macrophages were identified to be dependent on AlgU binding [174].

It had been proposed that the mutations stated above had already started to occur in the early colonization of airways, especially at the paranasal sinuses in young CF patients [175]. The paranasal sinuses were observed to serve as important reservoirs by *P. aeruginosa* in the early infections, thus subjecting the lungs to intermittent seeding by *P. aeruginosa* and leading to the eventual chronic lung infection in the future.

17.15 Transcriptomics of *P. aeruginosa* in Chronic CF Infections

Other than comparative genomics, transcriptomic analyses are widely used to compare the physiology and adaptation of multiple CF *P. aeruginosa* isolates. In our previous work, mRNAs extracted from logarithmic phase cultures of CF *P. aeruginosa* isolates from the DK2 clonal lineages were used for microarray-based transcriptomic analysis [165]. The *P. aeruginosa* isolates used in this study includes early- and late-stage CF isolates from the same patients as well as isolates from newborn CF children (Figure 17.2). Even though these CF isolates had different adaptation backgrounds, unsupervised bioinformatics analysis (hierarchical clustering) was able to clearly classify the transcriptomes into three groups (early-stage, late-stage and mucoid isolates) (Figure 17.3). In accordance with the comparative genomics, independent component analysis was also able to reveal the importance

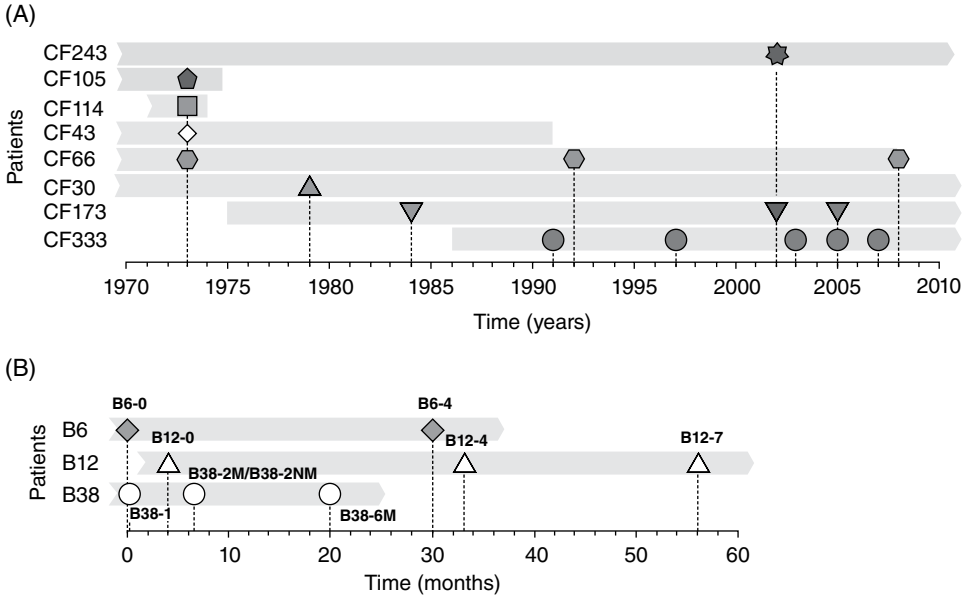


Figure 17.2 Sampling points of isolates and life expectancy of patients. *P. aeruginosa* isolates were isolated from 11 CF patients over 35 years. Each isolate is labelled by different symbols, whereas patient life expectancy is shown with grey bars. Figure adapted from Yang, L. *et al.*, 2011. Evolutionary dynamics of bacteria in a human host environment. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7481–7486 (2011); published online EpubMay 3 (10.1073/pnas.1018249108) [119].

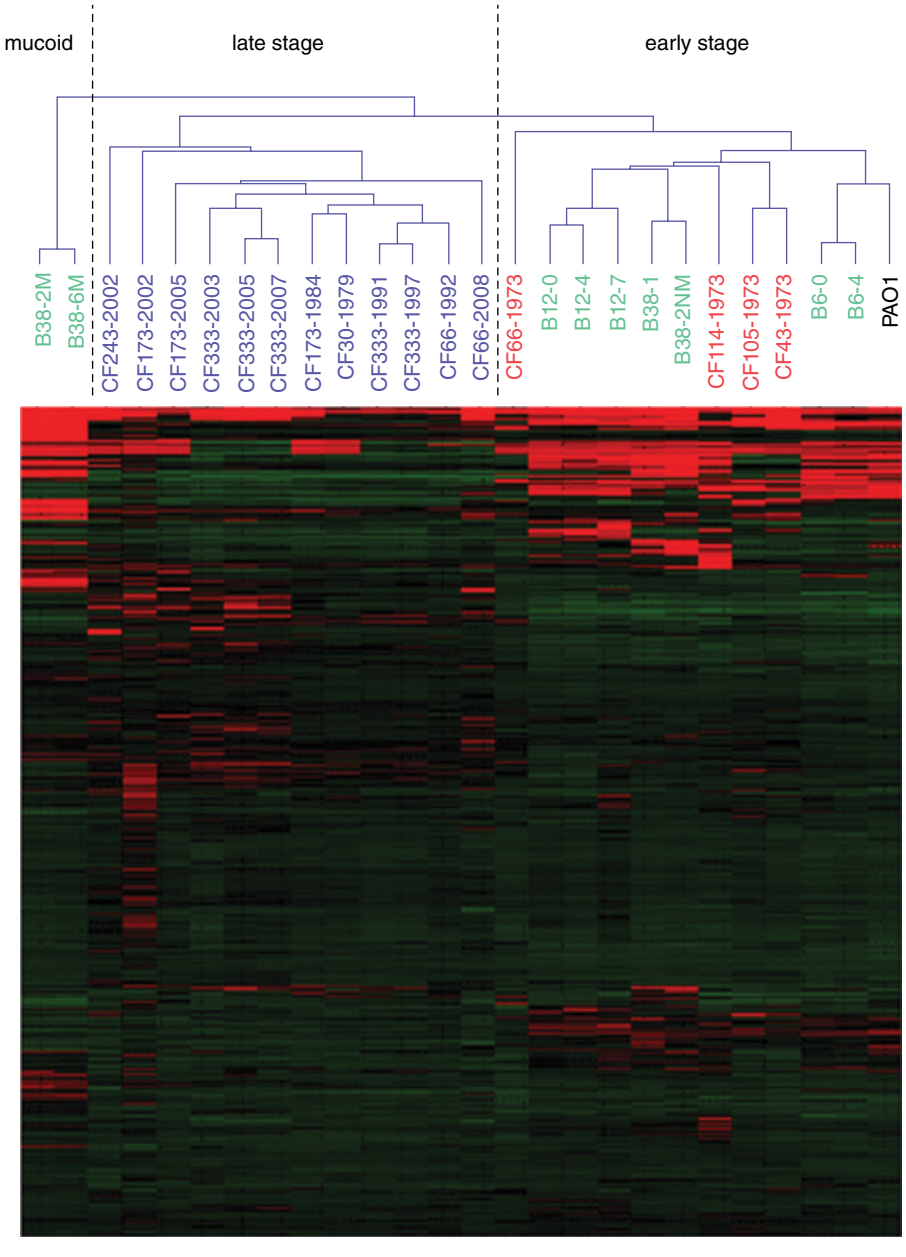


Figure 17.3 Hierarchical clustering of the RNA transcripts using Euclidean distances. The red block represents an increase in signal, whereas the green block shows a reduction in signal. Figure adapted from Yang, L. *et al.*, 2011. Bacterial adaptation during chronic infection revealed by independent component analysis of transcriptomic data. *BMC Microbiology* **11**, 184 (2011)10.1186/1471-2180-11-184) [165].

of losing QS and gain mucoidy and antibiotic resistance in the adaptation of *P. aeruginosa* in the CF lung [165].

RNA sequencing was also employed in the study of important genes in the adaptation of *P. aeruginosa* in the CF lung. Clonal *P. aeruginosa* isolates were sequentially collected from different patients ranging from months to years, and their transcriptomes were compared by using RNA sequencing to study the adaptive traits [176]. Twenty-four common genes, such as the type IV pilus biosynthesis genes, were identified as showing similar changes among three clonal lineages, suggesting they play an important role in CF *P. aeruginosa* adaptation and highlighting the parallel evolution trait of *P. aeruginosa* in the CF lung [176].

Transcriptomic analysis of sequential isolates from patients over three to five years showed that, in addition to the QS, pili and antibiotic-resistant genes, metabolic pathways important in the *P. aeruginosa* adaptation in the lungs were also affected. Increased expression of genes important in anaerobic (e.g. *anr*) and microaerobic respiration (e.g. cytochrome oxidase *cbb3*) was found in the late CF *P. aeruginosa* isolates compared to early isolates from a hypermutable linage [177]. This observation showed that *P. aeruginosa* persisted in the hypoxic or anaerobic conditions of mucus in the lungs [178–180].

17.16 Proteomics of *P. aeruginosa* in Chronic CF Infections

Although genomics and transcriptomics had helped in answering how clinical isolates from CF patients evolve over time, comparative proteomics of CF *P. aeruginosa* isolates and defined laboratory strains can reveal the functionality of the metabolic pathways as well as biomarkers of CF infectious stages.

17.17 Applications of Proteomics to *P. aeruginosa* Characterization

Because transcriptomes provide changes at the gene level only, which may not lead to changes in metabolic pathways and phenotypes, proteomics is being used to provide phenotype evidence and biomarkers for pathogenesis of *P. aeruginosa* CF lung infections. It is essential to have physiologically relevant in vitro and in vivo models to perform proteomics so that the proteomic analysis can reflect true disease-causing mechanisms of *P. aeruginosa*. Initial studies mainly used 2-DE combined with mass spectrometry to identify proteins and compare proteins spots between samples [181]. With the advent of quantitative proteomics by isobaric tag for relative and absolute quantification (iTRAQ), proteome data of higher sensitivity and accuracy had been obtained to study the relative proportions of certain peptides between differentially labelled samples [182].

Comparative proteomics are used in but not limited to the following models for characterizing *P. aeruginosa* infections: (1) comparing non-mucoid and mucoid CF *P. aeruginosa* isolates to understand the physiological impact of mucoidy [183]; (2) comparing laboratory strains and their mutants with CF-like phenotypic traits to identify regulation of the CF phenotypes [184]; (3) comparing drug-treated and -untreated *P. aeruginosa* isolates to study their resistant mechanisms [185]; revealing

the stress response of *P. aeruginosa* isolates towards ROS to identify the factors required for survival during host immune response [186]; (5) comparing *P. aeruginosa* isolates cultivated under CF-like physiological conditions to laboratory conditions to reveal their metabolic adaptation during infections [187,188]; (6) comparing planktonic and biofilm cells to identify factors contributing to biofilm formation and factors required specifically for acute and chronic infections [189]; and (7) comparing subcellular proteomes to identify membrane-associated and secreted virulence factors [190,191].

Here we give examples of the application of comparative proteomics to reveal the physiology of *P. aeruginosa* strains with CF-like phenotypes.

17.18 Comparative Proteomic Investigation of Bis-(3′-5′)-Cyclic-Dimeric-GMP (C-Di-GMP) Regulation in *P. aeruginosa*

C-di-GMP is a global secondary messenger employed by many species of bacteria for regulating biofilm formation and dispersal [192]. The diguanylatecyclases (DGCs) catalyze c-di-GMP formation, whereas phosphodiesterases (PDEs) cause the degradation of c-di-GMP; thus, bacteria can adjust the intracellular concentration of c-di-GMP according to the changes in environment [193]. Increased concentration of c-di-GMP will promote EPS production and reduce motility for enhancing biofilm formation [194,195]. Reducing c-di-GMP content will cause biofilm dispersal and induce virulence factors for acute infections [105,196,197].

The rugose small-colony variants (RSCV) are often observed in late-stage CF *P. aeruginosa* isolates, and they possess high intracellular c-di-GMP content [198]. RSCV isolates produce large amounts of biofilms with the help of Pel and Psl exopolysaccharides compared to other CF isolates. These phenotypes are caused by enhanced intracellular c-di-GMP content via the *wsp* operon. Mutation in the *wspF* regulator gene can cause the constitutive activation of WspR, a DGC, resulting in high intracellular accumulation of c-di-GMP and formation of RSCV [161,194]. We have compared the proteomes of a *wspF* mutant (with high intracellular c-di-GMP content) and a wild-type strain containing pYhjH, which encodes a constitutively expressing PDE and thus degrades intracellular c-di-GMP [184]. Very interestingly, our data showed that high intracellular c-di-GMP not only induces synthesis of Pel and Psl exopolysaccharides, but also enhances the production of iron siderophore pyoverdine [184], which is an important factor for virulence and biofilm formation in *P. aeruginosa* [199]. Comparative proteomics of RSCV isolates with wild-type *P. aeruginosa* strains by other groups showed that drug efflux pumps such as MexAB, MexXY and OprM had been upregulated by RSCV strains, suggesting that c-di-GMP plays a role in antibiotic resistance [200].

Comparing the proteomes of a *wspF* mutant and the pYhjH-containing strain also showed that c-di-GMP positively regulates alginate synthesis [184]. This data is in accordance with a previous finding showing that *mucR*, an important regulator in the alginate synthesis, acts as a membrane-localized DGC [201].The Δ*mucR* mutant is deficient in alginate production, while its over-expression enhanced alginate synthesis and secretion. Proteomic analysis also revealed that MucR altered the outer membrane profile, especially by increasing the expression of AlgE. AlgE is an outer membrane protein, important in the export of alginate out of bacterial cells [202]. C-di-GMP produced by MucR was shown to bind specifically to its receptor Alg44 via its PilZ domain [203].

17.19 Comparative Proteomics of Mucoid and Non-Mucoid *P. aeruginosa* Strains

As mucoidy is the most important and most studied phenotype in CF *P. aeruginosa* isolates [183,190,204–208], it will be highly desirable to study the proteomes of mucoid strains and discover proteins important for mucoidy. Comparative proteomic analysis of mucoid and non-mucoid CF isolates showed that AmrZ (alginate motility regulator Z) and alginate biosynthesis proteins such as AlgC were upregulated in mucoid strains [183]. AmrZ was also found to be upregulated transcriptionally in the same mucoid strain and was a DNA-binding protein that enhances AlgD expression [209–211]. AlgC was a phosphomannomutase that is important in alginate and LPS synthesis [212,213]. A periplasmic porin, OprF, was also found to be induced in mucoid strains in many proteomic studies [191,204,206], implying a possible function in secretion of small molecules relating to alginate biosynthesis.

Proteins downregulated in mucoid strains compared to non-mucoid strains include many motility related proteins, resulting in absence of flagellar motility [183,214]. AlgU (AlgT, σ^{22}) was able to suppress expression of flagella via FleQ inhibition [214]. This caused the initially motile strains to convert to non-motile ones as infections progressed from acute to chronic. Loss of motility was essential to evade the host immune response as flagella can elicit an immunogenic response [215].

Interestingly, expression of type VI secretion system (T6SS) was also observed to be reduced in mucoid strains by comparative proteomics [183], which was corroborated with transcriptomic data that T6SS genes such as *hcp1*, *tssB1* and *tssC1* were downregulated in mucoid CF isolates [145,214]. It was hypothesized that T6SS was important in early clinical non-mucoid isolates for competing against other bacteria, but it was not necessary later when chronic infection by mucoidy had been established [72,183]. The above studies about mucoidy show that proteomics and transcriptomics could complement and correlate with each other very well.

17.20 Proteogenomics Reveal Shifting in Iron Uptake of CF *P. aeruginosa*

Another good example of combining genomics, transcriptomics and proteomics for investigation of CF *P. aeruginosa* adaptation is the examination of its iron uptake pathways. Iron is essential for growth of bacterial pathogens. Bacteria have evolved multiple specific iron scavenge systems to obtain iron from nature and host infection sites. Most iron in the human body is locked in hemoglobin and lactoferritin produced by the immune system that can sequester iron away from pathogens [216,217]. These factors might create a selection pressure on pathogens during chronic infections, allowing them to adapt their iron uptake systems.

P. aeruginosa is able to utilize two types of well-established systems, siderophore based and heme acquisition based, to scavenge iron from the environment (Figure17.4). The siderophores, namely pyoverdine [218] and pyochelin [219], are low-molecular-weight compounds that can chelate iron at very high affinities from the milieu and return to bacterial cells after binding to specific surface receptors for iron absorption.

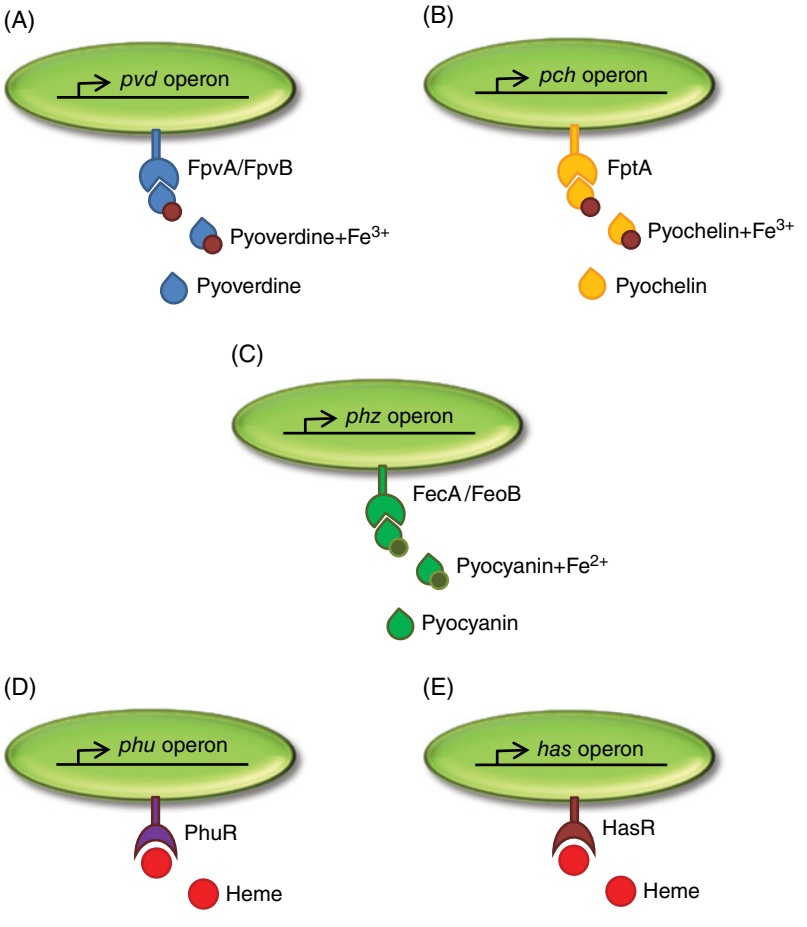


Figure 17.4 Iron acquisition systems employed by *P. aeruginosa*. Siderophore-based systems including *pvd* [220,221] (A) and *pch* [222] (B) are both involved in Fe^{3+} uptake; *phz* (C) is involved in Fe^{2+} uptake [223]; heme-acquisition-based systems including *phu* (D) and *has* (E), are involved in uptake of hemeproteins.

Similarly, heme-sequestering carriers are usually proteins that chelate and transport heme across bacterial membrane after binding with specific bacterial surface receptors (Figure 17.4).

Pyoverdine and pyochelin are required in the pathogenesis of various animal infection models as they scavenge Fe^{3+} from host cells during infections [218,225,226]. When pyoverdine binds to the FpvA membrane receptor, it activates downstream virulence genes via PvdS, allowing the production of ExoA [227] and an endoprotease PrpL [228]. While pyoverdine is important in acute infections for iron uptake and regulation of virulence factors, genomic and transcriptomic analyses suggested that the siderophore-mediated iron uptake seemed to play a less major role in CF infections [165]. Several studies showed that siderophore-mediated iron uptake genes were still expressed in most clinical samples [229,230], but with samples having low levels [229] or gaining mutations in pyoverdine genes over time [231,232], this implies that the late-stage CF

P. aeruginosa isolates either have high concentrations of free iron or utilize alternative iron uptake systems.

A recent work used comparative genomics and transcriptomics to show that loss of siderophore production coincides with increased expression of the *phu* heme uptake system [233]. Genomic analysis of CF *P. aeruginosa* isolates revealed a 38-fold higher frequency of mutations in the promoters of the PhuR receptor than would occur by chance, resulting in high expression level of *phuR* in late-stage CF isolates [233]. Furthermore, CF isolates from different patients but from the same lineage showed convergent evolution towards upregulation of *phuR*, implying that the heme system is subjected to evolutionary adaptation [233]. Our recent studies showed that the PhuR protein is one of the highest expressed proteins in late-stage CF isolates, and its expression level is very low in the early-stage ancestor strain of this isolate from the same CF patient (data not shown). Hence, it is very likely that heme-sequestering systems are employed by *P. aeruginosa* isolates during the late-stage infections in CF patients [231,233]. This might confer the late-stage isolates advantages over early-stage isolates when growing in the presence hemoglobin, whose bioavailability in the CF lung is high due to micro-bleeding [234] and abundance in alveolar epithelial cells [235].

Loss-of-function mutations of siderophore synthesis and receptor genes were also found to lead to induction of an alginate synthesis gene *algD* [169]. Moreover, growing mucoid strains in simple iron sources such as iron (III) citrate and iron (III) chloride abrogated alginate production, whereas heme sources allowed isolates to retain mucoidy, suggesting that the use of heme as an iron source has less repressive effects on alginate production than the use of other iron sources [169]. Hence, shifts in iron uptake pathways can influence alginate production and might allow better survival of *P. aeruginosa* in the CF lung in response to immune clearance. Targeting the heme uptake pathways in late-stage CF infections might be an alternative approach to control *P. aeruginosa* infections.

Although most studies investigate the uptake of ferric (Fe^{3+}), few studies showed the role of ferrous iron (Fe^{2+}) in CF *P. aeruginosa* isolates. As infections progressed, Fe^{2+} was also recently found to be abundant in the mucus of CF lungs [236]. Interestingly, *P. aeruginosa* could also uptake Fe^{2+} using phenazines via the FeoB receptor [229,237].

Because mucoidy was shown to link with differential iron utilization systems, the underline mechanisms between mucoidy and iron uptake remain unclear. It is possible that heme utilization accumulates intracellular ROS after being uptaken by *P. aeruginosa* [238]. This can select for mucoidy, as mucoidy was shown to protect *P. aeruginosa* from oxidative damage [239]. Further studies are required to be conducted to prove the links for such complex relationships.

17.21 Conclusion and Future Perspectives

The application of genomic, transcriptomic and proteomic technologies has vastly improved our understanding of the adaptation and evolution of *P. aeruginosa* at the single-nucleotide, gene and protein expression levels, respectively. This information helped provide important targets for future therapeutic interventions. These techniques can identify important biomarkers in the transition from early- to late-stage CF infections. This knowledge allows us to target relevant genes and proteins at the early

stage, when there is a better chance to eradicate *P. aeruginosa* infections before the disease progresses to the chronic phase.

Nonetheless, there is still room for improvement in the depth, resolution and sensitivity of genomic sequencing platforms and mass-spectrometry-based proteomic platforms. Future sequencing platforms (such as PacBio RS II) will provide high-resolution sequences with fewer gaps and other nucleotide modification (e.g. DNA methylation) information than the current sequencing platform, and allow more accurate genome organization and SNP analyses than what is available today. More sensitive mass spectrometers will enable us to detect low-abundant proteins from proteomic samples.

Despite the large number of studies carried for CF *P. aeruginosa* infections, current research still presents a simplistic view of the infection itself. Most infections are polymicrobial, with the CF airway containing a complex microbiome [240,241], which consists of more than 20 genera such as *Streptococcus*, *Staphylococcus* and *Achromobacter*. It was further observed that patients with disease progression possessed higher microbiome diversity as compared to patients with stable chronic disease [240]. Predictably, the declining diversity of bacterial communities was attributed to antibiotics treatment [240]. The interactions of those species are highly complex and diverse; thus, it is important not to neglect the other species and focus on *P. aeruginosa* alone. Investigation of interspecies interactions will further elucidate the mechanisms of *P. aeruginosa* CF adaptation [242].

Most proteogenomic studies only applied to a single genome, and mass spectrometry data was limited to a single proteome. The complexity of the CF microbiome has raised the need to study microbial communities holistically by using meta-proteogenomics. With the increased quantity and quality of metagenomic data, and the increased accuracy and sensitivity of mass spectrometers, it is possible to quantify and identify proteins expressed in the microbiome. To study protein abundance and function, metaproteomic data from MS will be searched against the metagenome database. As metagenomics cannot decide which species or gene is important in carrying out a certain function, protein identification and quantification can show its active role in real time. Such techniques have been applied in studying other diseases, for example, in the case of gut microbiota in Crohn's disease [243]. Hence, meta-proteogenomics can lead to improved annotations of genomes for gene function and classification [244].

To further complicate the study of CF infections, the human airway consists of different anatomical areas, including the nasal cavities, bronchi and lungs. Such a differentiated spatial organization will favour different specialized microbiota to reside in [245]. This is partly due to the varied environments along the airways, favouring differential growth of bacteria [179,246]. Most sampling methods had been restricted to getting sputum expectorated from patients [247] and fluid from bronchoalveolar lavage (BAL) [248]. Hence, it is important to take into account that the CF infections are polymicrobial and multi-factorial [249,250], requiring customized treatments for different areas of infections. If not, antibiotic treatment of certain microbial species may be effective only in certain regions of the airways.

Other than collecting isolates from human samples, both genomics and proteomics should also be used in defined chronic animal infection models to study the adaptations of *P. aeruginosa* to CF-like conditions. As there were large variations across patient demographics, treatment and disease progression, laboratory-based experiments should be employed to confirm results from the clinical samples. This will confer

control and reproducibility to the study of *P. aeruginosa* evolution, and provide a means of testing new drugs or varied conditions. Animal models including the pig [68,251] and sheep [252] will also provide suitable replacements in studying *P. aeruginosa*–mediated lung infections, given their resemblance in airway anatomy and genetics to humans.

References

1 C. K. Stover, X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, M. V. Olson, Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* **406**, 959–964 (2000); published online EpubAug 31 (10.1038/35023079).

2 G. P. Bodey, R. Bolivar, V. Fainstein, L. Jadeja, Infections caused by Pseudomonas aeruginosa. *Reviews of Infectious Diseases* **5**, 279–313 (1983); published online EpubMar-Apr (10.1093/clinids/5.2.279).

3 T. Bjarnsholt, P. O. Jensen, M. J. Fiandaca, J. Pedersen, C. R. Hansen, C. B. Andersen, T. Pressler, M. Givskov, N. Hoiby, Pseudomonas aeruginosa biofilms in the respiratory tract of cystic fibrosis patients. *Pediatric Pulmonology* **44**, 547–558 (2009); published online EpubJun (10.1002/ppul.21011).

4 M. Fazli, T. Bjarnsholt, K. Kirketerp-Moller, A. Jorgensen, C. B. Andersen, M. Givskov, T. Tolker-Nielsen, Quantitative analysis of the cellular inflammatory response against biofilm bacteria in chronic wounds. *Wound Repair and Regeneration: Official Publication of the Wound Healing Society [and] the European Tissue Repair Society* **19**, 387–391 (2011); published online EpubMay–Jun (10.1111/j.1524-475X.2011.00681.x).

5 K. H. Turner, J. Everett, U. Trivedi, K. P. Rumbaugh, M. Whiteley, Requirements for Pseudomonas aeruginosa acute burn and chronic surgical wound infection. *PLoS Genetics* **10**, e1004518 (2014); published online EpubJul (10.1371/journal.pgen.1004518).

6 L. D. Hazlett, Corneal response to Pseudomonas aeruginosa infection. *Progress in Retinal and Eye Research* **23**, 1–30 (2004); published online EpubJan (10.1016/j.preteyeres.2003.10.002).

7 G. P. Bodey, L. Jadeja, L. Elting, Pseudomonas bacteremia. Retrospective analysis of 410 episodes. *Archives of Internal Medicine* **145**, 1621–1629 (1985); published online EpubSep (10.1001/archinte.1985.00360090089015).

8 C. R. Huang, C. H. Lu, Y. C. Chuang, N. W. Tsai, C. C. Chang, S. F. Chen, H. C. Wang, C. C. Chien, W. N. Chang, Adult Pseudomonas aeruginosa meningitis: High incidence of underlying medical and/or postneurosurgical conditions and high mortality rate. *Japanese Journal of Infectious Diseases* **60**, 397–399 (2007); published online EpubNov 9.

9 C. Van Delden, B. H. Iglewski, Cell-to-cell signaling and Pseudomonas aeruginosa infections. *Emerging Infectious Diseases* **4**, 551–560 (1998); published online Epub Oct–Dec (10.3201/eid0404.980405).

10 B. Frederiksen, C. Koch, N. Hoiby, Antibiotic treatment of initial colonization with Pseudomonas aeruginosa postpones chronic infection and prevents deterioration of pulmonary function in cystic fibrosis. *Pediatric Pulmonology* **23**, 330–335 (1997);

published online EpubMay (10.1002/(SICI)1099-0496(199705)23:5<330::AID-PPUL4>3.0.CO;2-O).

11 T. S. Murray, M. Egan, B. I. Kazmierczak, Pseudomonas aeruginosa chronic colonization in cystic fibrosis patients. *Current Opinion in Pediatrics* **19**, 83–88 (2007); published online EpubFeb (10.1097/MOP.0b013e3280123a5d).

12 G. H. Talbot, J. Bradley, J. E. Edwards, Jr., D. Gilbert, M. Scheld, J. G. Bartlett, A. Antimicrobial Availability Task Force of the Infectious Diseases Society of, Bad bugs need drugs: An update on the development pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* **42**, 657–668 (2006); published online EpubMar 1 (10.1086/499819).

13 L. C. Tsui, M. Buchwald, D. Barker, J. C. Braman, R. Knowlton, J. W. Schumm, H. Eiberg, J. Mohr, D. Kennedy, N. Plavsic *et al.*, Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* **230**, 1054–1057 (1985); published online EpubNov 29 (10.1126/science.2997931).

14 F. Ratjen, G. Doring, Cystic fibrosis. *Lancet* **361**, 681–689 (2003); published online EpubFeb 22 (10.1016/S0140-6736(03)12567-6).

15 P. A. Flume, P. J. Mogayzel, Jr., K. A. Robinson, R. L. Rosenblatt, L. Quittell, B. C. Marshall, C. Clinical practice guidelines for pulmonary therapies, C. Cystic Fibrosis Foundation Pulmonary Therapies, Cystic fibrosis pulmonary guidelines: Pulmonary complications: Hemoptysis and pneumothorax. *American Journal of Respiratory and Critical Care Medicine* **182**, 298–306 (2010); published online EpubAug 1 (10.1164/rccm.201002-0157CI10.1164/rccm.201002-0157OC).

16 P. M. Quinton, Chloride impermeability in cystic fibrosis. *Nature* **301**, 421–422 (1983); published online EpubFeb 3 (10.1038/301421a0).

17 S. M. Rowe, S. Miller, E. J. Sorscher, Cystic fibrosis. *The New England Journal of Medicine* **352**, 1992–2001 (2005); published online EpubMay 12 (10.1056/NEJMra043184).

18 T. MacKenzie, A. H. Gifford, K. A. Sabadosa, H. B. Quinton, E. A. Knapp, C. H. Goss, B. C. Marshall, Longevity of patients with cystic fibrosis in 2000 to 2010 and beyond: survival analysis of the cystic fibrosis foundation patient registry. *Annals of Internal Medicine* **161**, 233–241 (2014); published online EpubAug 19 (10.7326/M13-0636).

19 F. Harrison, Microbial ecology of the cystic fibrosis lung. *Microbiology* **153**, 917–923 (2007); published online EpubApr (10.1099/mic.0.2006/004077-0).

20 L. Saiman, Microbiology of early CF lung disease. *Paediatric Respiratory Reviews* **5 Suppl A**, S367–369 (2004).

21 M. F. Tosi, H. Zakem-Cloud, C. A. Demko, J. R. Schreiber, R. C. Stern, M. W. Konstan, M. Berger, Cross-sectional and longitudinal studies of naturally occurring antibodies to Pseudomonas aeruginosa in cystic fibrosis indicate absence of antibody-mediated protection and decline in opsonic quality after infection. *The Journal of Infectious Diseases* **172**, 453–461 (1995); published online EpubAug (10.1093/infdis/172.2.453).

22 N. Hoiby, O. Ciofu, T. Bjarnsholt, Pseudomonas aeruginosa biofilms in cystic fibrosis. *Future Microbiol* **5**, 1663–1674 (2010); published online EpubNov (10.2217/fmb.10.125).

23 C. F. Foundation, "Patient Registry, Annual Data Report 2009," (2009).

24 P. K. Singh, A. L. Schaefer, M. R. Parsek, T. O. Moninger, M. J. Welsh, E. P. Greenberg, Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **407**, 762–764 (2000); published online EpubOct 12 (10.1038/35037627).

25 D. M. Ramsey, D. J. Wozniak, Understanding the control of Pseudomonas aeruginosa alginate synthesis and the prospects for management of chronic infections in cystic fibrosis. *Molecular Microbiology* **56**, 309–322 (2005); published online EpubApr (10.1111/j.1365-2958.2005.04552.x).

26 M. Kolpen, C. R. Hansen, T. Bjarnsholt, C. Moser, L. D. Christensen, M. van Gennip, O. Ciofu, L. Mandsberg, A. Kharazmi, G. Doring, M. Givskov, N. Hoiby, P. O. Jensen, Polymorphonuclear leucocytes consume oxygen in sputum from chronic Pseudomonas aeruginosa pneumonia in cystic fibrosis. *Thorax* **65**, 57–62 (2010); published online EpubJan (10.1136/thx.2009.114512).

27 J. Hull, P. Vervaart, K. Grimwood, P. Phelan, Pulmonary oxidative stress response in young children with cystic fibrosis. *Thorax* **52**, 557–560 (1997); published online EpubJun 28.

28 J. A. Simpson, S. E. Smith, R. T. Dean, Scavenging by alginate of free radicals released by macrophages. *Free Radical Biology & Medicine* **6**, 347–353 (1989).

29 J. G. Leid, C. J. Willson, M. E. Shirtliff, D. J. Hassett, M. R. Parsek, A. K. Jeffers, The exopolysaccharide alginate protects Pseudomonas aeruginosa biofilm bacteria from IFN-gamma-mediated macrophage killing. *Journal of Immunology* **175**, 7512–7518 (2005); published online EpubDec 1 (10.4049/jimmunol.175.11.7512).

30 S. S. Pedersen, A. Kharazmi, F. Espersen, N. Hoiby, Pseudomonas aeruginosa alginate in cystic fibrosis sputum and the inflammatory response. *Infection and Immunity* **58**, 3363–3368 (1990); published online EpubOct 31.

31 M. Alhede, T. Bjarnsholt, P. O. Jensen, R. K. Phipps, C. Moser, L. Christophersen, L. D. Christensen, M. van Gennip, M. Parsek, N. Hoiby, T. B. Rasmussen, M. Givskov, Pseudomonas aeruginosa recognizes and responds aggressively to the presence of polymorphonuclear leukocytes. *Microbiology* **155**, 3500–3508 (2009); published online EpubNov (10.1099/mic.0.031443-0mic.0.031443-0 [pii]).

32 H. K. Johansen, L. Norregaard, P. C. Gotzsche, T. Pressler, C. Koch, N. Hoiby, Antibody response to Pseudomonas aeruginosa in cystic fibrosis patients: a marker of therapeutic success? – A 30-year cohort study of survival in Danish CF patients after onset of chronic P. aeruginosa lung infection. *Pediatric Pulmonology* **37**, 427–432 (2004); published online EpubMay (10.1002/ppul.10457).

33 S. J. Pamp, M. Gjermansen, H. K. Johansen, T. Tolker-Nielsen, Tolerance to the antimicrobial peptide colistin in Pseudomonas aeruginosa biofilms is linked to metabolically active cells, and depends on the pmr and mexAB-oprM genes. *Molecular Microbiology* **68**, 223–240 (2008); published online EpubApr (10.1111/j.1365-2958.2008.06152.x).

34 D. L. MacLeod, L. E. Nelson, R. M. Shawar, B. B. Lin, L. G. Lockwood, J. E. Dirk, G. H. Miller, J. L. Burns, R. L. Garber, Aminoglycoside-resistance mechanisms for cystic fibrosis Pseudomonas aeruginosa isolates are unchanged by long-term, intermittent, inhaled tobramycin treatment. *The Journal of Infectious Diseases* **181**, 1180–1184 (2000); published online EpubMar (10.1086/315312).

35 F. Bert, C. Branger, N. Lambert-Zechovsky, Identification of PSE and OXA beta-lactamase genes in Pseudomonas aeruginosa using PCR-restriction fragment length polymorphism. *The Journal of Antimicrobial Chemotherapy* **50**, 11–18 (2002); published online EpubJul (10.1093/jac/dkf069).

36 M. A. Alkawash, J. S. Soothill, N. L. Schiller, Alginate lyase enhances antibiotic killing of mucoid Pseudomonas aeruginosa in biofilms. *APMIS: Acta Pathologica, Microbiologica,*

et Immunologica Scandinavica **114**, 131–138 (2006); published online EpubFeb (10.1111/j.1600-0463.2006.apm_356.x).

37 H. H. Diaz E, Lai A, and Yadana J, Role of alginate in gentamicin antibiotic susceptibility during the early stages of Pseudomonas aeruginosa PAO1 Biofilm Establishment *JEMI* **15**, 71–78 (2011).

38 K. Lewis, Persister cells. *Annual Review of Microbiology* **64**, 357–372 (2010)10.1146/annurev.micro.112408.134306).

39 J. C. Boucher, H. Yu, M. H. Mudd, V. Deretic, Mucoid Pseudomonas aeruginosa in cystic fibrosis: Characterization of muc mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection. *Infection and immunity* **65**, 3838–3846 (1997); published online EpubSep 40.

40 L. F. Wood, D. E. Ohman, Identification of genes in the sigma(2)(2) regulon of Pseudomonas aeruginosa required for cell envelope homeostasis in either the planktonic or the sessile mode of growth. *mBio* **3**, (2012)10.1128/mBio.00094-12).

41 L. F. Wood, D. E. Ohman, Use of cell wall stress to characterize sigma 22 (AlgT/U) activation by regulated proteolysis and its regulon in Pseudomonas aeruginosa. *Molecular Microbiology* **72**, 183–201 (2009); published online EpubApr (10.1111/j.1365-2958.2009.06635.x).

42 G. J. Fyfe J, A revised chromosomal location for muc: A locus involved in the control of alginate production by mucoid Pseudomonas aeruginosa. *Society for General Microbiology Quarterly* **8**, 250–251 (1981).

43 A. Brooun, S. Liu, K. Lewis, A dose-response study of antibiotic resistance in Pseudomonas aeruginosa biofilms. *Antimicrobial Agents and Chemotherapy* **44**, 640–646 (2000); published online EpubMar (10.1128/AAC.44.3.640-646.2000).

44 J. W. Costerton, P. S. Stewart, E. P. Greenberg, Bacterial biofilms: A common cause of persistent infections. *Science* **284**, 1318–1322 (1999); published online EpubMay 21 (10.1126/science.284.5418.1318).

45 A. L. Spoering, K. Lewis, Biofilms and planktonic cells of Pseudomonas aeruginosa have similar resistance to killing by antimicrobials. *Journal of Bacteriology* **183**, 6746–6751 (2001); published online EpubDec (10.1128/JB.183.23.6746-6751.2001).

46 E. Drenkard, F. M. Ausubel, Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature* **416**, 740–743 (2002); published online EpubApr 18 (10.1038/416740a).

47 T. Strateva, I. Mitov, Contribution of an arsenal of virulence factors to pathogenesis of Pseudomonas aeruginosa infections. *Annals of Microbiology* **61**, 717–732 (2011).

48 M. B. Miller, B. L. Bassler, Quorum sensing in bacteria. *Annual Review of Microbiology* **55**, 165–199 (2001)10.1146/annurev.micro.55.1.165).

49 N. A. Whitehead, A. M. Barnard, H. Slater, N. J. Simpson, G. P. Salmond, Quorum-sensing in Gram-negative bacteria. *FEMS Microbiology Reviews* **25**, 365–404 (2001); published online EpubAug (10.1111/j.1574-6976.2001.tb00583.x).

50 J. P. Pearson, E. C. Pesci, B. H. Iglewski, Roles of Pseudomonas aeruginosa las and rhl quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes. *Journal of Bacteriology* **179**, 5756–5767 (1997); published online EpubSep (10.1128/jb.179.18.5756-5767.1997).

51 S. P. Diggle, S. Matthijs, V. J. Wright, M. P. Fletcher, S. R. Chhabra, I. L. Lamont, X. Kong, R. C. Hider, P. Cornelis, M. Camara, P. Williams, The Pseudomonas aeruginosa 4-quinolone signal molecules HHQ and PQS play multifunctional roles in

quorum sensing and iron entrapment. *Chemistry & Biology* **14**, 87–96 (2007); published online EpubJan (10.1016/j.chembiol.2006.11.014).

52 A. Latifi, M. Foglino, K. Tanaka, P. Williams, A. Lazdunski, A hierarchical quorum-sensing cascade in Pseudomonas aeruginosa links the transcriptional activators LasR and RhIR (VsmR) to expression of the stationary-phase sigma factor RpoS. *Molecular Microbiology***21**, 1137–1146 (1996); published online EpubSep.

53 D. S. Wade, M. W. Calfee, E. R. Rocha, E. A. Ling, E. Engstrom, J. P. Coleman, E. C. Pesci, Regulation of Pseudomonas quinolone signal synthesis in Pseudomonas aeruginosa. *Journal of Bacteriology* **187**, 4372–4380 (2005); published online EpubJul (10.1128/JB.187.13.4372-4380.2005).

54 S. L. McKnight, B. H. Iglewski, E. C. Pesci, The Pseudomonas quinolone signal regulates rhl quorum sensing in Pseudomonas aeruginosa. *Journal of Bacteriology* **182**, 2702–2708 (2000); published online EpubMay (10.1128/JB.182.10.2702-2708.2000).

55 Z. Kuang, Y. Hao, B. E. Walling, J. L. Jeffries, D. E. Ohman, G. W. Lau, Pseudomonas aeruginosa elastase provides an escape from phagocytosis by degrading the pulmonary surfactant protein-A. *PLoS ONE* **6**, e27091 (2011)10.1371/journal.pone.0027091).

56 B. Wretlind, O. R. Pavlovskis, Pseudomonas aeruginosa elastase and its role in pseudomonas infections. *Reviews of Infectious Diseases* **5 Suppl 5**, S998–1004 (1983); published online EpubNov–Dec.

57 N. Moker, C. R. Dean, J. Tao, Pseudomonas aeruginosa increases formation of multidrug-tolerant persister cells in response to quorum-sensing signaling molecules. *Journal of Bacteriology* **192**, 1946–1955 (2010); published online EpubApr (10.1128/JB.01231-09).

58 N. Bhargava, P. Sharma, N. Capalash, Pyocyanin stimulates quorum sensing-mediated tolerance to oxidative stress and increases persister cell populations in Acinetobacter baumannii. *Infection and Immunity* **82**, 3417–3425 (2014); published online EpubAug (10.1128/IAI.01600-14).

59 B. E. Britigan, T. L. Roeder, G. T. Rasmussen, D. M. Shasby, M. L. McCormick, C. D. Cox, Interaction of the Pseudomonas aeruginosa secretory products pyocyanin and pyochelin generates hydroxyl radical and causes synergistic damage to endothelial cells. Implications for Pseudomonas-associated tissue injury. *The Journal of Clinical Investigation* **90**, 2187–2196 (1992); published online EpubDec (10.1172/JCI116104).

60 R. Wilson, D. A. Sykes, D. Watson, A. Rutman, G. W. Taylor, P. J. Cole, Measurement of Pseudomonas aeruginosa phenazine pigments in sputum and assessment of their contribution to sputum sol toxicity for respiratory epithelium. *Infection and Immunity* **56**, 2515–2517 (1988); published online EpubSep.

61 H. Ran, D. J. Hassett, G. W. Lau, Human targets of Pseudomonas aeruginosa pyocyanin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14315–14320 (2003); published online EpubNov 25 (10.1073/pnas.2332354100).

62 Y. Li, H. P. Qu, J. L. Liu, H. Y. Wan, Correlation between group behavior and quorum sensing in Pseudomonas aeruginosa isolated from patients with hospital-acquired pneumonia. *Journal of Thoracic Disease* **6**, 810–817 (2014); published online EpubJun (10.3978/j.issn.2072-1439.2014.03.37).

63 G. A. O'Toole, R. Kolter, Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Molecular Microbiology* **30**, 295–304 (1998); published online EpubOct (10.1046/j.1365-2958.1998.01062.x).

64 R. A. Alm, J. S. Mattick, Genes involved in the biogenesis and function of type-4 fimbriae in Pseudomonas aeruginosa. *Gene* **192**, 89–98 (1997); published online EpubJun 11 (10.1016/S0378-1119(96)00805-0).

65 N. Dasgupta, M. C. Wolfgang, A. L. Goodman, S. K. Arora, J. Jyot, S. Lory, R. Ramphal, A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in Pseudomonas aeruginosa. *Molecular microbiology***50**, 809–824 (2003); published online EpubNov (10.1046/j.1365-2958.2003.03740.x).

66 Z. Kuang, Y. Hao, S. Hwang, S. Zhang, E. Kim, H. T. Akinbi, M. J. Schurr, R. T. Irvin, D. J. Hassett, G. W. Lau, The Pseudomonas aeruginosa flagellum confers resistance to pulmonary surfactant protein-A by impacting the production of exoproteases through quorum-sensing. *Molecular Microbiology* **79**, 1220–1235 (2011); published online EpubMar (10.1111/j.1365-2958.2010.07516.x).

67 L. M. Cobb, J. C. Mychaleckyj, D. J. Wozniak, Y. S. Lopez-Boado, Pseudomonas aeruginosa flagellin and alginate elicit very distinct gene expression patterns in airway epithelial cells: implications for cystic fibrosis disease. *Journal of Immunology* **173**, 5659–5670 (2004); published online EpubNov 1 (10.4049/jimmunol.173.9.5659).

68 X. Luan, V. A. Campanucci, M. Nair, O. Yilmaz, G. Belev, T. E. Machen, D. Chapman, J. P. Ianowski, Pseudomonas aeruginosa triggers CFTR-mediated airway surface liquid secretion in swine trachea. *Proceedings of the National Academy of Sciences of the United States of America* **111**(35):12930–12935 (2014); published online EpubAug 18 (10.1073/pnas.1406414111).

69 A. Stintzi, K. Evans, J. M. Meyer, K. Poole, Quorum-sensing and siderophore biosynthesis in Pseudomonas aeruginosa: lasR/lasI mutants exhibit reduced pyoverdine biosynthesis. *FEMS Microbiology Letters* **166**, 341–345 (1998); published online EpubSep 15 (10.1111/j.1574-6968.1998.tb13910.x).

70 P. Cornelis, S. Aendekerk, A new regulator linking quorum sensing and iron uptake in Pseudomonas aeruginosa. *Microbiology* **150**, 752–756 (2004); published online EpubApr (10.1099/mic.0.27086-0).

71 T. G. Sana, A. Hachani, I. Bucior, C. Soscia, S. Garvis, E. Termine, J. Engel, A. Filloux, S. Bleves, The second type VI secretion system of Pseudomonas aeruginosa strain PAO1 is regulated by quorum sensing and Fur and modulates internalization in epithelial cells. *The Journal of Biological Chemistry* **287**, 27095–27105 (2012); published online EpubAug 3 (10.1074/jbc.M112.376368).

72 R. D. Hood, P. Singh, F. Hsu, T. Guvener, M. A. Carl, R. R. Trinidad, J. M. Silverman, B. B. Ohlson, K. G. Hicks, R. L. Plemel, M. Li, S. Schwarz, W. Y. Wang, A. J. Merz, D. R. Goodlett, J. D. Mougous, A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell Host & Microbe* **7**, 25–37 (2010); published online EpubJan 21 (10.1016/j.chom.2009.12.007).

73 N. S. Lossi, E. Manoli, P. Simpson, C. Jones, K. Hui, R. Dajani, S. J. Coulthurst, P. Freemont, A. Filloux, The archetype Pseudomonas aeruginosa proteins TssB and TagJ form a novel subcomplex in the bacterial type VI secretion system. *Molecular Microbiology* **86**, 437–456 (2012); published online EpubOct (10.1111/j.1365-2958.2012.08204.x).

74 H. Mikkelsen, M. Sivaneson, A. Filloux, Key two-component regulatory systems that control biofilm formation in Pseudomonas aeruginosa. *Environmental Microbiology* **13**, 1666–1681 (2011); published online EpubJul (10.1111/j.1462-2920.2011.02495.x).

75 J. A. Moscoso, H. Mikkelsen, S. Heeb, P. Williams, A. Filloux, The Pseudomonas aeruginosa sensor RetS switches type III and type VI secretion via c-di-GMP signalling.

Environmental Microbiology **13**, 3128–3138 (2011); published online EpubDec (10.1111/j.1462-2920.2011.02595.x).

76 H. Mikkelsen, R. McMullan, A. Filloux, The Pseudomonas aeruginosa reference strain PA14 displays increased virulence due to a mutation in ladS. *PLoS ONE* **6**, e29113 (2011)10.1371/journal.pone.0029113).

77 I. Ventre, A. L. Goodman, I. Vallet-Gely, P. Vasseur, C. Soscia, S. Molin, S. Bleves, A. Lazdunski, S. Lory, A. Filloux, Multiple sensors control reciprocal expression of Pseudomonas aeruginosa regulatory RNA and virulence genes. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 171–176 (2006); published online EpubJan 3 (10.1073/pnas.0507407103).

78 A. L. Goodman, B. Kulasekara, A. Rietsch, D. Boyd, R. S. Smith, S. Lory, A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in Pseudomonas aeruginosa. *Developmental Cell* **7**, 745–754 (2004); published online EpubNov (10.1016/j.devcel.2004.08.020).

79 A. L. Goodman, M. Merighi, M. Hyodo, I. Ventre, A. Filloux, S. Lory, Direct interaction between sensor kinase proteins mediates acute and chronic disease phenotypes in a bacterial pathogen. *Genes & Development* **23**, 249–259 (2009); published online EpubJan 15 (10.1101/gad.1739009).

80 A. Brencic, K. A. McFarland, H. R. McManus, S. Castang, I. Mogno, S. L. Dove, S. Lory, The GacS/GacA signal transduction system of Pseudomonas aeruginosa acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. *Molecular Microbiology* **73**, 434–445 (2009); published online EpubAug (10.1111/j.1365-2958.2009.06782.x).

81 C. Valverde, S. Heeb, C. Keel, D. Haas, RsmY, a small regulatory RNA, is required in concert with RsmZ for GacA-dependent expression of biocontrol traits in Pseudomonas fluorescens CHA0. *Molecular Microbiology* **50**, 1361–1379 (2003); published online EpubNov (10.1046/j.1365-2958.2003.03774.x).

82 E. Kay, B. Humair, V. Denervaud, K. Riedel, S. Spahr, L. Eberl, C. Valverde, D. Haas, Two GacA-dependent small RNAs modulate the quorum-sensing response in Pseudomonas aeruginosa. *Journal of bacteriology***188**, 6026–6033 (2006); published online EpubAug (188/16/6026 [pii]10.1128/JB.00409-06).

83 C. Reimmann, M. Beyeler, A. Latifi, H. Winteler, M. Foglino, A. Lazdunski, D. Haas, The global activator GacA of Pseudomonas aeruginosa PAO positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. *Molecular Microbiology* **24**, 309–319 (1997); published online EpubApr.

84 G. Pessi, D. Haas, Dual control of hydrogen cyanide biosynthesis by the global activator GacA in Pseudomonas aeruginosa PAO1. *FEMS Microbiology Letters* **200**, 73–78 (2001); published online EpubJun 12.

85 M. C. Maiden, J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, B. G. Spratt, Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 3140–3145 (1998); published online EpubMar 17.

86 H. Rohde, J. Qin, Y. Cui, D. Li, N. J. Loman, M. Hentschke, W. Chen, F. Pu, Y. Peng, J. Li, F. Xi, S. Li, Y. Li, Z. Zhang, X. Yang, M. Zhao, P. Wang, Y. Guan, Z. Cen, X. Zhao, M. Christner, R. Kobbe, S. Loos, J. Oh, L. Yang, A. Danchin, G. F. Gao, Y. Song, Y. Li,

H. Yang, J. Wang, J. Xu, M. J. Pallen, J. Wang, M. Aepfelbacher, R. Yang, E. c. O. H. G. A. C.-S. Consortium, Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *The New ENGLAND Journal of Medicine* **365**, 718–724 (2011); published online EpubAug 25 (10.1056/NEJMoa1107643).

87 T. Zhang, M. F. Shao, L. Ye, 454 pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. *The ISME Journal* **6**, 1137–1147 (2012); published online EpubJun (10.1038/ismej.2011.188).

88 S. Junemann, K. Prior, R. Szczepanowski, I. Harks, B. Ehmke, A. Goesmann, J. Stoye, D. Harmsen, Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon sequencing. *PLoS ONE* **7**, e41606 (2012)10.1371/journal.pone.0041606).

89 S. H. Lin, Y. C. Liao, CISA: Contig integrator for sequence assembly of bacterial genomes. *PLoS ONE* **8**, e60843 (2013)10.1371/journal.pone.0060843).

90 R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008) 10.1186/1471-2164-9-75).

91 D. J. Edwards, K. E. Holt, Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microbial Informatics and Experimentation* **3**, 2 (2013)10.1186/2042-5783-3-2).

92 B. P. Howden, C. R. McEvoy, D. L. Allen, K. Chua, W. Gao, P. F. Harrison, J. Bell, G. Coombs, V. Bennett-Wood, J. L. Porter, R. Robins-Browne, J. K. Davies, T. Seemann, T. P. Stinear, Evolution of multidrug resistance during Staphylococcus aureus infection involves mutation of the essential two component regulator WalKR. *PLoS Pathogens* **7**, e1002359 (2011); published online EpubNov (10.1371/journal.ppat.1002359).

93 E. S. Snitkin, A. M. Zelazny, P. J. Thomas, F. Stock, N. C. S. P. Group, D. K. Henderson, T. N. Palmore, J. A. Segre, Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Science Translational Medicine* **4**, 148ra116 (2012); published online EpubAug 22 (10.1126/scitranslmed.3004129).

94 S. R. Harris, E. J. Cartwright, M. E. Torok, M. T. Holden, N. M. Brown, A. L. Ogilvy-Stuart, M. J. Ellington, M. A. Quail, S. D. Bentley, J. Parkhill, S. J. Peacock, Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: A descriptive study. *The Lancet. Infectious Diseases* **13**, 130–136 (2013); published online EpubFeb (10.1016/S1473-3099(12)70268-2).

95 K. E. Holt, S. Baker, F. X. Weill, E. C. Holmes, A. Kitchen, J. Yu, V. Sangal, D. J. Brown, J. E. Coia, D. W. Kim, S. Y. Choi, S. H. Kim, W. D. da Silveira, D. J. Pickard, J. J. Farrar, J. Parkhill, G. Dougan, N. R. Thomson, Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nature Genetics* **44**, 1056–1059 (2012); published online EpubSep (10.1038/ng.2369).

96 S. P. Gygi, G. L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9390–9395 (2000); published online EpubAug 15 (10.1073/pnas.160270797).

97 P. Seng, M. Drancourt, F. Gouriet, B. La Scola, P. E. Fournier, J. M. Rolain, D. Raoult, Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted

laser desorption ionization time-of-flight mass spectrometry. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* **49**, 543–551 (2009); published online EpubAug 15 (10.1086/600885).

98 P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D. J. Pappin, Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics: MCP* **3**, 1154–1169 (2004); published online EpubDec (10.1074/mcp.M400129-MCP200).

99 Y. Ni, L. Song, X. Qian, Z. Sun, Proteomic analysis of Pseudomonas putida reveals an organic solvent tolerance-related gene mmsB. *PLoS ONE* **8**, e55858 (2013)10.1371/journal.pone.0055858).

100 A. Coelho, E. de Oliveira Santos, M. L. Faria, D. P. de Carvalho, M. R. Soares, W. M. von Kruger, P. M. Bisch, A proteome reference map for Vibrio cholerae El Tor. *Proteomics* **4**, 1491–1504 (2004); published online EpubMay (10.1002/pmic.200300685).

101 I. Catrein, R. Herrmann, The proteome of Mycoplasma pneumoniae, a supposedly "simple" cell. *Proteomics* **11**, 3614–3632 (2011); published online EpubSep (10.1002/pmic.201100076).

102 N. E. Sherman, B. Stefansson, J. W. Fox, J. B. Goldberg, Pseudomonas aeruginosa and a proteomic approach to bacterial pathogenesis. *Disease Markers* **17**, 285–293 (2001).

103 G. L. Winsor, D. K. Lam, L. Fleming, R. Lo, M. D. Whiteside, N. Y. Yu, R. E. Hancock, F. S. Brinkman, Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. *Nucleic Acids Research* **39**, D596–600 (2011); published online EpubJan (10.1093/nar/gkq869gkq869 [pii]).

104 S. Tan, H. T. Tan, M. C. Chung, Membrane proteins and membrane proteomics. *Proteomics* **8**, 3924–3932 (2008); published online EpubOct (10.1002/pmic.200800597).

105 S. L. Chua, Y. Liu, J. K. Yam, Y. Chen, R. M. Vejborg, B. G. Tan, S. Kjelleberg, T. Tolker-Nielsen, M. Givskov, L. Yang, Dispersed cells represent a distinct stage in the transition from bacterial biofilm to planktonic lifestyles. *Nature Communications* **5**, 4462 (2014)10.1038/ncomms5462).

106 F. Xu, H. Zhao, X. Feng, L. Chen, D. Chen, Y. Zhang, F. Nan, J. Liu, B. F. Liu, Single-cell chemical proteomics with an activity-based probe: Identification of low-copy membrane proteins on primary neurons. *Angewandte Chemie* **53**, 6730–6733 (2014); published online EpubJun 23 (10.1002/anie.201402363).

107 B. F. Cravatt, G. M. Simon, J. R. Yates, 3rd, The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000 (2007); published online EpubDec 13 (10.1038/nature06525).

108 K. Mathee, G. Narasimhan, C. Valdes, X. Qiu, J. M. Matewish, M. Koehrsen, A. Rokas, C. N. Yandava, R. Engels, E. Zeng, R. Olavarietta, M. Doud, R. S. Smith, P. Montgomery, J. R. White, P. A. Godfrey, C. Kodira, B. Birren, J. E. Galagan, S. Lory, Dynamics of Pseudomonas aeruginosa genome evolution. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3100–3105 (2008); published online EpubFeb 26 (10.1073/pnas.0711982105).

109 C. Winstanley, M. G. Langille, J. L. Fothergill, I. Kukavica-Ibrulj, C. Paradis-Bleau, F. Sanschagrin, N. R. Thomson, G. L. Winsor, M. A. Quail, N. Lennard, A. Bignell,

L. Clarke, K. Seeger, D. Saunders, D. Harris, J. Parkhill, R. E. Hancock, F. S. Brinkman, R. C. Levesque, Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome Research* **19**, 12–23 (2009); published online EpubJan (10.1101/gr.086082.108).

110 X. Qiu, B. R. Kulasekara, S. Lory, Role of horizontal gene transfer in the evolution of Pseudomonas aeruginosa virulence. *Genome Dynamics* **6**, 126–139 (2009)10.1159/000235767).

111 D. Medini, C. Donati, H. Tettelin, V. Masignani, R. Rappuoli, The microbial pan-genome. *Current opinion in Genetics & Development* **15**, 589–594 (2005); published online EpubDec (10.1016/j.gde.2005.09.006).

112 U. Romling, K. D. Schmidt, B. Tummler, Large genome rearrangements discovered by the detailed analysis of 21 Pseudomonas aeruginosa clone C isolates found in environment and disease habitats. *Journal of Molecular Biology* **271**, 386–404 (1997); published online EpubAug 22 (10.1006/jmbi.1997.1186).

113 A. U. Kresse, S. D. Dinesh, K. Larbig, U. Romling, Impact of large chromosomal inversions on the adaptation and evolution of Pseudomonas aeruginosa chronically colonizing cystic fibrosis lungs. *Molecular Microbiology* **47**, 145–158 (2003); published online EpubJan.

114 R. K. Ernst, D. A. D'Argenio, J. K. Ichikawa, M. G. Bangera, S. Selgrade, J. L. Burns, P. Hiatt, K. McCoy, M. Brittnacher, A. Kas, D. H. Spencer, M. V. Olson, B. W. Ramsey, S. Lory, S. I. Miller, Genome mosaicism is conserved but not unique in Pseudomonas aeruginosa isolates from the airways of young children with cystic fibrosis. *Environmental Microbiology* **5**, 1341–1349 (2003); published online EpubDec (10.1111/j.1462-2920.2003.00518.x).

115 J. Klockgether, N. Cramer, L. Wiehlmann, C. F. Davenport, B. Tummler, Pseudomonas aeruginosa genomic structure and diversity. *Frontiers in Microbiology* **2**, 150 (2011)10.3389/fmicb.2011.00150).

116 D. H. Spencer, A. Kas, E. E. Smith, C. K. Raymond, E. H. Sims, M. Hastings, J. L. Burns, R. Kaul, M. V. Olson, Whole-genome sequence variation among multiple isolates of Pseudomonas aeruginosa. *Journal of Bacteriology* **185**, 1316–1325 (2003); published online EpubFeb.

117 A. Folkesson, L. Jelsbak, L. Yang, H. K. Johansen, O. Ciofu, N. Hoiby, S. Molin, Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: An evolutionary perspective. *Nature Reviews. Microbiology* **10**, 841–851 (2012); published online EpubDec (10.1038/nrmicro2907).

118 E. E. Smith, D. G. Buckley, Z. Wu, C. Saenphimmachak, L. R. Hoffman, D. A. D'Argenio, S. I. Miller, B. W. Ramsey, D. P. Speert, S. M. Moskowitz, J. L. Burns, R. Kaul, M. V. Olson, Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8487–8492 (2006); published online EpubMay 30 (10.1073/pnas.0602138103).

119 L. Yang, L. Jelsbak, R. L. Marvig, S. Damkiaer, C. T. Workman, M. H. Rau, S. K. Hansen, A. Folkesson, H. K. Johansen, O. Ciofu, N. Hoiby, M. O. Sommer, S. Molin, Evolutionary dynamics of bacteria in a human host environment. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7481–7486 (2011); published online EpubMay 3 (10.1073/pnas.1018249108).

120 E. Mahenthiralingam, M. E. Campbell, D. P. Speert, Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis. *Infection and Immunity* **62**, 596–605 (1994); published online EpubFeb 121.

121 L. Wiehlmann, G. Wagner, N. Cramer, B. Siebert, P. Gudowius, G. Morales, T. Kohler, C. van Delden, C. Weinel, P. Slickers, B. Tummler, Population structure of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8101–8106 (2007); published online EpubMay 8 (10.1073/pnas.0609213104).

122 C. M. Shaver, A. R. Hauser, Relative contributions of Pseudomonas aeruginosa ExoU, ExoS, and ExoT to virulence in the lung. *Infection and Immunity* **72**, 6969–6977 (2004); published online EpubDec (10.1128/IAI.72.12.6969-6977.2004).

123 K. D. Larbig, A. Christmann, A. Johann, J. Klockgether, T. Hartsch, R. Merkl, L. Wiehlmann, H. J. Fritz, B. Tummler, Gene islands integrated into tRNA(Gly) genes confer genome diversity on a Pseudomonas aeruginosa clone. *Journal of Bacteriology* **184**, 6665–6680 (2002); published online EpubDec.

124 U. Romling, J. Wingender, H. Muller, B. Tummler, A major Pseudomonas aeruginosa clone common to patients and aquatic habitats. *Applied and Environmental Microbiology* **60**, 1734–1738 (1994); published online EpubJun.

125 N. Cramer, J. Klockgether, K. Wrasman, M. Schmidt, C. F. Davenport, B. Tummler, Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs. *Environmental Microbiology* **13**, 1690–1704 (2011); published online EpubJul (10.1111/j.1462-2920.2011.02483.x).

126 J. Klockgether, N. Miethke, P. Kubesch, Y. S. Bohn, I. Brockhausen, N. Cramer, L. Eberl, J. Greipel, C. Herrmann, S. Herrmann, S. Horatzek, M. Lingner, L. Luciano, P. Salunkhe, D. Schomburg, M. Wehsling, L. Wiehlmann, C. F. Davenport, B. Tummler, Intraclonal diversity of the Pseudomonas aeruginosa cystic fibrosis airway isolates TBCF10839 and TBCF121838: distinct signatures of transcriptome, proteome, metabolome, adherence and pathogenicity despite an almost identical genome sequence. *Environmental Microbiology* **15**, 191–210 (2013); published online EpubJan (10.1111/j.1462-2920.2012.02842.x).

127 S. E. Finkel, R. Kolter, Evolution of microbial diversity during prolonged starvation. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4023–4027 (1999); published online EpubMar 30 (10.1073/pnas.96.7.4023).

128 L. Jelsbak, H. K. Johansen, A. L. Frost, R. Thogersen, L. E. Thomsen, O. Ciofu, L. Yang, J. A. Haagensen, N. Hoiby, S. Molin, Molecular epidemiology and dynamics of Pseudomonas aeruginosa populations in lungs of cystic fibrosis patients. *Infection and immunity***75**, 2214–2224 (2007); published online EpubMay (10.1128/IAI.01282-06).

129 M. Anthony, B. Rose, M. B. Pegler, M. Elkins, H. Service, K. Thamotharampillai, J. Watson, M. Robinson, P. Bye, J. Merlino, C. Harbour, Genetic analysis of Pseudomonas aeruginosa isolates from the sputa of Australian adult cystic fibrosis patients. *Journal of Clinical Microbiology* **40**, 2772–2778 (2002); published online EpubAug (10.1128/JCM.40.8.2772-2778.2002).

130 M. Denton, K. Kerr, L. Mooney, V. Keer, A. Rajgopal, K. Brownlee, P. Arundel, S. Conway, Transmission of colistin-resistant Pseudomonas aeruginosa between

patients attending a pediatric cystic fibrosis center. *Pediatric Pulmonology* **34**, 257–261 (2002); published online EpubOct (10.1002/ppul.10166).

131 F. W. Scott, T. L. Pitt, Identification and characterization of transmissible Pseudomonas aeruginosa strains in cystic fibrosis patients in England and Wales. *Journal of Medical Microbiology* **53**, 609–615 (2004); published online EpubJul.

132 E. T. Jensen, B. Giwercman, B. Ojeniyi, J. M. Bangsborg, A. Hansen, C. Koch, N. E. Fiehn, N. Hoiby, Epidemiology of Pseudomonas aeruginosa in cystic fibrosis and the possible role of contamination by dental equipment. *The Journal of Hospital Infection* **36**, 117–122 (1997); published online EpubJun (10.1016/S0195-6701(97)90117-1).

133 E. Mahenthiralingam, M. E. Campbell, J. Foster, J. S. Lam, D. P. Speert, Random amplified polymorphic DNA typing of Pseudomonas aeruginosa isolates recovered from patients with cystic fibrosis. *Journal of Clinical Microbiology* **34**, 1129–1135 (1996); published online EpubMay.

134 U. Romling, B. Fiedler, J. Bosshammer, D. Grothues, J. Greipel, H. von der Hardt, B. Tummler, Epidemiology of chronic Pseudomonas aeruginosa infections in cystic fibrosis. *The Journal of Infectious Diseases* **170**, 1616–1621 (1994); published online EpubDec.

135 K. Cheng, R. L. Smyth, J. R. Govan, C. Doherty, C. Winstanley, N. Denning, D. P. Heaf, H. van Saene, C. A. Hart, Spread of beta-lactam-resistant Pseudomonas aeruginosa in a cystic fibrosis clinic. *Lancet* **348**, 639–642 (1996); published online EpubSep 7 (10.1016/S0140-6736(96)05169-0).

136 S. D. Aaron, K. L. Vandemheen, K. Ramotar, T. Giesbrecht-Lewis, E. Tullis, A. Freitag, N. Paterson, M. Jackson, M. D. Lougheed, C. Dowson, V. Kumar, W. Ferris, F. Chan, S. Doucette, D. Fergusson, Infection with transmissible strains of Pseudomonas aeruginosa and clinical outcomes in adults with cystic fibrosis. *JAMA* **304**, 2145–2153 (2010); published online EpubNov 17 (10.1001/jama.2010.1665).

137 A. Ashish, M. Shaw, C. Winstanley, M. J. Ledson, M. J. Walshaw, Increasing resistance of the Liverpool Epidemic Strain (LES) of Pseudomonas aeruginosa (Psa) to antibiotics in cystic fibrosis (CF) – a cause for concern? *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **11**, 173–179 (2012); published online EpubMay (10.1016/j.jcf.2011.11.004).

138 P. Salunkhe, C. H. Smart, J. A. Morgan, S. Panagea, M. J. Walshaw, C. A. Hart, R. Geffers, B. Tummler, C. Winstanley, A cystic fibrosis epidemic strain of Pseudomonas aeruginosa displays enhanced virulence and antimicrobial resistance. *Journal of Bacteriology* **187**, 4908–4920 (2005); published online EpubJul (10.1128/JB.187.14.4908-4920.2005).

139 M. Al-Aloul, J. Crawley, C. Winstanley, C. A. Hart, M. J. Ledson, M. J. Walshaw, Increased morbidity associated with chronic infection by an epidemic Pseudomonas aeruginosa strain in CF patients. *Thorax* **59**, 334–336 (2004); published online EpubApr.

140 M. E. Carter, J. L. Fothergill, M. J. Walshaw, K. Rajakumar, A. Kadioglu, C. Winstanley, A subtype of a Pseudomonas aeruginosa cystic fibrosis epidemic strain exhibits enhanced virulence in a murine model of acute respiratory infection. *The Journal of Infectious Diseases* **202**, 935–942 (2010); published online EpubSep 15 (10.1086/655781).

141 J. L. Fothergill, S. Panagea, C. A. Hart, M. J. Walshaw, T. L. Pitt, C. Winstanley, Widespread pyocyanin over-production among isolates of a cystic fibrosis epidemic strain. *BMC Microbiology* **7**, 45 (2007)10.1186/1471-2180-7-45).

142 J. Jeukens, B. Boyle, I. Kukavica-Ibrulj, M. M. Ouellet, S. D. Aaron, S. J. Charette, J. L. Fothergill, N. P. Tucker, C. Winstanley, R. C. Levesque, Comparative genomics of isolates of a Pseudomonas aeruginosa epidemic strain associated with chronic lung infections of cystic fibrosis patients. *PLoS ONE* **9**, e87611 (2014)10.1371/journal.pone.0087611).

143 J. L. Fothergill, E. Mowat, M. J. Ledson, M. J. Walshaw, C. Winstanley, Fluctuations in phenotypes and genotypes within populations of Pseudomonas aeruginosa in the cystic fibrosis lung during pulmonary exacerbations. *Journal of Medical Microbiology* **59**, 472–481 (2010); published online EpubApr (10.1099/jmm.0.015875-0).

144 T. Hindre, C. Knibbe, G. Beslon, D. Schneider, New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Reviews. Microbiology* **10**, 352–365 (2012); published online EpubMay (10.1038/nrmicro2750).

145 M. H. Rau, S. K. Hansen, H. K. Johansen, L. E. Thomsen, C. T. Workman, K. F. Nielsen, L. Jelsbak, N. Hoiby, L. Yang, S. Molin, Early adaptive developments of Pseudomonas aeruginosa after the transition from life in the environment to persistent colonization in the airways of human cystic fibrosis hosts. *Environmental Microbiology* **12**, 1643–1658 (2010); published online EpubJun (10.1111/j.1462-2920.2010.02211.x).

146 E. Mowat, S. Paterson, J. L. Fothergill, E. A. Wright, M. J. Ledson, M. J. Walshaw, M. A. Brockhurst, C. Winstanley, Pseudomonas aeruginosa population diversity and turnover in cystic fibrosis chronic infections. *American Journal of Respiratory and Critical Care Medicine* **183**, 1674–1679 (2011); published online EpubJun 15 (10.1164/rccm.201009-1430OC).

147 S. A. West, A. S. Griffin, A. Gardner, S. P. Diggle, Social evolution theory for microorganisms. *Nature Reviews. Microbiology* **4**, 597–607 (2006); published online EpubAug (10.1038/nrmicro1461).

148 C. N. Wilder, S. P. Diggle, M. Schuster, Cooperation and cheating in Pseudomonas aeruginosa: the roles of the las, rhl and pqs quorum-sensing systems. *The ISME Journal* **5**, 1332–1343 (2011); published online EpubAug (10.1038/ismej.2011.13).

149 K. M. Sandoz, S. M. Mitzimberg, M. Schuster, Social cheating in Pseudomonas aeruginosa quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 15876–15881 (2007); published online EpubOct 2 (10.1073/pnas.0705653104).

150 A. A. Dandekar, S. Chugani, E. P. Greenberg, Bacterial quorum sensing and metabolic incentives to cooperate. *Science* **338**, 264–266 (2012); published online EpubOct 12 (10.1126/science.1227289).

151 R. L. Marvig, H. K. Johansen, S. Molin, L. Jelsbak, Genome analysis of a transmissible lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genetics* **9**, e1003741 (2013)10.1371/journal.pgen.1003741).

152 M. R. Weigand, G. W. Sundin, General and inducible hypermutation facilitate parallel adaptation in Pseudomonas aeruginosa despite divergent mutation spectra. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13680–13685 (2012); published online EpubAug 21 (10.1073/pnas.1205357109).

153 A. Oliver, R. Canton, P. Campo, F. Baquero, J. Blazquez, High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science* **288**, 1251–1254 (2000); published online EpubMay 19 (10.1126/science.288.5469.1251).

154 C. Lopez-Causape, E. Rojo-Molinero, X. Mulet, G. Cabot, B. Moya, J. Figuerola, B. Togores, J. L. Perez, A. Oliver, Clonal dissemination, emergence of mutator lineages and antibiotic resistance evolution in Pseudomonas aeruginosa cystic fibrosis chronic lung infection. *PLoS ONE* **8**, e71001 (2013)10.1371/journal.pone.0071001).

155 L. R. Hoffman, H. D. Kulasekara, J. Emerson, L. S. Houston, J. L. Burns, B. W. Ramsey, S. I. Miller, Pseudomonas aeruginosa lasR mutants are associated with cystic fibrosis lung disease progression. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **8**, 66–70 (2009); published online EpubJan (10.1016/j.jcf.2008.09.006).

156 C. N. Wilder, G. Allada, M. Schuster, Instantaneous within-patient diversity of Pseudomonas aeruginosa quorum-sensing populations from cystic fibrosis lung infections. *Infection and Immunity* **77**, 5631–5639 (2009); published online EpubDec (10.1128/IAI.00755-09).

157 D. A. D'Argenio, M. Wu, L. R. Hoffman, H. D. Kulasekara, E. Deziel, E. E. Smith, H. Nguyen, R. K. Ernst, T. J. Larson Freeman, D. H. Spencer, M. Brittnacher, H. S. Hayden, S. Selgrade, M. Klausen, D. R. Goodlett, J. L. Burns, B. W. Ramsey, S. I. Miller, Growth phenotypes of Pseudomonas aeruginosa lasR mutants adapted to the airways of cystic fibrosis patients. *Molecular Microbiology* **64**, 512–533 (2007); published online EpubApr (10.1111/j.1365-2958.2007.05678.x).

158 T. Nishijyo, D. Haas, Y. Itoh, The CbrA-CbrB two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in Pseudomonas aeruginosa. *Molecular Microbiology* **40**, 917–931 (2001); published online EpubMay.

159 A. M. Lujan, A. J. Moyano, I. Segura, C. E. Argarana, A. M. Smania, Quorum-sensing-deficient (lasR) mutants emerge at high frequency from a Pseudomonas aeruginosa mutS strain. *Microbiology* **153**, 225–237 (2007); published online EpubJan (10.1099/mic.0.29021-0).

160 K. L. Palmer, L. M. Mashburn, P. K. Singh, M. Whiteley, Cystic fibrosis sputum supports growth and cues key aspects of Pseudomonas aeruginosa physiology. *Journal of Bacteriology* **187**, 5267–5277 (2005); published online EpubAug (10.1128/JB.187.15.5267-5277.2005).

161 D. A. D'Argenio, M. W. Calfee, P. B. Rainey, E. C. Pesci, Autolysis and autoaggregation in Pseudomonas aeruginosa colony morphology mutants. *Journal of Bacteriology* **184**, 6481–6489 (2002); published online EpubDec (10.1128/JB.184.23.6481-6489.2002).

162 A. L. Barth, T. L. Pitt, The high amino-acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic Pseudomonas aeruginosa. *Journal of Medical Microbiology* **45**, 110–119 (1996); published online EpubAug (10.1099/00222615-45-2-110).

163 K. L. Palmer, L. M. Aye, M. Whiteley, Nutritional cues control Pseudomonas aeruginosa multicellular behavior in cystic fibrosis sputum. *Journal of Bacteriology* **189**, 8079–8087 (2007); published online EpubNov (10.1128/JB.01138-07).

164 M. J. Gambello, B. H. Iglewski, Cloning and characterization of the Pseudomonas aeruginosa lasR gene, a transcriptional activator of elastase expression. *Journal of Bacteriology* **173**, 3000–3009 (1991); published online EpubMay.

165 L. Yang, M. H. Rau, L. Yang, N. Hoiby, S. Molin, L. Jelsbak, Bacterial adaptation during chronic infection revealed by independent component analysis of transcriptomic data. *BMC Microbiology* **11**, 184 (2011)10.1186/1471-2180-11-184).

166 P. A. Totten, J. C. Lara, S. Lory, The rpoN gene product of Pseudomonas aeruginosa is required for expression of diverse genes, including the flagellin gene. *Journal of Bacteriology* **172**, 389–396 (1990); published online EpubJan 167.

167 L. S. Thompson, J. S. Webb, S. A. Rice, S. Kjelleberg, The alternative sigma factor RpoN regulates the quorum sensing gene rhlI in Pseudomonas aeruginosa. *FEMS Microbiology Letters* **220**, 187–195 (2003); published online EpubMar 28 ((10.1016/S0378-1097(03)00097-1).

168 E. Mahenthiralingam, D. P. Speert, Nonopsonic phagocytosis of Pseudomonas aeruginosa by macrophages and polymorphonuclear leukocytes requires the presence of the bacterial flagellum. *Infection and Immunity* **63**, 4519–4523 (1995); published online EpubNov.

169 J. R. Wiens, A. I. Vasil, M. J. Schurr, M. L. Vasil, Iron-regulated expression of alginate production, mucoid phenotype, and biofilm formation by Pseudomonas aeruginosa. *mBio* **5**, e01010–01013 (2014)10.1128/mBio.01010-13).

170 S. Feliziani, A. M. Lujan, A. J. Moyano, C. Sola, J. L. Bocco, P. Montanaro, L. F. Canigia, C. E. Argarana, A. M. Smania, Mucoidy, quorum sensing, mismatch repair and antibiotic resistance in Pseudomonas aeruginosa from cystic fibrosis chronic airways infections. *PLoS ONE* **5**, 1–12 (2010) 10.1371/journal.pone.0012669).

171 J. R. Aires, T. Kohler, H. Nikaido, P. Plesiat, Involvement of an active efflux system in the natural resistance of Pseudomonas aeruginosa to aminoglycosides. *Antimicrobial Agents and Chemotherapy* **43**, 2624–2628 (1999); published online EpubNov 172.

172 S. Westbrock-Wadman, D. R. Sherman, M. J. Hickey, S. N. Coulter, Y. Q. Zhu, P. Warrener, L. Y. Nguyen, R. M. Shawar, K. R. Folger, C. K. Stover, Characterization of a Pseudomonas aeruginosa efflux pump contributing to aminoglycoside impermeability. *Antimicrobial Agents and Chemotherapy* **43**, 2975–2983 (1999); published online EpubDec 173.

173 M. L. Sobel, G. A. McKay, K. Poole, Contribution of the MexXY multidrug transporter to aminoglycoside resistance in Pseudomonas aeruginosa clinical isolates. *Antimicrobial Agents and Chemotherapy* **47**, 3202–3207 (2003); published online EpubOct (10.1128/AAC.47.10.3202-3207.2003).

174 A. M. Firoved, J. C. Boucher, V. Deretic, Global genomic analysis of AlgU (sigma(E))-dependent promoters (sigmulon) in Pseudomonas aeruginosa and implications for inflammatory processes in cystic fibrosis. *Journal of Bacteriology* **184**, 1057–1064 (2002); published online EpubFeb (10.1128/jb.184.4.1057-1064.2002).

175 S. K. Hansen, M. H. Rau, H. K. Johansen, O. Ciofu, L. Jelsbak, L. Yang, A. Folkesson, H. O. Jarmer, K. Aanaes, C. von Buchwald, N. Hoiby, S. Molin, Evolution and diversification of Pseudomonas aeruginosa in the paranasal sinuses of cystic fibrosis children have implications for chronic lung infection. *The ISME Journal* **6**, 31–45 (2012); published online EpubJan (10.1038/ismej.2011.83).

176 H. K. Huse, T. Kwon, J. E. Zlosnik, D. P. Speert, E. M. Marcotte, M. Whiteley, Parallel evolution in Pseudomonas aeruginosa over 39,000 generations in vivo. *mBio* **1**, (2010)10.1128/mBio.00199-10).

177 C. Hoboth, R. Hoffmann, A. Eichner, C. Henke, S. Schmoldt, A. Imhof, J. Heesemann, M. Hogardt, Dynamics of adaptive microevolution of hypermutable Pseudomonas

aeruginosa during chronic pulmonary infection in patients with cystic fibrosis. *The Journal of Infectious Diseases* **200**, 118–130 (2009); published online EpubJul 1 (10.1086/599360).

178 D. Worlitzsch, R. Tarran, M. Ulrich, U. Schwab, A. Cekici, K. C. Meyer, P. Birrer, G. Bellon, J. Berger, T. Weiss, K. Botzenhart, J. R. Yankaskas, S. Randell, R. C. Boucher, G. Doring, Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. *The Journal of Clinical Investigation* **109**, 317–325 (2002); published online EpubFeb (10.1172/JCI13870).

179 M. Kolpen, M. Kuhl, T. Bjarnsholt, C. Moser, C. R. Hansen, L. Liengaard, A. Kharazmi, T. Pressler, N. Hoiby, P. O. Jensen, Nitrous oxide production in sputum from cystic fibrosis patients with chronic Pseudomonas aeruginosa lung infection. *PLoS ONE* **9**, e84353 (2014)10.1371/journal.pone.0084353).

180 S. S. Yoon, R. F. Hennigan, G. M. Hilliard, U. A. Ochsner, K. Parvatiyar, M. C. Kamani, H. L. Allen, T. R. DeKievit, P. R. Gardner, U. Schwab, J. J. Rowe, B. H. Iglewski, T. R. McDermott, R. P. Mason, D. J. Wozniak, R. E. Hancock, M. R. Parsek, T. L. Noah, R. C. Boucher, D. J. Hassett, Pseudomonas aeruginosa anaerobic respiration in biofilms: relationships to cystic fibrosis pathogenesis. *Developmental Cell* **3**, 593–603 (2002); published online EpubOct (10.1016/S1534-5807(02)00295-2).

181 M. Quadroni, P. James, P. Dainese-Hatt, M. A. Kertesz, Proteome mapping, mass spectrometric sequencing and reverse transcription-PCR for characterization of the sulfate starvation-induced response in Pseudomonas aeruginosa PAO1. *European Journal of Biochemistry/FEBS* **266**, 986–996 (1999); published online EpubDec (10.1046/j.1432-1327.1999.00941.x).

182 S. Wiese, K. A. Reidegeld, H. E. Meyer, B. Warscheid, Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7**, 340–350 (2007); published online EpubFeb (10.1002/pmic.200600422).

183 J. Rao, F. H. Damron, M. Basler, A. Digiandomenico, N. E. Sherman, J. W. Fox, J. J. Mekalanos, J. B. Goldberg, Comparisons of two proteomic analyses of non-mucoid and mucoid Pseudomonas aeruginosa clinical isolates from a cystic fibrosis patient. *Frontiers in Microbiology* **2**, 162 (2011)10.3389/fmicb.2011.00162).

184 S. L. Chua, S. Y. Tan, M. T. Rybtke, Y. Chen, S. A. Rice, S. Kjelleberg, T. Tolker-Nielsen, L. Yang, M. Givskov, Bis-(3′-5′)-cyclic dimeric GMP regulates antimicrobial peptide resistance in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy* **57**, 2066–2075 (2013); published online EpubMay (10.1128/AAC.02499-12AAC.02499-12 [pii]).

185 H. Mikkelsen, J. E. Swatton, K. S. Lilley, M. Welch, Proteomic analysis of the adaptive responses of Pseudomonas aeruginosa to aminoglycoside antibiotics. *FEMS Microbiology Letters,* **302**(1), 1–7 (2010); published online EpubAug 19 (10.1111/j.1574-6968.2009.01729.x).

186 T. Vinckx, Q. Wei, S. Matthijs, J. P. Noben, R. Daniels, P. Cornelis, A proteome analysis of the response of a Pseudomonas aeruginosa oxyR mutant to iron limitation. *Biometals: An International Journal on the Role of Metal ions in Biology, Biochemistry, and Medicine* **24**, 523–532 (2011); published online EpubJun (10.1007/s10534-010-9403-4).

187 M. D. Platt, M. J. Schurr, K. Sauer, G. Vazquez, I. Kukavica-Ibrulj, E. Potvin, R. C. Levesque, A. Fedynak, F. S. Brinkman, J. Schurr, S. H. Hwang, G. W. Lau,

P. A. Limbach, J. J. Rowe, M. A. Lieberman, N. Barraud, J. Webb, S. Kjelleberg, D. F. Hunt, D. J. Hassett, Proteomic, microarray, and signature-tagged mutagenesis analyses of anaerobic Pseudomonas aeruginosa at pH 6.5, likely representing chronic, late-stage cystic fibrosis airway conditions. *Journal of Bacteriology* **190**, 2739–2758 (2008); published online EpubApr (10.1128/JB.01683-07).

188 M. Wu, T. Guina, M. Brittnacher, H. Nguyen, J. Eng, S. I. Miller, The Pseudomonas aeruginosa proteome during anaerobic growth. *Journal of Bacteriology* **187**, 8185–8190 (2005); published online EpubDec (10.1128/JB.187.23.8185-8190.2005).

189 A. J. Park, K. Murphy, J. R. Krieger, D. Brewer, P. Taylor, M. Habash, C. M. Khursigara, A temporal examination of the planktonic and biofilm proteome of whole cell Pseudomonas aeruginosa PAO1 using quantitative mass spectrometry. *Molecular & Cellular Proteomics: MCP* **13**, 1095–1105 (2014); published online EpubApr (10.1074/mcp.M113.033985).

190 D. Wehmhoner, S. Haussler, B. Tummler, L. Jansch, F. Bredenbruch, J. Wehland, I. Steinmetz, Inter- and intraclonal diversity of the Pseudomonas aeruginosa proteome manifests within the secretome. *Journal of Bacteriology* **185**, 5807–5814 (2003); published online EpubOct.

191 F. Imperi, F. Ciccosanti, A. B. Perdomo, F. Tiburzi, C. Mancone, T. Alonzi, P. Ascenzi, M. Piacentini, P. Visca, G. M. Fimia, Analysis of the periplasmic proteome of Pseudomonas aeruginosa, a metabolically versatile opportunistic pathogen. *Proteomics* **9**, 1901–1915 (2009); published online EpubApr (10.1002/pmic.200800618).

192 U. Romling, M. Gomelsky, M. Y. Galperin, C-di-GMP: The dawning of a novel bacterial signalling system. *Molecular Microbiology* **57**, 629–639 (2005); published online EpubAug (MMI4697 [pii]10.1111/j.1365-2958.2005.04697.x).

193 R. Hengge, Principles of c-di-GMP signalling in bacteria. *Nature Reviews. Microbiology* **7**, 263–273 (2009); published online EpubApr (10.1038/nrmicro2109).

194 J. W. Hickman, D. F. Tifrea, C. S. Harwood, A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14422–14427 (2005); published online EpubOct 4 (10.1073/pnas.0507170102).

195 H. Kulasakara, V. Lee, A. Brencic, N. Liberati, J. Urbach, S. Miyata, D. G. Lee, A. N. Neely, M. Hyodo, Y. Hayakawa, F. M. Ausubel, S. Lory, Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis-(3′-5′)-cyclic-GMP in virulence. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2839–2844 (2006); published online EpubFeb 21 (0511090103 [pii]10.1073/pnas.0511090103).

196 Y. Li, S. Heine, M. Entian, K. Sauer, N. Frankenberg-Dinkel, NO-induced biofilm dispersion in Pseudomonas aeruginosa is mediated by an MHYT domain-coupled phosphodiesterase. *Journal of Bacteriology* **195**, 3531–3542 (2013); published online EpubAug (10.1128/JB.01156-12).

197 A. B. Roy, O. E. Petrova, K. Sauer, The phosphodiesterase DipA (PA5017) is essential for Pseudomonas aeruginosa biofilm dispersion. *Journal of Bacteriology* **194**, 2904–2915 (2012); published online EpubJun (10.1128/JB.05346-11JB.05346-11 [pii]).

198 M. Starkey, J. H. Hickman, L. Ma, N. Zhang, S. De Long, A. Hinz, S. Palacios, C. Manoil, M. J. Kirisits, T. D. Starner, D. J. Wozniak, C. S. Harwood, M. R. Parsek, Pseudomonas aeruginosa rugose small-colony variants have adaptations that likely

promote persistence in the cystic fibrosis lung. *Journal of Bacteriology* **191**, 3492–3503 (2009); published online EpubJun (10.1128/JB.00119-09).

199 L. Yang, M. Nilsson, M. Gjermansen, M. Givskov, T. Tolker-Nielsen, Pyoverdine and PQS mediated subpopulation interactions involved in Pseudomonas aeruginosa biofilm formation. *Molecular Microbiology* **74**, 1380–1392 (2009); published online EpubDec (10.1111/j.1365-2958.2009.06934.x).

200 Q. Wei, S. Tarighi, A. Dotsch, S. Haussler, M. Musken, V. J. Wright, M. Camara, P. Williams, S. Haenen, B. Boerjan, A. Bogaerts, E. Vierstraete, P. Verleyen, L. Schoofs, R. Willaert, V. N. De Groote, J. Michiels, K. Vercammen, A. Crabbe, P. Cornelis, Phenotypic and genome-wide analysis of an antibiotic-resistant small colony variant (SCV) of Pseudomonas aeruginosa. *PLoS ONE* **6**, e29276 (2011)10.1371/journal.pone.0029276).

201 I. D. Hay, U. Remminghorst, B. H. Rehm, MucR, a novel membrane-associated regulator of alginate biosynthesis in Pseudomonas aeruginosa. *Applied and Environmental Microbiology* **75**, 1110–1120 (2009); published online EpubFeb (10.1128/AEM.02416-08).

202 B. H. Rehm, G. Boheim, J. Tommassen, U. K. Winkler, Overexpression of algE in Escherichia coli: Subcellular localization, purification, and ion channel properties. *Journal of Bacteriology* **176**, 5639–5647 (1994); published online EpubSep.

203 M. Merighi, V. T. Lee, M. Hyodo, Y. Hayakawa, S. Lory, The second messenger bis-(3′-5′)-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in Pseudomonas aeruginosa. *Molecular Microbiology* **65**, 876–895 (2007); published online EpubAug (10.1111/j.1365-2958.2007.05817.x).

204 S. Malhotra, L. A. Silo-Suh, K. Mathee, D. E. Ohman, Proteome analysis of the effect of mucoid conversion on global protein expression in Pseudomonas aeruginosa strain PAO1 shows induction of the disulfide bond isomerase, dsbA. *Journal of Bacteriology* **182**, 6999–7006 (2000); published online EpubDec (10.1128/JB.182.24.6999-7006.2000).

205 D. D. Sriramulu, M. Nimtz, U. Romling, Proteome analysis reveals adaptation of Pseudomonas aeruginosa to the cystic fibrosis lung environment. *Proteomics* **5**, 3712–3721 (2005); published online EpubSep (10.1002/pmic.200401227).

206 S. L. Hanna, M. T. Kinter, J. B. Goldberg, Comparison of proteins expressed by Pseudomonas aeruginosa strains representing initial and chronic isolates from a cystic fibrosis patient: An analysis by 2-D gel electrophoresis and capillary column liquid chromatography-tandem mass spectrometry. *Microbiology* **146**, 2495–2508 (2000).

207 F. H. Damron, D. Qiu, H. D. Yu, The Pseudomonas aeruginosa sensor kinase KinB negatively controls alginate production through AlgW-dependent MucA proteolysis. *Journal of Bacteriology* **191**, 2285–2295 (2009); published online EpubApr (10.1128/JB.01490-08).

208 L. F. Wood, A. J. Leech, D. E. Ohman, Cell wall-inhibitory antibiotics activate the alginate biosynthesis operon in Pseudomonas aeruginosa: Roles of sigma (AlgT) and the AlgW and Prc proteases. *Molecular Microbiology* **62**, 412–426 (2006); published online EpubOct (10.1111/j.1365-2958.2006.05390.x).

209 P. J. Baynham, A. L. Brown, L. L. Hall, D. J. Wozniak, Pseudomonas aeruginosa AlgZ, a ribbon-helix-helix DNA-binding protein, is essential for alginate synthesis and algD transcriptional activation. *Molecular Microbiology* **33**, 1069–1080 (1999); published online EpubSep (10.1046/j.1365-2958.1999.01550.x).

210 P. J. Baynham, D. M. Ramsey, B. V. Gvozdyev, E. M. Cordonnier, D. J. Wozniak, The Pseudomonas aeruginosa ribbon-helix-helix DNA-binding protein AlgZ (AmrZ) controls twitching motility and biogenesis of type IV pili. *Journal of Bacteriology* **188**, 132–140 (2006); published online EpubJan (10.1128/JB.188.1.132-140.2006).

211 J. Rao, A. DiGiandomenico, J. Unger, Y. Bao, R. K. Polanowska-Grabowska, J. B. Goldberg, A novel oxidized low-density lipoprotein-binding protein from Pseudomonas aeruginosa. *Microbiology* **154**, 654–665 (2008); published online EpubFeb (10.1099/mic.0.2007/011429-0).

212 M. J. Coyne, Jr., K. S. Russell, C. L. Coyle, J. B. Goldberg, The Pseudomonas aeruginosa algC gene encodes phosphoglucomutase, required for the synthesis of a complete lipopolysaccharide core. *Journal of Bacteriology* **176**, 3500–3507 (1994); published online EpubJun (10.1128/jb.176.12.3500-3507.1994).

213 N. A. Zielinski, A. M. Chakrabarty, A. Berry, Characterization and regulation of the Pseudomonas aeruginosa algC gene encoding phosphomannomutase. *The Journal of Biological Chemistry* **266**, 9754–9763 (1991); published online EpubMay 25.

214 A. H. Tart, M. C. Wolfgang, D. J. Wozniak, The alternative sigma factor AlgT represses Pseudomonas aeruginosa flagellum biosynthesis by inhibiting expression of fleQ. *Journal of Bacteriology* **187**, 7955–7962 (2005); published online EpubDec (10.1128/JB.187.23.7955-7962.2005).

215 Y. R. Patankar, R. R. Lovewell, M. E. Poynter, J. Jyot, B. I. Kazmierczak, B. Berwin, Flagellar motility is a key determinant of the magnitude of the inflammasome response to Pseudomonas aeruginosa. *Infection and Immunity* **81**, 2043–2052 (2013); published online EpubJun (10.1128/IAI.00054-13).

216 M. Nairz, A. Schroll, T. Sonnweber, G. Weiss, The struggle for iron – a metal at the host-pathogen interface. *Cellular Microbiology* **12**, 1691–1702 (2010); published online EpubDec (10.1111/j.1462-5822.2010.01529.x).

217 B. R. Otto, A. M. Verweij-van Vught, D. M. MacLaren, Transferrins and heme-compounds as iron sources for pathogenic bacteria. *Critical Reviews in Microbiology* **18**, 217–233 (1992)10.3109/10408419209114559).

218 P. Nadal Jimenez, G. Koch, E. Papaioannou, M. Wahjudi, J. Krzeslak, T. Coenye, R. H. Cool, W. J. Quax, Role of PvdQ in Pseudomonas aeruginosa virulence under iron-limiting conditions. *Microbiology* **156**, 49–59 (2010); published online EpubJan (10.1099/mic.0.030973-0).

219 O. Cunrath, V. Gasser, F. Hoegy, C. Reimmann, L. Guillon, I. J. Schalk, A cell biological view of the siderophore pyochelin iron uptake pathway in Pseudomonas aeruginosa. *Environmental Microbiology* (2014); published online EpubJun 20 (10.1111/1462-2920.12544).

220 J. M. Meyer, A. Stintzi, K. Poole, The ferripyoverdine receptor FpvA of Pseudomonas aeruginosa PAO1 recognizes the ferripyoverdines of P. aeruginosa PAO1 and P. fluorescens ATCC 13525. *FEMS Microbiology Letters* **170**, 145–150 (1999); published online EpubJan 1 (10.1016/S0378-1097(98)00537-0).

221 B. Ghysels, B. T. Dieu, S. A. Beatson, J. P. Pirnay, U. A. Ochsner, M. L. Vasil, P. Cornelis, FpvB, an alternative type I ferripyoverdine receptor of Pseudomonas aeruginosa. *Microbiology* **150**, 1671–1680 (2004); published online EpubJun (10.1099/mic.0.27035-0).

222 R. G. Ankenbauer, H. N. Quan, FptA, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a phenolate siderophore receptor homologous to hydroxamate siderophore receptors. *Journal of Bacteriology* **176**, 307–319 (1994); published online EpubJan.

223 B. Marshall, A. Stintzi, C. Gilmour, J. M. Meyer, K. Poole, Citrate-mediated iron uptake in Pseudomonas aeruginosa: Involvement of the citrate-inducible FecA receptor and the FeoB ferrous iron transporter. *Microbiology* **155**, 305–315 (2009); published online EpubJan (10.1099/mic.0.023531-0).

224 U. A. Ochsner, Z. Johnson, M. L. Vasil, Genetics and regulation of two distinct haem-uptake systems, phu and has, in Pseudomonas aeruginosa. *Microbiology* **146 (Pt 1)**, 185–198 (2000); published online EpubJan (10.1099/00221287-146-1-185).

225 J. M. Meyer, A. Neely, A. Stintzi, C. Georges, I. A. Holder, Pyoverdin is essential for virulence of Pseudomonas aeruginosa. *Infection and Immunity* **64**, 518–523 (1996); published online EpubFeb 226.

226 C. D. Cox, Effect of pyochelin on the virulence of Pseudomonas aeruginosa. *Infection and Immunity* **36**, 17–23 (1982); published online EpubApr 227.

227 T. A. Hunt, W. T. Peng, I. Loubens, D. G. Storey, The Pseudomonas aeruginosa alternative sigma factor PvdS controls exotoxin A expression and is expressed in lung infections associated with cystic fibrosis. *Microbiology* **148**, 3183–3193 (2002); published online EpubOct (10.1099/00221287-148-10-3183).

228 P. J. Wilderman, A. I. Vasil, Z. Johnson, M. J. Wilson, H. E. Cunliffe, I. L. Lamont, M. L. Vasil, Characterization of an endoprotease (PrpL) encoded by a PvdS-regulated gene in Pseudomonas aeruginosa. *Infection and Immunity* **69**, 5385–5394 (2001); published online EpubSep (10.1128/IAI.69.9.5385-5394.2001).

229 A. F. Konings, L. W. Martin, K. J. Sharples, L. F. Roddam, R. Latham, D. W. Reid, I. L. Lamont, Pseudomonas aeruginosa uses multiple pathways to acquire iron during chronic infection in cystic fibrosis lungs. *Infection and Immunity* **81**, 2697–2704 (2013); published online EpubAug (10.1128/IAI.00418-13).

230 L. W. Martin, D. W. Reid, K. J. Sharples, I. L. Lamont, Pseudomonas siderophores in the sputum of patients with cystic fibrosis. *Biometals: An International Journal on the Role of Metal Ions in Biology, Biochemistry, and Medicine* **24**, 1059–1067 (2011); published online EpubDec (10.1007/s10534-011-9464-z).

231 A. T. Nguyen, M. J. O'Neill, A. M. Watts, C. L. Robson, I. L. Lamont, A. Wilks, A. G. Oglesby-Sherrouse, Adaptation of Iron Homeostasis Pathways by a Pseudomonas aeruginosa Pyoverdine Mutant in the Cystic Fibrosis Lung. *Journal of Bacteriology* **196**, 2265–2276 (2014); published online EpubJun 15 (10.1128/JB.01491-14).

232 D. De Vos, M. De Chial, C. Cochez, S. Jansen, B. Tummler, J. M. Meyer, P. Cornelis, Study of pyoverdine type and production by Pseudomonas aeruginosa isolated from cystic fibrosis patients: prevalence of type II pyoverdine isolates and accumulation of pyoverdine-negative mutations. *Arch Microbiol* **175**, 384–388 (2001); published online EpubMay.

233 R. L. Marvig, S. Damkiaer, S. M. Khademi, T. M. Markussen, S. Molin, L. Jelsbak, Within-host evolution of Pseudomonas aeruginosa reveals adaptation toward iron acquisition from hemoglobin. *mBio* **5**, e00966–00914 (2014)10.1128/mBio.00966-14).

234 S. Cosgrove, S. H. Chotirmall, C. M. Greene, N. G. McElvaney, Pulmonary proteases in the cystic fibrosis lung induce interleukin 8 expression from bronchial epithelial cells via a heme/meprin/epidermal growth factor receptor/Toll-like receptor pathway.

The Journal of Biological Chemistry **286**, 7692–7704 (2011); published online EpubMar 4 (10.1074/jbc.M110.183863).

235 D. A. Newton, K. M. Rao, R. A. Dluhy, J. E. Baatz, Hemoglobin is expressed by alveolar epithelial cells. *The Journal of Biological Chemistry* **281**, 5668–5676 (2006); published online EpubMar 3 (10.1074/jbc.M509314200).

236 R. C. Hunter, F. Asfour, J. Dingemans, B. L. Osuna, T. Samad, A. Malfroot, P. Cornelis, D. K. Newman, Ferrous iron is a significant component of bioavailable iron in cystic fibrosis airways. *mBio* **4**(4), e00557–13 (2013)10.1128/mBio.00557-13).

237 Y. Wang, J. C. Wilks, T. Danhorn, I. Ramos, L. Croal, D. K. Newman, Phenazine-1-carboxylic acid promotes bacterial biofilm development via ferrous iron acquisition. *Journal of Bacteriology* **193**, 3606–3617 (2011); published online EpubJul (10.1128/JB.00396-11).

238 N. Jiang, N. S. Tan, B. Ho, J. L. Ding, Respiratory protein-generated reactive oxygen species as an antimicrobial strategy. *Nature Immunology* **8**, 1114–1122 (2007); published online EpubOct (10.1038/ni1501).

239 J. R. Govan, V. Deretic, Microbial pathogenesis in cystic fibrosis: Mucoid Pseudomonas aeruginosa and Burkholderia cepacia. *Microbiological Reviews* **60**, 539–574 (1996); published online EpubSep240.

240 J. Zhao, P. D. Schloss, L. M. Kalikin, L. A. Carmody, B. K. Foster, J. F. Petrosino, J. D. Cavalcoli, D. R. VanDevanter, S. Murray, J. Z. Li, V. B. Young, J. J. LiPuma, Decade-long bacterial community dynamics in cystic fibrosis airways. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 5809–5814 (2012); published online EpubApr 10 (10.1073/pnas.1120577109).

241 V. Klepac-Ceraj, K. P. Lemon, T. R. Martin, M. Allgaier, S. W. Kembel, A. A. Knapp, S. Lory, E. L. Brodie, S. V. Lynch, B. J. Bohannan, J. L. Green, B. A. Maurer, R. Kolter, Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. *Environmental Microbiology* **12**, 1293–1303 (2010); published online EpubMay (10.1111/j.1462-2920.2010.02173.x).

242 L. Yang, Y. Liu, T. Markussen, N. Hoiby, T. Tolker-Nielsen, S. Molin, Pattern differentiation in co-culture biofilms formed by Staphylococcus aureus and Pseudomonas aeruginosa. *FEMS Immunology and Medical Microbiology* **62**, 339–347 (2011); published online EpubAug (10.1111/j.1574-695X.2011.00820.x).

243 A. R. Erickson, B. L. Cantarel, R. Lamendella, Y. Darzi, E. F. Mongodin, C. Pan, M. Shah, J. Halfvarson, C. Tysk, B. Henrissat, J. Raes, N. C. Verberkmoes, C. M. Fraser, R. L. Hettich, J. K. Jansson, Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS ONE* **7**, e49138 (2012)10.1371/journal.pone.0049138).

244 N. Gupta, J. Benhamida, V. Bhargava, D. Goodman, E. Kain, I. Kerman, N. Nguyen, N. Ollikainen, J. Rodriguez, J. Wang, M. S. Lipton, M. Romine, V. Bafna, R. D. Smith, P. A. Pevzner, Comparative proteogenomics: Combining mass spectrometry and comparative genomics to analyze multiple genomes. *Genome Research* **18**, 1133–1142 (2008); published online EpubJul (10.1101/gr.074344.107).

245 D. Willner, M. R. Haynes, M. Furlan, R. Schmieder, Y. W. Lim, P. B. Rainey, F. Rohwer, D. Conrad, Spatial distribution of microbial communities in the cystic fibrosis lung. *The ISME Journal* **6**, 471–474 (2012); published online EpubFeb (10.1038/ismej.2011.104).

246 H. Matsui, V. E. Wagner, D. B. Hill, U. E. Schwab, T. D. Rogers, B. Button, R. M. Taylor, 2nd, R. Superfine, M. Rubinstein, B. H. Iglewski, R. C. Boucher, A physical linkage between cystic fibrosis airway surface dehydration and Pseudomonas aeruginosa biofilms. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18131–18136 (2006); published online EpubNov 28 (10.1073/pnas.0606428103).

247 B. K. Rubin, Mucus, phlegm, and sputum in cystic fibrosis. *Respiratory Care* **54**, 726–732; discussion 732 (2009); published online EpubJun.

248 B. Ahmed, A. Bush, J. C. Davies, How to use: bacterial cultures in diagnosing lower respiratory tract infections in cystic fibrosis. *Archives of Disease in childhood. Education and Practice Edition* **99**, 181–187 (2014); published online EpubOct (10.1136/archdischild-2012-303408).

249 F. Armougom, F. Bittar, N. Stremler, J. M. Rolain, C. Robert, J. C. Dubus, J. Sarles, D. Raoult, B. La Scola, Microbial diversity in the sputum of a cystic fibrosis patient studied with 16S rDNA pyrosequencing. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology* **28**, 1151–1154 (2009); published online EpubSep (10.1007/s10096-009-0749-x).

250 M. J. Cox, M. Allgaier, B. Taylor, M. S. Baek, Y. J. Huang, R. A. Daly, U. Karaoz, G. L. Andersen, R. Brown, K. E. Fujimura, B. Wu, D. Tran, J. Koff, M. E. Kleinhenz, D. Nielson, E. L. Brodie, S. V. Lynch, Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS ONE* **5**, e11044 (2010)10.1371/journal.pone.0011044).

251 M. J. Hoegger, A. J. Fischer, J. D. McMenimen, L. S. Ostedgaard, A. J. Tucker, M. A. Awadalla, T. O. Moninger, A. S. Michalski, E. A. Hoffman, J. Zabner, D. A. Stoltz, M. J. Welsh, Cystic fibrosis. Impaired mucus detachment disrupts mucociliary transport in a piglet model of cystic fibrosis. *Science* **345**, 818–822 (2014); published online EpubAug 15 (10.1126/science.1255825).

252 D. Collie, J. Govan, S. Wright, E. Thornton, P. Tennant, S. Smith, C. Doherty, G. McLachlan, A lung segmental model of chronic Pseudomonas infection in sheep. *PLoS ONE* **8**, e67677 (2013)10.1371/journal.pone.0067677).

18

Top-Down Proteomics in the Study of Microbial Pathogenicity

Joseph Gault, Egor Vorontsov, Mathieu Dupré, Valeria Calvaresi, Magalie Duchateau, Diogo B. Lima, Christian Malosse and Julia Chamot-Rooke

Structural Mass Spectrometry and Proteomics Unit, Institut Pasteur, Paris, France

18.1 Introduction

Top-down proteomics is an emerging technology based on the analysis of intact proteins using high-resolution mass spectrometry.[1,2] This is in contrast to the better-established bottom-up strategy, which relies on the analysis of peptide fragments obtained after enzymatic digestion.

A great advantage of top-down proteomics is its ability to address and characterize protein variations such as alternative splicing, allelic variation, or posttranslational modification (PTM) that are essential for proper protein function, but are not directly encoded in the genome.[3] By examining intact proteins directly, top-down proteomics provides a bird's eye view of all protein species (proteoforms) and thus reveals information that may be closely connected to complex disease phenotypes.[4] By studying intact proteins rather than peptides, top-down proteomics also neatly sidesteps the so-called "inference problem" that is inherent in the bottom-up proteomics workflow (the same peptide can be shared by several proteins, which can lead to ambiguous identification); (see comparison, Figure 18.1).

Two distinct analysis strategies exist in top-down proteomics (Figure 18.2). Choosing which is the most appropriate generally depends on the sample complexity.

The first strategy – *targeted mode (or offline)* – addresses the analysis of purified proteins or simple mixtures of several proteins and/or proteoforms. Proteins are introduced into the mass spectrometer by direct infusion nanoelectrospray (nanoESI). The use of the chip-based infusion technology developed by Advion (Triversa Nanomate)[5] is a very good option to obtain an nESI spray that is both readily reproducible and allows the measurement of samples for the long scan times required to obtain high-quality data. An MS profiling experiment is first performed. This MS-level characterization provides an accurate mass for all proteoforms present in the sample. Following

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

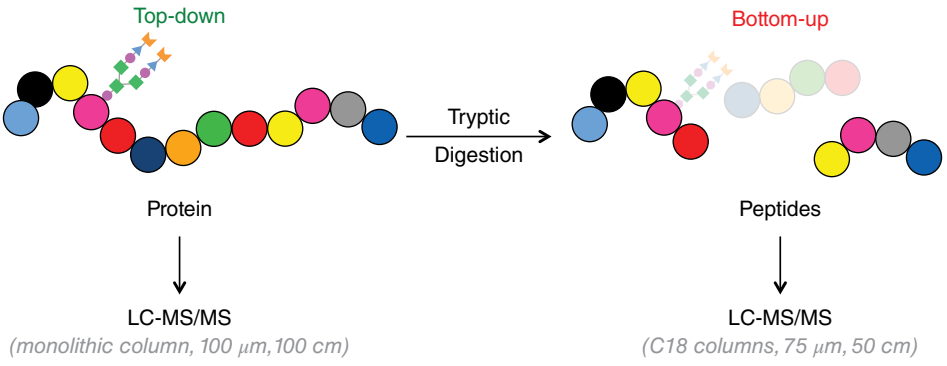


Figure 18.1 Bottom-up versus top-down proteomics.

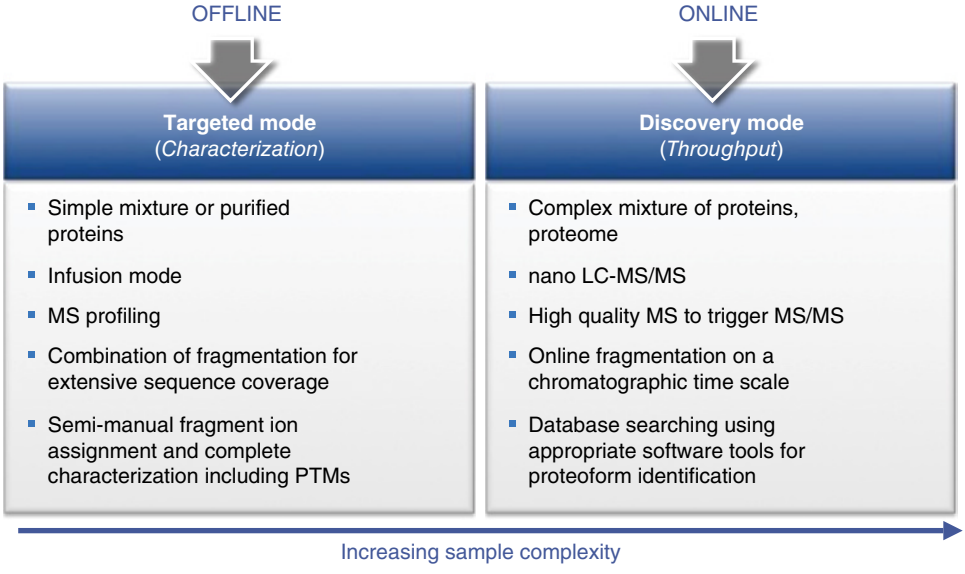


Figure 18.2 Targeted versus discovery mode in top-down proteomics.

this measurement, the different proteoforms are individually selected inside the mass spectrometer and fragmented to reveal further structural information. In most cases, a combination of several complementary fragmentation techniques (collision activated dissociation or CAD, electron transfer dissociation or ETD, electron capture dissociation or ECD, etc.) can be employed to maximize sequence coverage of the target proteoform.[6] CAD (as well as HCD, high collision energy dissociation) leads to the formation of *b/y* ions, whereas ETD and ECD give rise to *c/z* fragments. A high sequence coverage is required to confidently localize potential modifications. This is particularly useful when there are multiple combinations of modifications at different sites on the protein

backbone that give rise to several proteoforms of the same mass. For data analysis, various software packages exist for matching spectral ions to a sequence and pinpointing known PTMs. There is to date no software capable of automatically assigning unknown PTMs in protein sequences associated with a score, and in most of these cases a semi-manual fragment ion assignment must be performed. Unfortunately, this can be very time consuming.

The second acquisition strategy is *discovery* mode (or *on-line*). In this case, the objective is to perform a large-scale analysis of many proteins present in a "complex" sample, ideally a whole proteome, using an LC-MS/MS strategy. This mode, which mimics bottom-up proteomics, requires on-line chromatographic separation of proteins prior to MS/MS and their subsequent identification by database searching. A possible way to increase the dynamic range in top-down proteomics is to include one or several prefractionation steps, using SCX or SAX (strong cation/anion exchange chromatography), or iso-electric focusing.[7] The on-line separation of intact proteins in conditions compatible with mass spectrometry (neither salt nor detergent) is not an easy task.[8] The usual C18 capillary columns used for the separation of peptides in shotgun proteomics are not suitable for the separation of intact proteins, and other types of separation are required. Among the different materials that have been tested, the most suitable appears to be the monolithic columns that emerged at the end of the 1980s. Monolithic columns are formed of a single piece of porous solid, rather than the particulate matter that is usually used in most reverse phase columns. Monoliths possess the advantages of both polymeric phases (low pressure drop and ease of handling) and conventional columns (high efficiency and resolution).[9] Because of the requirement for good chromatographic separation at reasonable gradient times (a few hours at the most), peak widths, and thus the time available for proteoform fragmentation, are usually less than 1 min. One cannot therefore expect sequence coverage as high as in targeted mode, where acquisitions may last much longer. However, this on-line approach allows for the analysis of much more complex samples, up to entire proteomes, and is complementary to the bottom-up approach as it provides protein-level information on proteoforms that would otherwise be lost in classical peptide-centric methods. Just as in bottom-up proteomics, specialized methods can be tailored to optimize fragmentation and thus sequence coverage of specific species if required.

An important step in the process is also data analysis, which requires the use of appropriate software allowing database searching to identify both proteins and proteoforms. To date, the most advanced software packages include ProSight[10], MS-Align + [11], MASH,[12] and pTop.[13] These software packages have become increasingly capable of processing the complex spectra data generated from fragmentation of entire protein species and matching spectral peaks' theoretical ions to achieve protein identification. However, there is still much room for improvement in the bioinformatics analysis of top-down proteomics data, and robust statistics are still required to attach appropriate confidence metrics or "scores" to identified proteoforms and for their quantification (p-Score, c-Score).[14] This is particularly important when the sample is very complex or data are sparse, and multiple assignments are possible. In such cases, manual validation of individual spectra often remains necessary.

In the field of microbiology, several papers have already demonstrated the strength of top-down approaches to study proteins involved in bacterial virulence. A summary of these studies, as well as future trends, is presented in this chapter.

18.2 Top-Down Analysis of Modified Bacterial Proteins in Targeted Mode

PTM in the bacterial world is becoming recognized as an important, yet still poorly understood method by which pathogens mediate interactions with their host. Because bacterial proteins present an average mass around 30 kDa and are available in large quantities, they represent ideal targets for top-down proteomics. Several studies have focused on the analysis of pilin proteins that are the major components of type IV pili. Type IV pili (Figure 18.3) are extracellular filamentous virulence factors frequently expressed by bacterial pathogens.

They play roles in a myriad of life processes such as DNA uptake, motility, host cell adhesion, and bacterial aggregation. The ability of pili from different bacteria to aggregate or bundle helps colonies stay together and resist external pressures such as shear stress.[15] The initial top-down study of the major pilin PilE, purified from *Neisseria meningitidis* NM8013 reference strain, led to the characterization of a previously unreported sugar called GATDH (glyceramidoacetamidotrideoxyhexose) that was found to be present in half of clinical isolates.[16] It was then demonstrated, using a combination of bottom-up and top-down approaches, that bacterial dissemination, a key step in virulence, is triggered by the addition of a single phosphoglycerol group (PG) on PilE, a PTM regulated upon cell contact. This PG group added to Ser^{93} introduces a negative charge in a patch of positively charged amino acids on the pilus surface and completely destabilizes the pilus–pilus interaction. This destabilization favors the detachment of single bacteria from the colony, their dissemination within the host and, crucially, their migration through the epithelial layer. These results shed new light on *N. meningitidis*

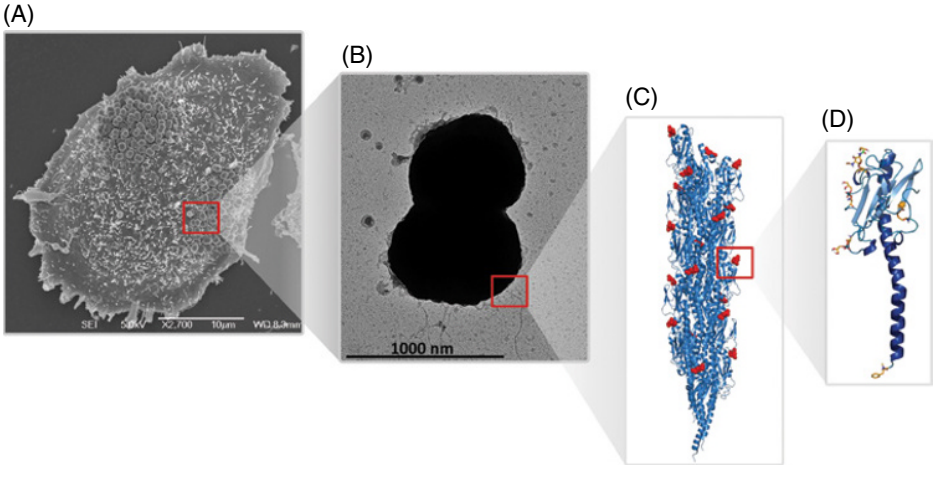


Figure 18.3 (A) *Neisseria meningitidis* colonies on an epithelial cell, (B) *N. meningitidis* bacterium and its type IV pili, (C) modeling of a type IV pilus, (D) PilE protein: the major component of type IV pili.*

* The authors thank Guillaume Dumenil (Unit of Pathogenesis of Vascular Infections at the Institut Pasteur), Gérard Péhau-Arnaudet (CNRS UMR3528 and Institut Pasteur Ultrapole) for providing the EM images of Figure 18.3 and M. Nilges (Structural Bioinformatics Unit, Institut Pasteur) for modeling of the fiber (Figure 18.3).

pathogenicity and show the added value that top-down approaches can bring to the analysis of PTMs.[17]

Following this study, all PTMs carried by the two major proteoforms of PilE from reference strain NM8013 were comprehensively mapped for the first time. The methodology used was a combination of mass profiling (intact mass measurement of the proteoforms) and tandem mass spectrometry on intact proteoforms (targeted mode top-down analysis). The modifications found include a processed and methylated N-terminus, disulfide bridge, glycosylation, and glycerophosphorylation at two different sites (Figure 18.4).[18]

In order to shed new light on the enigmatic link between PTMs and virulence, the analysis of PilE was recently extended to previously uncharacterized hypervirulent *N. meningitidis* strains collected in a hospital from patients with evidence of meningitis. *N. meningitidis* strains can be classified into two groups. The strains expressing class I pilins contain a genetic recombination system that promotes variation of the pilin sequence and is thought to aid immune escape.[19] However, numerous hypervirulent clinical isolates (such as the ones collected in this study) lack this property and express pilins belonging to a second class (class II) with an invariable primary structure. This raised the question of how strains expressing class II pilins evade immunity targeting type IV pili? Although class I pilins carry a single glycan, it was unexpectedly found, using dedicated top-down approaches, that class II pilins display up to five glycosylation sites. The combination of these results with molecular modeling led to a new model where strains expressing class II pilins evade the immune system by changing the chemical composition of the sugar rather than the pilin primary structure. These results highlight the fact that bacterial glycans can have extensive functional and immunological

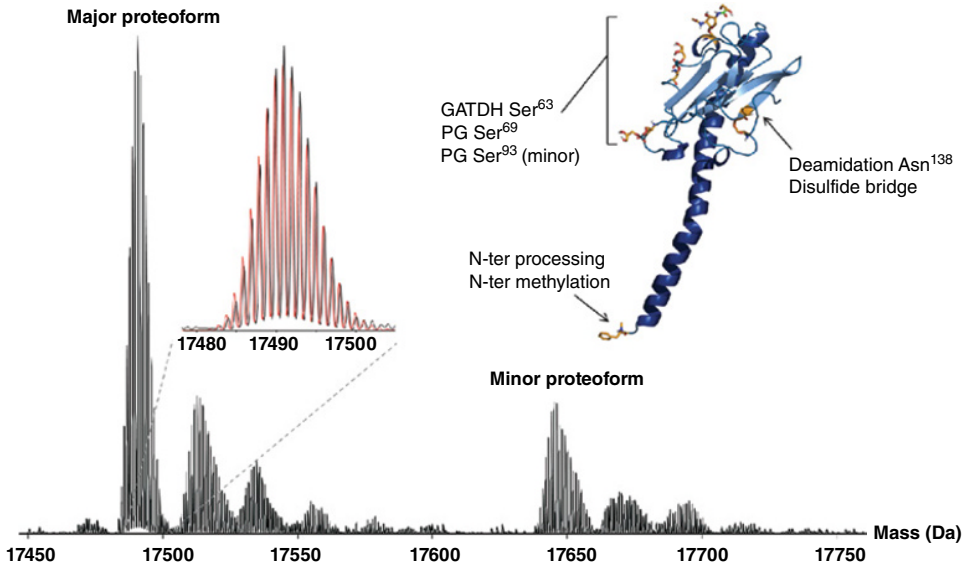


Figure 18.4 High-resolution FT-ICR mass profile of PilE alongside a model showing all PTMs (orange) on the protein structure. Once all PTMs are taken into account, the experimental spectrum (black) correlates very well with a theoretically generated isotope pattern (red).[18]

consequences and that top-down approaches are particularly suitable when more than two proteoforms are present in a single sample.[20,21]

A similar top-down study was performed on ComGC, which is the major pilin of the first type IV pilus ever described for a gram-positive bacterium (in *Streptococcus pneumonia*). In that case, the top-down analysis only revealed a processed and methylated N-terminus and no further PTM.[22,23]

All these results have been obtained on purified proteins, using direct infusion of samples, exemplifying the fact that a targeted approach is suitable for purified proteins or very simple mixtures but not for large-scale analysis. To improve throughput and move from the analysis of intact proteins to intact proteomes, on-line LC-MS/MS is required.

18.3 Top-Down Analysis of Bacterial Proteins in Discovery Mode

The first study describing the large-scale analysis of bacterial proteins using LC-MS dates back to 2008 and concerns the analysis of the soluble fraction of an *Escherichia coli* lysate.[24] Strong anion exchange (SAX) chromatography was used to fractionate the sample into 36 fractions that were further analyzed by LC-MS/MS with ETD. The experiments were undertaken on a linear ion trap mass spectrometer using a C4 separation. Using a dedicated data processing pipeline, 174 proteins, corresponding to 322 proteoforms, could be identified with an FDR < 1%. Among these proteins, 43 of the 53 ribosomal subunits were identified, representing 141 proteoforms. The majority of the species observed had some degree of processing when compared to the protein expected from the genome. N-terminal methionine excision, which is a common PTM for prokaryotes, was observed in 47 proteins. Thiolation as well as signal peptide removal were also observed for several proteins. In this study, 20 µg of protein was injected into the mass spectrometer, and data collection was completed in two days.

Similar numbers (154 proteins, 201 proteoforms) were obtained in 2009 by Tsai *et al.*[23] for the analysis of *Salmonella typhimurium* outer membrane extract. In that case, an LTQ-Orbitrap instrument was used with a C4 chromatographic separation. Data were obtained using CAD on the most abundant precursor ions selected over seven different ranges in the ion trap (gas phase fractionation) with detection of fragments in the Orbitrap. In this paper, the authors developed a precursor ion independent algorithm (called PIITA) to process their data. Compared to a bottom-up analysis performed on the same sample, 73 proteins were identified exclusively using the top-down approach. As with the top-down characterization of the *E. coli* soluble proteome, the most frequently found PTM was methionine excision (for 71 proteins). Acetylation, methylation, and thiolation were also observed and confirmed previous findings.

More recently an LTQ-Orbitrap was also used for the study of the periplasmic *Novosphingobium aromaticivorans* proteome.[25] These gram-negative bacteria are known for their ability to degrade aromatic hydrocarbons. The genome of *Novosphingobium aromaticivorans* comprises 3917 proteins, with 30% annotated as "hypothetical." The single-dimension LC-MS/MS analysis of the sample, using a (80 cm × 75 µm) column packed with C5 particles led to the confident identification of 55 proteins, compared

with 87 proteins by a bottom-up approach. Most of the proteins were found to be modified, mostly by N-terminal methionine excision or signal peptide removal. Some of these modifications were identified exclusively by top-down analysis, as the corresponding peptides were not retrieved in the bottom-up analysis.

An interesting paper on the analysis of *Salmonella typhimurium* proteome was published by the same group shortly after, using the same setup (single-dimension LC-MS/MS on LTQ-Orbitrap Velos).[26] In that case, 563 proteins, corresponding to 1655 proteoforms, were identified (the largest microbial top-down dataset reported so far). Of particular interest, the authors reported the differential utilization of protein S-thiolation forms in *Salmonella* in response to growth under infection-like conditions. Under infection-like conditions, *Salmonella* preferentially uses S-cysteinylation as a mechanism of thiol protection and/or environmental sensing, whereas under basal conditions the pathogen preferentially uses S-gluthationylation. These data are corroborated both by bottom-up proteomics analysis and transcriptomics data. Most of the proteins were found expressed as one or two proteoforms, but interestingly one protein was present as more than 50 different proteoforms.

These studies demonstrate that top-down proteomics has recently achieved an important milestone and can now be employed to characterize bacterial proteins or proteomes in a large-scale manner. Therefore, one may think of using this approach in clinical microbiology to characterize bacterial pathogens.

18.4 Top-Down Proteomics: The Next Step in Clinical Microbiology?

The true pioneer of the use of mass spectrometry for bacterial identification is undoubtedly Catherine Fenselau. In 1975, long before the first published description of MALDI-TOF MS, she combined pyrolysis and mass spectrometry to characterize pathogenic gram-negative bacteria based on the analysis of small molecules such as phospholipids or ubiquinones.[27] The soft ionization techniques developed in the late 1980s then made it possible to analyze large biomolecules such as intact proteins. In 1996, two different papers demonstrated that it was possible to obtain MALDI-TOF spectral profiles from whole bacterial cells.[28,29] Since then, multiple studies have reported the utility of MALDI-TOF MS for the routine identification of microorganisms in clinical microbiology laboratories. Many hospitals worldwide are equipped with this technology, which has been broadly adopted for routine diagnostics.[30] In MALDI-TOF MS, a cultured colony is directly spotted onto the MALDI plate and overlaid with the matrix solution (alpha-cyano-4-hydroxycinnamic acid).[31] The plate is air-dried and inserted into the mass spectrometer for automated measurement. The desorption/ionization of the most abundant proteins (mostly ribosomal proteins) from the colony leads to a spectral profile (MS profile), which is recorded in the 2,000–20,000 Da range. The commercial systems are equipped with time-of-flight analyzers having a resolution that is deliberately reduced to allow the measurement of a *unique average molecular mass* per protein, even for proteins with very low molecular weights. The spectral profile is then compared to a library of spectra previously recorded under the same conditions for known microorganisms. A score ranging between 0 and 3 is

attributed to the pathogen identification and reflects the match between the experimental profile and reference spectra.

It is worth noting that the pathogen identification is done without any information on the identity of the proteins used to create the spectral profile. Overall, confident identification (at the genus or species level) is obtained for only 90% of the samples, which means that 10% remain highly problematic and require additional experiments. This important limitation is due either to the absence of reference spectra in the database or an inability to identify very closely related species with very similar MS profiles (a typical example includes the failure to discriminate between *E. coli* and *Shigella*). Another important limitation of MALDI-TOF MS is its poor discriminatory power, which is insufficient for reliably differentiating subspecies within species, or clones within subspecies, which would be very useful for early identification of epidemics. More importantly, virulence or resistance determinants cannot be characterized with MALDI-TOF MS, which is a severe obstacle for appropriate patient care and prescription of antibiotics in hospitals. Thus, new approaches allowing more precise identification of bacterial strains are strongly needed, if possible with sample analysis performed very rapidly after sample collection.[32] Currently, genomic methods (either targeted or large scale) are widely used for this purpose but have substantial drawbacks. Most significant is the intrinsic, and thus intractable problem, that some proteins may not actually be expressed although the corresponding gene is present. Therefore, proteomic approaches, that directly target expressed proteins (gene products rather than genes) are closer to phenotypes and can provide more relevant information. A few studies have demonstrated that bottom-up approaches can make a contribution to achieving this goal.[33,34] However, similar to MALDI-TOF MS, the recent development of robust top-down approaches offers a neat alternative targeting strategy, enabling access to a unique "proteoform fingerprint" that could provide a signature of a particular genus, species, or phenotype.

To illustrate such an experiment, a top-down LC-MS/MS analysis of intact proteins extracted from total cell lysates of *E. coli* and *Shigella sonnei* performed on an Orbitrap Fusion mass spectrometer is shown in Figure 18.5. It is important to note that these two species are indistinguishable using MALDI-TOF MS.

In both cases, more than 200 proteins and 500 proteoforms could be identified. An example of the excellent sequence coverage typically obtained for a ribosomal protein is shown in Figure 18.6. Processing of the initiation methionine and acetylation of N-terminal serine could easily be evidenced.

In both lysates, specific proteins (marked with an asterisk in Figure 18.4) with unique masses and retention times could be identified. These particular "protein markers" in the proteome profiles can serve to unambiguously identify each bacterial species. This brief example shows that top-down proteomics is capable of going beyond MALDI-TOF for the accurate characterization of bacterial pathogens.

Although this and other examples[35,36] show the added value of top-down proteomics in the field of clinical microbiology, every aspect of the technology has not yet matured sufficiently to the point that it can be routinely used for the deep characterization of clinical samples. To be largely embraced by a wider community of microbiologists and clinicians, the method will have to prove as cheap and simple to use as MALDI-TOF MS, which will be a major, but achievable challenge in the near future.

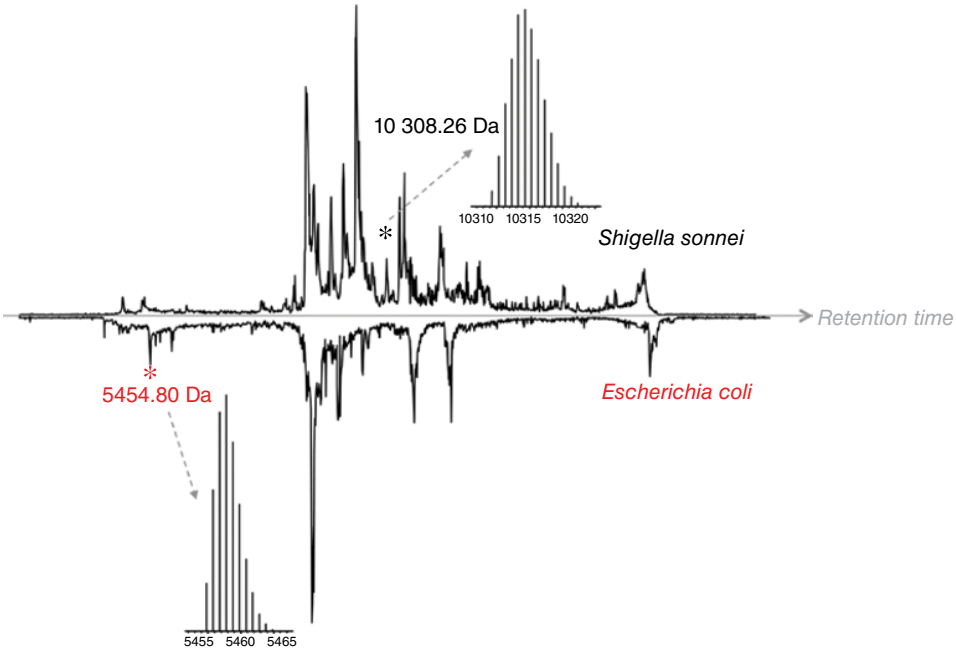


Figure 18.5 Comparison of the total ion chromatograms obtained for the LC-MS analysis of intact proteins extracted from *S. sonnei* and *E. coli* lysates.

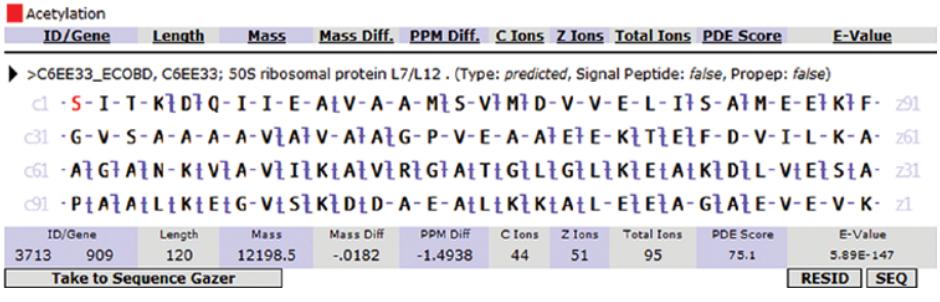


Figure 18.6 Sequence coverage (*c*/*z* ions) obtained for the 50S L7/L12 ribosomal protein of mass 12,198.5 Da (using ETD fragmentation).

References

1 Yates, J. R., 3rd, and Kelleher, N. L. Top down proteomics. *Anal Chem*, 2013, 85: 6151.

2 Catherman, A. D., Skinner, O. S., and Kelleher, N. L. Top down proteomics: Facts and perspectives. *Biochem Biophys Res Commun*, 2014, 445: 683–693.

3 Smith, L. M., and Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat Methods*, 2013, 10: 186–187.

4 Savaryn, J. P., Catherman, A. D., Thomas, P. M., Abecassis, M. M., and Kelleher, N. L. The emergence of top-down proteomics in clinical research. *Genome Med*, 2013, 5: 53.

5 Van Pelt, C. K., Zhang, S., and Henion, J. D. Characterization of a fully automated nanoelectrospray system with mass spectrometric detection for proteomic analyses. *J Biomol Tech*, 2002, 13: 72–84.

6 Brunner, A. M., Lossl, P., Liu, F., Huguet, R., Mullen, C., Yamashita, M., Zabrouskov, V., Makarov, A., Altelaar, A. F., and Heck, A. J. Benchmarking multiple fragmentation methods on an Orbitrap Fusion for top-down phospho-proteoform characterization. *Anal Chem*, 2015, 87: 4152–4158.

7 Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E., Catherman, A. D., Durbin, K. R., Tipton, J. D., Vellaichamy, A., Kellie, J. F., Li, M., Wu, C., Sweet, S. M., Early, B. P., Siuti, N., LeDuc, R. D., Compton, P. D., Thomas, P. M., and Kelleher, N. L. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*, 2011, 480: 254–258.

8 Doucette, A. A., Tran, J. C., Wall, M. J., and Fitzsimmons, S. Intact proteome fractionation strategies compatible with mass spectrometry. *Expert Rev Proteomics*, 2011, 8: 787–800.

9 Capriotti, A. L., Cavaliere, C., Foglia, P., Samperi, R., and Lagana, A. Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. *J Chromatogr A*, 2011, 1218: 8760–8776.

10 Zamdborg, L., LeDuc, R. D., Glowacz, K. J., Kim, Y. B., Viswanathan, V., Spaulding, I. T., Early, B. P., Bluhm, E. J., Babai, S., and Kelleher, N. L. ProSight PTM 2.0: Improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res*, 2007, 35: W701–6.

11 Liu, X., Sirotkin, Y., Shen, Y., Anderson, G., Tsai, Y. S., Ting, Y. S., Goodlett, D. R., Smith, R. D., Bafna, V., and Pevzner, P. A. Protein identification using top-down. *Mol Cell Proteomics*, 2012, 11: M111 008524.

12 Cai, W. X., Guner, H., Gregorich, Z. R., Chen, A. J., Ayaz-Guner, S., Peng, Y., Valeja, S. G., Liu, X. W., and Ge, Y. MASH Suite Pro: A comprehensive software tool for top-down proteomics. *Mol Cell Proteomics*, 2016, 15: 703–714.

13 Sun, R. X., Luo, L., Wu, L., Wang, R. M., Zeng, W. F., Chi, H., Liu, C., and He, S. M. pTop 1.0: A high-accuracy and high-efficiency search engine for intact protein identification. *Anal Chem*, 2016, 88: 3082–3090.

14 LeDuc, R. D., Fellers, R. T., Early, B. P., Greer, J. B., Thomas, P. M., and Kelleher, N. L. The C-Score: A Bayesian framework to sharply improve proteoform scoring in high-throughput top down proteomics. *J Proteome Res*, 2014, 13: 3231–3240.

15 Craig, L., Pique, M. E., and Tainer, J. A. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol*, 2004, 2: 363–378.

16 Chamot-Rooke, J., Rousseau, B., Lanternier, F., Mikaty, G., Mairey, E., Malosse, C., Bouchoux, G., Pelicic, V., Camoin, L., Nassif, X., and Dumenil, G. Alternative Neisseria spp. type IV pilin glycosylation with a glyceramido acetamido trideoxyhexose residue. *Proc Natl Acad Sci U S A*, 2007, 104: 14783–14788.

17 Chamot-Rooke, J., Mikaty, G., Malosse, C., Soyer, M., Dumont, A., Gault, J., Imhaus, A. F., Martin, P., Trellet, M., Clary, G., Chafey, P., Camoin, L., Nilges, M., Nassif, X., and Dumenil, G. Posttranslational modification of pili upon cell contact triggers N. meningitidis dissemination. *Science*, 2011, 331: 778–782.

18 Gault, J., Malosse, C., Dumenil, G., and Chamot-Rooke, J. A combined mass spectrometry strategy for complete posttranslational modification mapping of Neisseria meningitidis major pilin. *J Mass Spectrom*, 2013, 48: 1199–1206.

19 Wormann, M. E., Horien, C. L., Bennett, J. S., Jolley, K. A., Maiden, M. C., Tang, C. M., Aho, E. L., and Exley, R. M. Sequence, distribution and chromosomal context of class I and class II pilin genes of *Neisseria meningitidis* identified in whole genome sequences. *BMC Genomics*, 2014, 15: 253.

20 Gault, J., Ferber, M., Machata, S., Imhaus, A. F., Malosse, C., Charles-Orszag, A., Millien, C., Bouvier, G., Bardiaux, B., Pehau-Arnaudet, G., Klinge, K., Podglajen, I., Ploy, M. C., Seifert, H. S., Nilges, M., Chamot-Rooke, J., and Dumenil, G. Neisseria meningitidis Type IV Pili composed of sequence invariable pilins are masked by multisite glycosylation. *PLoS Pathog*, 2015, 11: e1005162.

21 Gault, J., Malosse, C., Machata, S., Millien, C., Podglajen, I., Ploy, M. C., Costello, C. E., Dumenil, G., and Chamot-Rooke, J. Complete posttranslational modification mapping of pathogenic *Neisseria meningitidis* pilins requires top-down mass spectrometry. *Proteomics*, 2014, 14: 1141–1151.

22 Laurenceau, R., Pehau-Arnaudet, G., Baconnais, S., Gault, J., Malosse, C., Dujeancourt, A., Campo, N., Chamot-Rooke, J., Le Cam, E., Claverys, J. P., and Fronzes, R. A type IV pilus mediates DNA binding during natural transformation in *Streptococcus pneumoniae*. *PLoS Pathog*, 2013, 9: e1003473.

23 Tsai, Y. S., Scherl, A., Shaw, J. L., MacKay, C. L., Shaffer, S. A., Langridge-Smith, P. R., and Goodlett, D. R. Precursor ion independent algorithm for top-down shotgun proteomics. *J Am Soc Mass Spectrom*, 2009, 20: 2154–2166.

24 Bunger, M. K., Cargile, B. J., Ngunjiri, A., Bundy, J. L., and Stephenson, J. L., Jr. Automated proteomics of *E. coli* via top-down electron-transfer dissociation mass spectrometry. *Anal Chem*, 2008, 80: 1459–1467.

25 Wu, S., Brown, R. N., Payne, S. H., Meng, D., Zhao, R., Tolic, N., Cao, L., Shukla, A., Monroe, M. E., Moore, R. J., Lipton, M. S., and Pasa-Tolic, L. Top-down characterization of the post-translationally modified intact periplasmic proteome from the bacterium Novosphingobium aromaticivorans. *Int J Proteomics*, 2013, 2013: 279590–279590.

26 Ansong, C., Wu, S., Meng, D., Liu, X., Brewer, H. M., Kaiser, B. L. D., Nakayasu, E. S., Cort, J. R., Pevzner, P., Smith, R. D., Heffron, F., Adkins, J. N., and Pasa-Tolic, L. Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella typhimurium* in response to infection-like conditions. *Proc. Natl Acad Sci U S A*, 2013, 110: 10153–10158.

27 Anhalt, J. P., and Fenselau, C. Identification of bacteria using mass-spectrometry. *Anal Chem*, 1975, 47: 219–225.

28 Claydon, M. A., Davey, S. N., Edwards-Jones, V., and Gordon, D. B. The rapid identification of intact microorganisms using mass spectrometry. *Nat Biotechnol*, 1996, 14: 1584–1586.

29 Holland, R. D., Wilkes, J. G., Rafii, F., Sutherland, J. B., Persons, C. C., Voorhees, K. J., and Lay, J. O., Jr. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*, 1996, 10: 1227–1232.

30 Nassif, X. A revolution in the identification of pathogens in clinical laboratories. *Clin Infect Dis*, 2009, 49: 552–553.

31 Mather, C. A., Rivera, S. F., and Butler-Wu, S. M. Comparison of the Bruker Biotyper and Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry systems for identification of mycobacteria using simplified protein extraction protocols. *J Clin Microbiol*, 2014, 52: 130–138.

32 Fournier, P. E., Drancourt, M., Colson, P., Rolain, J. M., La Scola, B., and Raoult, D. Modern clinical microbiology: New challenges and solutions. *Nat Rev Microbiol*, 2013, 11: 574–585.

33 Charretier, Y., Dauwalder, O., Franceschi, C., Degout-Charmette, E., Zambardi, G., Cecchini, T., Bardet, C., Lacoux, X., Dufour, P., Veron, L., Rostaing, H., Lanet, V., Fortin, T., Beaulieu, C., Perrot, N., Dechaume, D., Pons, S., Girard, V., Salvador, A., Durand, G., Mallard, F., Theretz, A., Broyer, P., Chatellier, S., Gervasi, G., Van Nuenen, M., Roitsch, C. A., Van Belkum, A., Lemoine, J., Vandenesch, F., and Charrier, J. P. Rapid bacterial identification, resistance, virulence and type profiling using selected reaction monitoring mass spectrometry. *Sci Rep*, 2015, 5.

34 Karlsson, R., Davidson, M., Svensson-Stadler, L., Karlsson, A., Olesen, K., Carlsohn, E., and Moore, E. R. Strain-level typing and identification of bacteria using mass spectrometry-based proteomics. *J Proteome Res*, 2012, 11: 2710–2720.

35 Wynne, C., Fenselau, C., Demirev, P. A., and Edwards, N. Top-down identification of protein biomarkers in bacteria with unsequenced genomes. *Anal Chem*, 2009, 81: 9633–9642.

36 Wynne, C., Edwards, N. J., and Fenselau, C. Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC-MS/MS on an Orbitrap. *Proteomics*, 2010, 10: 3631–3643.

19

Tandem Mass Spectrometry in Resolving Complex Gut Microbiota Functions

Carolin Kolmeder,[1] Kaarina Lähteenmäki,[2] Pirjo Wacklin,[3] Annika Kotovuori,[4] Ilja Ritamo,[4] Jaana Mättö,[2] Willem M. de Vos[1,5] and Leena Valmu[4]

[1] *University of Helsinki, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland*
[2] *Finnish Red Cross Blood Service, Helsinki, Finland*
[3] *Salwe Ltd, Espoo, Finland*
[4] *ThermoFisher Scientific, Vantaa, Finland*
[5] *Laboratory of Microbiology Wageningen University, Wageningen, The Netherlands*

19.1 Introduction

19.1.1 Scope

We are all colonized by a massive microbiome, a complex set of microbes that interact actively with our body, constantly influencing our health and well-being. Over a thousand different bacterial species constitute the highly complex ecosystem in the human intestine. High-throughput sequencing approaches have recently provided large datasets of metagenomic information but are limited in the ability to draw conclusions about the functionality of the intestinal microbiome. Rapid development of high-accuracy mass spectrometers with deep resolving power, together with enhanced protein separation technology, have promoted the field of metaproteomics, that is, the study of the collective proteome of microbial communities. We review here the technological development of high-resolution tandem mass spectrometry and its capability in analysis of complex microbial systems, particularly in the human gastrointestinal tract, and we present two relevant examples of the metaproteomic approach.

19.1.2 Strategies to Study Intestinal Microbiome

Microbiology is a constantly developing area of science with an extremely wide field of research, one of the biggest among different sciences. Our planet is estimated to accommodate an astronomical number of microorganisms, including more than 10^{30} prokaryotes, namely, bacteria and archaea [1]. We humans provide an excellent environment for microbes in our own body. Among scientists and physicians, there is an increasing awareness of the impact of our colonized microbiome on our health [2]. A significant amount of microbes can be found in the oral cavity, stomach, as well as in the vagina,

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

but the most diverse spectrum of microorganisms can be found in our gastrointestinal (GI) tract. The GI microbiota of healthy subjects contributes to the control of nutrient intake, immune homeostasis, as well as gut development, while an unbalanced gut microbiota is often associated with pathogenic conditions, such as infectious diarrhea, inflammatory bowel diseases, obesity, diabetes, and colon cancer [3, 4]. The human GI microbiota may amount to approximately 10^{14} microbes, mainly bacteria and archaea, which are derived from several thousands of species or species-like taxa, most of which have not yet been cultured [4, 5]. The majority of the intestinal microbiota belong to the limited set of phyla such as Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, and Verrucomicrobia [5, 6]. However, it is already well established that each human has a unique microbial community living in his or her GI tract [4–7]. Most of the initial studies of the human GI microbiota included culture-based approaches [4, 5]. Although a first insight into the GI ecosystem could be obtained using this technology, it became rapidly apparent that the human intestinal microbiota was far too complex to be analyzed using these methods of classical microbiology. Moreover, many of the GI microbes are strict anaerobes that are hard to culture [5, 8].

Revolutionary progress in high-throughput technologies has brought new tools for studying the complex GI microbiota. These mainly nucleotide-based approaches are culture independent and can be performed in parallel with high efficiency and rapid computation [3, 4]. These tools include quantitative polymerase chain reaction (Q-PCR) analysis, PCR-based DNA profiling techniques, phylogenetic microarray and sequencing studies, flow cytometry, and next-generation DNA sequencing techniques [3]. All of these methods are associated with novel bioinformatics platforms. Notably, the next-generation sequencing platforms have allowed for a dramatically increased dataset; a first baseline study of the collective genomes of the human GI microbiota, the metagenome, provided over 3 M unique genes [9] and has been used as a starting point for further developments (see below).

The next step after generating the inventory of the GI metagenome is to determine which genes are expressed at what time in the human body [4]. This is an important step as phylogenetic and metagenomic approaches may reveal candidate species and genes that may be important in health or disease. However, they do not provide evidence on the actual involvement of these species or genes. Hence, functional approaches are needed that aim to identify the active species and molecules. These also provide information on interactions between the microbes and the host, which are extremely relevant for the health impact. The experimental approaches to address microbial functions, however, are more challenging than collecting metagenomic data. The first possibility of the omics-based approach is transcriptomics, where mRNA expression levels are studied to reveal the functionality of selected genes [10]. Metatranscriptomics studies have turned out to generate only limited meaningful information, mainly due to the instability of the prokaryotic mRNA, which requires rapid sampling and processing. Another possible approach for community-level analysis is metabolomics, where the functionality of microbial species can be seen in metabolic products of different biological processes [11]. Metabolite analysis is quite challenging and needs high-throughput instruments as well as advanced data analysis tools. Moreover, metabolites produced by the GI microbiota are often very quickly absorbed, and thus their analysis does not fully correlate with the dynamic information of the ecosystem. Metaproteomics does not suffer so much from these problems as it addresses relatively stable proteins

and hence is currently considered a powerful alternative for analyzing the functionality of the complex GI microbiota as will be described here.

Proteomics is a term of the "omics" era, describing methods to analyze the proteome, that is, a set of proteins expressed by a genome in a given time and under certain conditions. The methodology development has been rapid, and both the development of protein separation technology as well as high-throughput mass spectrometry (MS) has stimulated the field of metaproteomics, the analysis of all proteins present in an ecosystem. Proteins are rather stable molecules, compared to RNA or metabolites, and they are related directly to the genetic code, thus making proteomics a template-driven technology. Therefore, metaproteomics has gained enormously from the rapidly growing metagenomic databases. Here, the simplicity of prokaryotic proteomes, which is limited in post-translational processing, is an advantage. During the last decade, only a few metaproteomic studies on the GI microbiota have been executed [12–14]. This renders metaproteomics a rapidly developing research field with high potential [8, 15]. Here, we present an update on the development of the MS technology that has been instrumental in advancing the field of GI metaproteomics. Furthermore, we introduce studies that apply this state-of-the-art methodology.

19.2 MS in Microbiology

The development of gentle ionization methods enabled MS of biological macromolecules three decades ago. Protein MS became possible due to the rather simultaneous inventions of matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) and is now adopted as a widely used tool in identification as well as characterization of microbiota [16]. The importance of these inventions is illustrated by the Nobel Prize in Chemistry in 2002 awarded to Koichi Tanaka [17] and John Fenn [18]. These innovations lead to the fast development of proteomics [19] and raised tremendous expectations for the clinical application of proteomic research.

Despite the increased attention on protein research and the over thousand articles published every year about plausible protein biomarker candidates [20], clinical implementation remains limited. By the year 2009, the FDA had approved only 109 protein biomarkers for clinical use, of which more than 80% was validated long before protein MS was even possible. Today only one or two protein biomarkers get FDA clearance annually [20]. The obstacles for clinical utility of protein biomarkers are diverse. They include complex sample matrices, a wide dynamic range of peptide/protein analytes, plausible preanalytical modifications of the analyte, and persistent challenges in quantitation. The enormous number of possible biomarker candidates limits further progress because their validation is a tedious and time-consuming process [20]. Hence, the high expectations set for protein biomarker clinical MS are currently not met.

For the majority of proteomic scientists, it may have come as a surprise to see protein MS having the first wide clinical use in the field of microbiology. Microbial identification has for long been time consuming and a laborious effort mainly relying on culturing and biochemical methods. In the past decade, nucleotide-based detection and identification systems have entered the field of clinical microbiology but still require significant time and investments [16]. Therefore, the entry of MS-based diagnostics with low cost, fast performance, and robustness has been very much welcomed. The

first MS-based identification of gram-negative bacteria was reported already in 1975 [21]. After development of soft ionization methods, the possibility of identifying microbes on the basis of their protein profiles became evident, as detailed elsewhere [22–24]. MALDI has widely been shown to enable protein ionization from intact microbial cells, and the time-of-flight (TOF) mass analyzer is capable of providing m/z information usable in pattern matching leading to microbial identification [24]. The MALDI-TOF technology utilized in clinical microbiology laboratories today relies mostly on the analysis of ribosomal proteins that are present under all growth conditions, as testified by comprehensive interlaboratory evaluations [25]. MALDI-based microbial identification has rapidly gained an important position in clinical microbiology laboratories due to commercially available MALDI-TOF systems and proprietary microbial databases [24]. One of the most appreciated properties of current MALDI-TOF technology is its ease of use. The biggest limitation is the limited amount of information obtained; that is, with MALDI-TOF technology there is no possibility of targeting specific protein biomarkers, such as resistance markers and virulence factors, or handling complex microbial samples.

Although MALDI-based MS has in recent decades become an established technology in the management and control of infectious diseases, tandem MS has become increasingly utilized in attempts to understand and unravel complex microbial systems. The LC ESI–based tandem MS setup is currently the leading technology in the microbiological research field, whereas MALDI has occupied the clinical market. Here, the main determining factor is sample complexity (see Table 19.1). Although MALDI can still retain a portion of microbial proteomic research owing to the low complexity of microbial cells and the gel-based sample preparation workflow, the scientific community has an increasing interest in microbial ecosystems having enormous complexity. Here, MALDI-based technology evidently loses its resolving capability, whereas ESI tandem MS can provide better sensitivity, increased dynamic range and thus resolving power, wider proteome coverage, and the possibility of analyzing challenging proteins, such as hydrophobic as well as glycosylated and phosphorylated proteins. An impressive illustration of the power of tandem MS in microbial proteome coverage has been provided by Haroun Shah in determining the pathotypes and toxin levels during the *Escherichia coli* O104:H4 outbreak in Europe through the summer of 2011 [26]. The analytical

Table 19.1 Items determining decision making when selecting either MALDI- or tandem-MS-based technology.

MALDI	Tandem MS
+ Ease of use	– Challenging usability
+ Speed	– Slow gradient in LC-MS
+ Low operational cost	– Moderate operational cost
– Limited information amount	+ High information amount
– Pattern matching approach	+ Protein sequencing capability
–Limited protein detection	+ Targeted protein analysis
– Simple sample	+ Complex sample

power in microbial surface protein characterization has also been demonstrated in multiple other studies. For *Staphylococcus aureus*, the proteomic study of the cell surface adhesions has clearly demonstrated the protein expression correlation to adherence phenotypes [27]. In a proteomic study of *Listeria monocytogenes* cell wall components, novel information was obtained from the cell wall remodeling process [28]. The study of *Tannerella forsythia* S-layer protein illustrated the challenging analytics of glycosylated proteoforms [29]. Proteogenomic studies of secreted proteins, the so-called exoproteomes, have also been performed. One excellent example can be found for the marine bacteria *Roseobacter* spp. [30]. All of the mentioned cases represent tandem mass spectrometric examples in the field of extremely diverse microbiology. The main obstacle in further development of the microbial proteomic technology is the dependence of sequence-based proteomics on genomic information. Even though the amount of genomic data is currently growing at an extraordinary rate, it is still far from enabling analysis of the complete coding capacity of the microbial biodiversity. However, microbial proteomics clearly benefits of the ongoing technical developments. Mass spectrometers and LC-based separation techniques are being advanced continuously, and sample preparation protocols are being standardized and automated. These developments are steadily increasing the user-friendliness of this technology and will finally promote applications in a clinical setting.

The development of high-resolution MS technology is one of the key issues in future implementation of tandem MS both in microbiological research as well as in clinical practice. Improvements in MS technology have indeed been enormous during the last two decades, and these have enabled all the time deepening analysis of complex proteomes. The entry of Orbitrap technology [31, 32] in the MS analyzer market ten years ago, following that of ion traps, quadrupole mass filters, and TOF mass analyzers, can be considered a remarkable milestone. As excellent Orbitrap performance can be achieved within a timeframe compatible with modern nano-LC instrumentation, the technology has already shown promising results in complex microbial studies. Very low-ppm mass accuracy has proved to be one of the main issues related to the resolving power of complex protein mixtures [33, 34]. The other critical factor here is the high resolution of the Orbitrap analyzer [35]. For fragmentation purposes, the Orbitrap can be linked to another mass analyzer, thus generating a hybrid tandem MS. The first Orbitrap tandem MS instrument in the market introduced a linear ion trap linked to the Orbitrap analyzer [36, 37]. The LTQ Orbitrap platform available for 10 years has become a globally widespread instrument, and the majority of published tandem mass spectrometric studies on microbes have been performed with this instrument [26–28, 30, 38]. Later, a modified release of the originally developed instrument called LTQ Orbitrap Velos was introduced [39]. This instrument possesses improved ion transmission, a dual linear ion trap, as well as a more efficient collision cell, and its power has also been demonstrated in challenging microbial proteomic setups [29]. We have utilized the LTQ Orbitrap in our analyses of complex gut microbiota, as shown in the two case studies in this chapter [14]. The newcomer in the Orbitrap family is a high-performance benchtop instrument, where a quadrupole mass filter is coupled to an Orbitrap analyzer [40]. This Q Exactive instrument has been on the market for about four years now, and it has been received very well, both in the general bioscientific community as well as by microbiological researchers [41]. This is likely due to the high performance coupled to a small size. The schematic representation of the instrument can be seen in Figure 19.1.

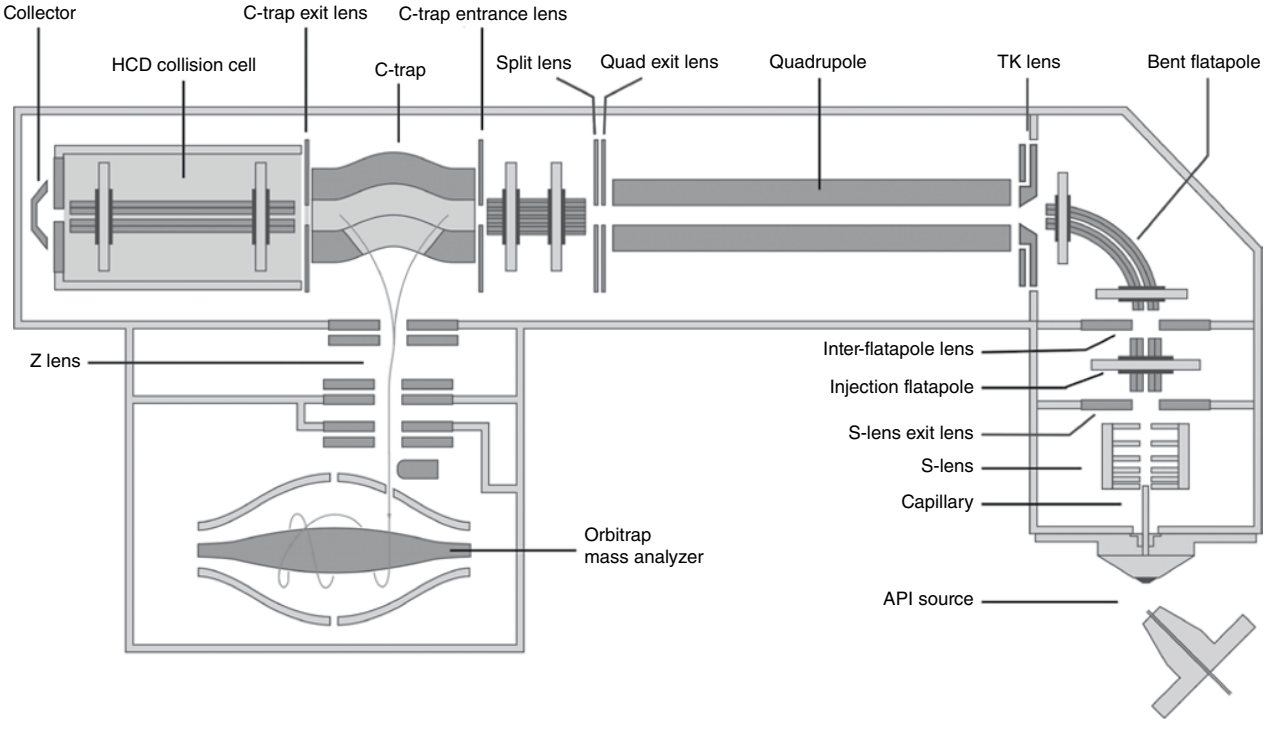


Figure 19.1 Q Exactive instrument layout. The instrument incorporates an S-lens, a selective quadrupole mass filter, a HCD collision cell interfaced to the
C-trap, and a high-resolution Orbitrap mass analyzer. The figure is reprinted with permission from Thermo Fisher Scientific.

The main parts of the instrument are the S-lens for ion transmission, a mass selective quadrupole, a collision cell directly facing the C-trap and, finally, the high-resolution Orbitrap analyzer. In the Q Exactive MS family, two modified models with improved performance have been introduced [42].

One option that high-resolution MS brings to the field of microbiology is top-down proteomics [43], namely, analyzing and identifying intact proteins instead of digested fragments, as is done in the bottom-up approach [44]. The top-down technology incorporates several critical features into the field of proteomics. These are speed and the possibility of accurately characterizing post-translational modifications (PTMs) of proteins. The speed is achieved because no time-consuming proteolytic digestion is needed to perform bottom-up proteomics. As top-down proteomics has already been used in the field of microbiology [45], we also demonstrated the resolving capability of the top-down approach for microbial proteomics [46]. Even in the chromatographically unresolved *Escherichia coli* extract, 66 unique proteoforms could be identified. When preanalytical LC was performed using 5, 15, 30, and 60 min gradients, the number of proteoforms identified was 722, 996, 1395, and 1964, respectively. The study was performed on a Q Exactive HF instrument, and its 1.7-fold increased resolving power was revealed in discriminative proteomic analysis of *Escherichia coli* and *Schigella sonnei* species, which are of clinical relevance. The capability of top-down proteomics in the analysis of protein PTMs has been demonstrated earlier [47], but lately also in the field of microbiology. In recent studies by Chamot-Rooke, it was shown for another relevant clinical bacterium, *Neisseria meningitides,* that protein modifications of the pilin, an important pathogenicity factor,can be revealed using top-down MS technology [48, 49]. Even though top-down proteomics has been shown to be a very powerful tool, it is not widely used. This is likely due to the need for an extremely high-resolution MS instrument, although an Orbitrap analyzer has been shown to perform well in top-down analyses [50]. Another reason for the limited use of the top-down approach is the lack of fully optimized tools for the process, both pre- and post-analytical. The former requires further development of intact protein LC methodology, and further development is also required for nanoscale use, which is actively under way among many suppliers. However, the latter issue has still not been clearly resolved. Despite the huge progress made in the field of bioinformatics [51], there is currently only one commercially available software tool for top-down proteomic data analysis [52]. Some noncommercial software packages are available, and it is likely that these and other tools will be of further use in high-resolution top-down proteomics.

Although proteomics in general is, and has been for more than a decade, in a transition state from simple MALDI-based technology to more advanced tandem MS, novel "state-of-the-art" strategies still appear in the field of microbial proteomics. One of the most outstanding ones is that of metaproteomics, where the focus of studies is put on microbial communities that can be extremely complex such as those found in the human GI tract (see above). Hence, here we describe complementary approaches focusing on this highly relevant ecosystem. First, a review on studies addressing all the proteins in fecal samples is provided. Second, a more targeted approach is described that focuses on surface proteins found in intestinal microbiomes.

19.3 Intestinal Metaproteomics Addressing All Proteins

As for gut metagenomics, the routine sample for studying the human intestinal microbiota has been fecal material (Table 19.2), which is rich in host, microbial, and diet-derived components. Hence, analyzing their proteins all together in a metaproteomics approach will provide functional information on the entire ecosystem. As in classical proteomics, various solutions are available for each of the steps of a metaproteomics experiment. When reviewing the first studies on fecal metaproteomics, a critical evaluation of these solutions has been reported [8]. However, direct comparisons of different methods and their combinations cannot be made, as the number of studies is limited. Hence, we provide here a summary of the pros and cons, while highlighting recently published work.

19.3.1 Preprocessing of the Sample

To analyze the proteins present in feces, two main approaches have been followed: (1) extracting proteins from fecal material without prior fractionation, or (2) preparing a bacterial pellet either by density gradient or differential centrifugation [6, 14, 53]. A recent study compared differential centrifugation to protein extraction of unfractionated material and found more microbial proteins covered by the separation approach but an underrepresentation of proteins from *Bacteroidaceae*, which may have been lost in the centrifugation step as they were bound to food particles in the spin-down pellets [54]. A decision on the planned sample fractionation, therefore, depends not only on the expected identification coverage but also on the available analysis time. The use of only fecal water, the liquid phase of feces, has been described as well and was claimed to be an adequate strategy to identify human proteins [55]. However, the number of

Table 19.2 Overview of fecal metaproteomic studies in humans.

Study subjects	Subject No./sample no. per subject	Analysis platform	Reference
Healthy twins and twins with CD	12/1	SCX-LC-MS/MS	[53, 64, 78]
Lean subjects	2/1	1D PAGE LC-MS/MS	[6]
Healthy subjects	3/2	1D PAGE LC-MS/MS	[14]
Subject at baseline, during antibiotic treatment, and afterward	1/6	1D PAGE LC-MS/MS	[79]
Lean and obese adolescent	2/1	1D PAGE LC-MS/MS	[56]
Healthy subjects	16/3	1D PAGE LC-MS/MS	[56]
Non-obese and obese subjects	29/1	1D PAGE LC-MS/MS	[60]
Healthy subjects and CD patients	12/1	2D-DIGE	[59]

identified proteins was not found to be critically higher than when using the unfractionated approach for protein extraction and amounted to approximately 10%–15% of the total spectral counts in adults [56]. Human mucosal biopsies may provide important information on microbes associated with this site, but these samples biopsies have not been used for untargeted metaproteomics.

19.3.2 Protein Extraction

Protein extraction has been performed by bead beating with either phosphate buffered saline (PBS, mimicking body fluids) or buffers containing detergents [8, 56]. The use of a detergent buffer in combination with sonication has also been applied [53]. While a direct comparison of these methods is not available, and spectra are processed in a different way and results have been reported in diverse ways, it appeared that the major functional proteins were not found to differ too much, suggesting that no direct bias emerges from the protein extraction. In other application fields of metaproteomics, such as the soil ecosystem, protein extraction methods have been directly compared to each other. A recent study found that a method yielding the most total protein amount may not guarantee the highest protein coverage [57].

19.3.3 Protein Digestion

For enzymatic digestion of proteins into peptides for LC-MSMS measurements, in-gel digestion has been used most widely, as also in-solution digestion [53] and filter-aided digestion have found application [58]. The latter two approaches might be more suitable for bacterial pellets, which contain less interfering substances than crude fecal material.

19.3.4 Peptide Fractionation

Due to the obvious complexity of fecal proteins, protein or peptide fractionation is required prior to LC-MSMS. Molecular-weight-based fractionation by 1D gel electrophoresis has been commonly used. This has the advantage of simultaneous sample cleaning. One research group used the MudPit approach, that is, separating the peptides via ion-exchange chromatography prior to reverse phase (RP) LC-MSMS [53]. Another group used long gradients on the RP chromatography for sample fractionation [58]. Finally, 2D-DIGE has been described, but this application is rather complex with respect to sample preparation and biased toward proteins getting separated by the two dimensions [59].

A common problem when analyzing a complex proteome is finding a balance between analysis time and analysis depth. We developed a reductionist approach that is applicable to several dozens to hundreds of samples; we focused on the ~40–80 kDa sized proteins which contained more than 80% of proteins detected over the whole range when separating a sample by 1D PAGE; we have demonstrated the reproducibility of this approach and applied it to over 50 biological samples [14, 56, 60].

19.4 LC-MSMS Analysis

The least variation in fecal metaproteomics studies is seen in the combination of LC and MS. Usually, peptides are separated by nano-LC, ionized by ESI, and measured by tandem MS. Gradient lengths varied from 50 min for each of the 20 fractions of a fecal sample [6], over 8 h for a single whole lysate [58], to 22 h for SCX fractions [53].

19.5 Data Analysis

One very crucial part of a metaproteomic experiment, which recently received considerable attention, is converting the measured mass-to-charge ratios into peptide sequences. There are three well-described options: de novo sequencing, spectral library analysis, and peptide spectral matching (PSM); nowadays, the latter is routinely used in most studies, and de novo sequencing has only been used in explorative studies of fecal metaproteomics but holds great promise [60]. Hence, we focus here on PSM and de novo sequencing.

19.5.1 Peptide Spectral Matching

The challenge in applying PSM in fecal metaproteomics comes from the large theoretical amount of proteins present in the sample. Whereas a protein sequence database in a classical proteomics experiment is as large as approximately 10,000 protein sequence entries, the fecal metaproteome is much more complex, and following the baseline report of over 3 M genes [9], currently already 10 M microbial genes have been identified in the fecal metagenome of adults [61]. It has been shown recently that these large sequence spaces challenge the search algorithms and the false discovery rate (FDR) approaches based on decoy searches, which is routinely applied for hit validation. To point to this problem, a simple *Pyrococcus* proteomic data was searched in a *Pyrococcus* dedicated database (with 10,000 proteins) and in our intestinal metagenomic database, which at that moment coded for approximately 6 M proteins. By applying the same stringent FDR, many hits were lost in the intestinal metagenome-based sequence database due to an over-estimation of the false positives [60]. Less stringent search parameters, like a higher number of allowed sequences or semi-specific enzyme settings, showed the same trend. To circumvent the loss of peptide identifications due to the large search space, several solutions have been proposed. The first was a so-called iterative workflow: a first search is performed against a collection of bacterial genomes called the synthetic metagenome [6]. These hits are used for a BLAST search against a large metagenome sequence database, and the hits from the BLAST search are used to create a second database. This second database is used for the final search. In a similar approach, first a large metagenome database was used for PSM, then all the hits from the first search were used to build a second database, which was then searched and a strict filter was applied on the results [62]. The caveats of this approach have been discussed recently [60]. Whereas the FDR is overestimated when the search space becomes too large, it becomes underestimated when there is an overrepresentation of very likely hits. Another proposed solution is having several protein sequence databases instead of one and searching them separately [63]. There is still an evident need for increasing the sensitivity and accuracy in metaproteomics PSM and for sophisticated software to help with this.

19.5.2 De Novo Sequencing

Due to the obvious obstacles encountered in applying PSM to metaproteomic data, de novo sequencing appears as welcome alternative. There are two application examples [60, 64]: the first used de novo sequencing as an exploratory approach and contrasted

the number of identified peptides to those obtained by PSM, and the latter matched the de novo sequencing hits with those obtained by PSM and cross-matched these with the original database used for PSM. A critical amount of de novo sequencing peptides was missed by the database searches but could be mapped to the in-silico digest of the protein sequence database used for PSM [60].

19.5.3 Protein Quantification

Most protein quantifications in metaproteomics are based on counting the identified peptides per protein, functional, or taxonomic group. Both protein-normalized and protein-non-normalized approaches have been applied [14, 53]. In addition, MS-based data have been used for quantification but less frequently than MSMS-based quantification. For a successful MS-based quantification, there are three requirements: (1) reliable software for normalization and aligning, (2) resolving the problem of protein inference, and (3) good computer power. Both the noncommercial Maxquant [56] as well as the commercial software Progenesis LC-MS have been used for this task [14].

19.5.4 Metaproteomic Pipelines

To streamline the complex metaproteomic data analysis, pipelines offering solutions for several steps at once are convenient. One such example is the MetaProteomeAnalyzer (MPA), which provides a platform for PSM with up to four search algorithms [60]. In the case of Uniprot headers contained in the used protein sequence database, taxonomic and functional analysis of the data is provided [60]. Another option is Maxquant, not specifically designed for metaproteomics but having been used in metaproteomics, which includes PSM- and MS-based quantification. Maxquant-derived data can be statistically analyzed with the software Perseus, developed in conjunction with Maxquant.

19.5.5 Data Storage

Nowadays metaproteomics mass spectrometric data are increasingly deposited in data repositories such as PRIDE and MASSive. The idea behind this deposition is to allow reanalysis with advanced databases or software. However, as even a single study containing dozens of samples can already be very complex, it is to be questioned who is taking the effort to reanalyze previous data; this is only possible when several issues, like optimizing the peptide identification rates and resolving the protein inference problem, are improved and metaproteomic data analysis becomes more straightforward than it is today.

19.6 Data Output and Interpretation

As with classical proteomics and other omics methods, a large variety of data utilization and interpretation possibilities are available. Both MS and MSMS profiles have been used for between-sample comparisons. For example, MS profiles have been compared on the basis of Pearson correlation coefficients, and subject specificity of the human intestinal fecal metaproteome was found; a similar result was obtained when applying the Jaccard Similarity Index on peptide identifications (MSMS data) [56].

In most studies, peptide identifications have been used as data output. Both functional and taxonomic information have been attributed to fecal proteins. To identify the major functionalities, the orthological databases KEGG [65], COG [66] and, recently eggNOG [67] have been used. After over 10 years, COG was updated in 2014, and therefore functional annotation is again a possibility. In the studies reporting a ratio for spectra with functional assignment, a study applying eggNOG reported the highest coverage (90% vs. 70%) [60].

Although metaproteomics is not the method of choice for describing the microbial composition, assigning a taxonomic unit to a peptide has been popular. With knowledge of its caveats, and excluding household proteins that are highly conserved between bacteria, phylogenetic studies of peptides can be informative. Simple approaches have reported taxonomic information based on the species origin of protein hits. One peptide can easily be derived from several dozen and up to even thousands of different proteins. For example, the COG database for glutamate dehydrogenase includes over 10,000 sequences, but the majority of COG protein entries contains only some dozens. Because not all of these options are routinely reported by search algorithms, this approach may produce erroneous results.

Recently, the lowest common ancestor analysis (LCA), which assigns the lowest taxonomic level on which all matches agree to the peptide, based on complete peptide matching, has been applied widely. In fecal metaproteomics studies, this was first applied with the help of Python-based in-house scripts; today the low-threshold web application Unipept is used [68]. This system is now also available as an application programming interface (API) for integration into other software. However, for this approach also, it has to be kept in mind that the databases used for the mapping are far from complete, and peptides with only a single amino acid difference will not obtain a taxonomic assignment.

19.7 Development of Surface Metaproteomics for Intestinal Microbiota

The vast number of proteins in complex microbial communities, such as intestinal microbiota, are a challenge in metaproteomics approaches. Methods that focus on specific fractions of proteomes, instead of total proteomes, could be easier to perform and provide improved sensitivity. Bacterial interactions with the host are often mediated by bacterial surface structures. Methods that focus on bacterial surface proteins could thus give important host–bacterial-interaction-targeted information on the functionality of microbial communities. So far, studies on bacterial surface proteomes have concentrated on analysis of single-species preparations. We demonstrated profiling of surface proteomes of complex microbial populations, represented by intestinal microbes in a fecal sample. The profiling protocol includes (1) separation of intact bacteria from a fecal sample, (2) enrichment of bacterial surface proteins, (3) identification of the proteins by MS, and (4) data analysis.

The surface profiling protocol showed promise as a proteomic tool in analysis of samples from healthy individuals. In the future, the goal is to utilize the protocol in studies of intestinal disorders also. Microbial surface proteome profiles as such could be useful in the diagnostics of intestinal inflammation. The protocol could also be used as

a discovery tool for identification of novel biomarkers for diagnostics and monitoring of intestinal disorders and their response to therapy.

19.7.1 Isolation of Bacteria from Fecal Samples

Separation of intact bacterial cells from fecal material was performed by the differential centrifugation method developed by Apajalahti *et al.* [69]. Fecal samples were diluted in a detergent-containing buffer, rotated thoroughly, and centrifuged. The pellet was resuspended in buffer, and the process was repeated three times. The recovery rate of bacterial cells, estimated by microscopic counting before and after the treatment, varied from 42% to 47% in different experiments. As indicated above, this separation may introduce some bias, but for the proof-of-concept approach described here, this was not further studied.

19.7.2 Enrichment of the Surface Proteome from Fecal Bacterial Extract

For the enrichment of surface proteins from the separated fecal bacteria, the methodology that we have set up for studying surface proteomes of single-species bacterial preparations [Lähteenmäki *manuscript in preparation*] was applied. Fecal bacteria were labeled with biotin and lysed by repeated bead beating. A bacterial lysate was centrifuged and the supernatant containing the whole-cell protein lysate collected. Biotinylated proteins, that is, the proteins expected to have originally been located on the cell surface, were isolated with streptavidin-conjugated magnetic beads. Gel electrophoresis analysis showed that protein profiles in the biotin-enriched fraction and the whole-cell protein lysate differed, and enrichment of specific proteins in the biotin-treated fraction could be detected (Figure 19.2A).

The reproducibility of the method was tested by extracting surface proteins from three replicate samples from three individual donors. From one donor, an additional sample was collected after one month to study temporal effects on the fecal bacterial surface protein profile. As estimated by 1D gel electrophoresis, replicate samples from each individual had a similar major protein profile, there were some differences in the protein profiles in samples from different individuals, and the protein profile seemed to remain constant over the one-month time period (Figure 19.2B). Thus, the biotin-based extraction method seems to reproducibly enrich specific proteins from a complex fecal-derived microbial matrix.

19.7.3 Detection of Surface Proteins by LC-MSMS

The proteins in biotin-enriched surface preparations were identified by RP-LC tandem MS. We compared two sample pretreatment method: (1) in-gel reduction, alkylation, and trypsin digestion after resolving the sample by 1D gel electrophoresis and (2) in-solution digestion. The digested peptides were loaded onto a precolumn and separated in an analytical column with a linear gradient of acetonitrile and a flow rate of 300 nl/min. Eluted peptides were introduced to an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific Inc.) via an ESI Chip interface (Advion BioSciences Inc.) in positive-ion mode. On the basis of a full MS scan acquired on the Orbitrap detector, six data-dependent MSMS scans were acquired on the LTQ. Data files from the mass spectrometer were processed with Mascot Distiller (Matrix Science Ltd., version 2.2.1.0)

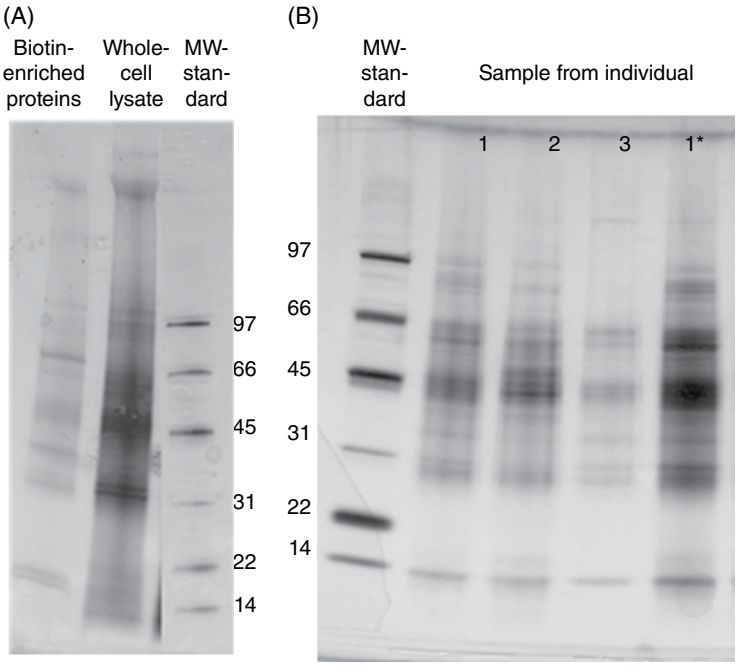


Figure 19.2 SDS-PAGE analysis of biotin-enriched surface proteins from fecal bacterial populations. (A) Comparison of biotin-enriched surface proteins and proteins in a whole-cell lysate of bacterial cells separated from a fecal sample by differential centrifugation. (B) Comparison of biotin-enriched surface proteins of bacteria from fecal samples collected from three healthy individuals (1–3). From donor 1, a sample (1*) was also collected one month after the first sample. For each sample, one example of three technical replicates treated separately is shown. Molecular weights (kDa) of standard proteins are indicated.

and searched against an in-house compiled database and the UNIProt KB database using Mascot Server (Matrix Science Ltd., version 2.2.04). The in-house database consisted of 490 intestinal bacterial genomes, the human genome, as well as selected genomes of eukaryotic species commonly present in food particles. Protein identifications were exported to ProteinCenter Software (Thermo Fisher Scientific Inc., version 3.7) for further analysis. For quantitative comparison of datasets, data files were preprocessed with Progenesis LC-MS software (Nonlinear Dynamics). Proteins having two or more peptide hits were included in the analysis.

The results showed that over 80% of the proteins detected after either in-gel or in-solution trypsin digestion originated from bacterial cells (Figure 19.3). The proportion of detected eukaryotic proteins was 15%–16%. Most of these probably originated from food particles, as only 2% of the proteins were estimated to be of human origin. This confirmed that the used centrifugation approach is not without bias. The detected bacterial proteins originated from bacterial taxa representing common fecal bacteria. Proteins from taxa including *Eubacterium, Faecalibacterium, Ruminococcus, Clostridium, Coprococcus, Bifidobacterium, Roseburia, Subdoligranulum, Dorea, Bacteroides*, and *Prevotella* were detected by both pretreatment methods. Thus, our surface protein profiling method seems suitable for detection of bacterial proteins.

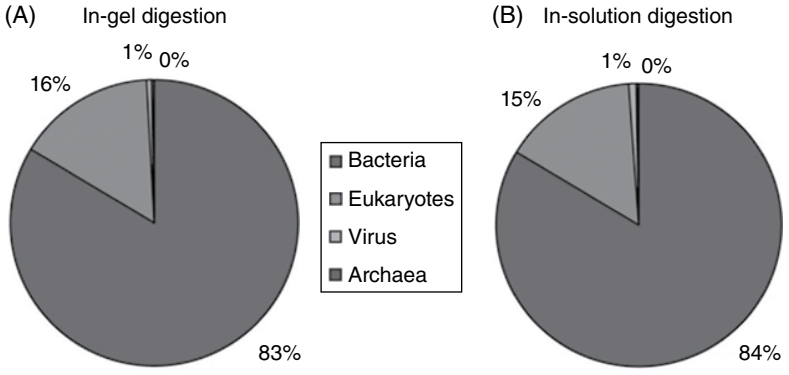


Figure 19.3 Origin of the proteins detected by the cell surface protein profiling method in a fecal sample. Ratios of proteins from bacteria, eukaryotes, viruses, and archaea detected after (A) in-gel trypsin digestion or (B) in-solution trypsin digestion are shown.

In total, we detected 533 bacterial proteins in the in-gel digested sample and 231 bacterial proteins in the in-solution digested sample. As expected, the gel-based separation of the proteins before MS-analysis yielded significantly more protein identifications by increasing the LC-MS/MS time spent by the sample. Only 14% the proteins were identified in both samples. However, the samples shared a considerably higher number (47%) of proteins with a high Mascot score (over 150 in the in-solution digested sample). This suggests that the most common proteins can be detected by both pretreatment methods, making the less laborious approach with in-solution digestion applicable for studies focusing on the dominant fraction of bacterial surface proteomes.

The proteins having significant amino acid homology (over 80% or 95%, in different experiments) with each other were clustered together in order to group the closely related proteins from different organisms. The 10 protein clusters with the highest Mascot scores are shown in Table 19.3. With the exception of the chaperonin GroEL, which was detected only in the sample resolved by SDS-PAGE, the most common protein clusters were detected in both samples irrespective of the pretreatment method. Characteristically for microbial samples, a relatively high amount of hypothetical proteins (22% in the in-gel digested and 26% in the in-solution digested sample) were detected, and 7%/14% of the protein clusters were yet unannotated. When cellular localization of the proteins was studied using the ProteinCenter database, only 36%/28% of the proteins were categorized as cell surface or extracellular proteins. However, the majority of the most commonly detected proteins were predicted to have membrane localization (Table 19.3). In the GO-ontology-based classification of biological processes and molecular functions, most of the detected proteins were classified to the categories of metabolic process (75%/89%) and catalytic activity (65%/84%). It should be noted that determination of the localization of bacterial proteins is not straightforward, because the location of many proteins is yet unknown. The ProteinCenter database information of protein localization is based on the presence of specific anchoring motifs, signal peptides, or transmembrane domains in the amino acid sequence. Many bacterial proteins found to be surface-localized have no such consensus motifs [70]. Furthermore, bacteria contain various "moonlighting proteins," which have an intracellular function but also can be localized on the cell surface [71]. Many moonlighting

Table 19.3 The 10 bacterial protein clusters detected with highest Mascot scores in biotin-enriched surface preparations of fecal bacteria. The proteins with over 80% homology with each other were clustered together. Mascot score shown is the maximal score of the proteins within a cluster. Cluster size indicates number of proteins clustered together.

Cluster size	Description	MASCOT score	Molecular functions	Cellular components	Biological processes
5	Transaldolase	1718	Catalytic activity	Cytoplasm/membrane	Metabolic process
23	Phosphoenolpyruvate carboxykinase*	1483	Catalytic activity/ Nucleotide binding	Cytoplasm/membrane	Metabolic process
5	Phosphoglycerate kinase	1400	Catalytic activity/ Nucleotide binding	Cytoplasm	Metabolic process
2	Hypothetical protein BIFADO_00189/ BIFPSEUDO_04022	1308	Catalytic activity	Membrane	Metabolic process
7	Glutamate dehydrogenase (NADP+)[a]	1260	Catalytic activity	Membrane	Metabolic process
6	Glucose-6-phosphate isomerase[a]	1249	Catalytic activity	Cytoplasm/membrane	Metabolic process
4	Chaperonin GroEL[a]	1215	Nucleotide binding /Protein binding	Cytoplasm/membrane	Metabolic process
3	Fructose-6-phosphate phosphoketolase[a]	1173	Catalytic activity	Unknown[b]	Metabolic process
4	Carbohydrate ABC transporter substrate binding protein/Tat pathway signal sequence	1114	Unknown (Transport activity)**	Unknown[b]	Unknown (Transport)**
3	Carbohydrate ABC transporter sustrate binding protein (CUT 1.1 family)	1090	Transport activity	Membrane	Transport

[a] Included also hypothetical proteins from one or more organisms.
[b] Reported unknown in ProteinCenter (information from other sources included when available).

proteins are metabolic enzymes or molecular chaperonins, and when localized on the cell surface, they often act in an adhesive capacity. Examples of moonlighting proteins that we identified with our surface metaproteomics protocol are phosphoglycerate kinase, glucose-6-phosphate isomerase, and the chaperonin GroEL (Table 19.3). These proteins have previously been reported in surface proteomic analysis of single species of lactic acid bacteria [72, 73], similar to various ABC transporter-associated proteins [73], which formed one group of proteins that are also enriched in our metaproteomic samples. Transaldolase, which also appeared in our samples with a high Mascot score, has been described as a surface protein in a proteomic analysis of a *Staphylococcus* species [74]. These results suggest that enrichment of surface proteomes of complex bacterial populations can be achieved with our protocol.

Reproducibility of the protocol was studied with three or more replicate fecal samples from three different individuals. The samples were analyzed by RP-LC-MSMS after in-solution trypsin digestion. Preliminary analysis showed that approximately 20% of the detected proteins were shared in samples from all three individuals. This is not considerably less than in replicate samples from each individual, in which also only 22%, 29%, or 31% of the detected proteins were shared. The high variation in replicate samples seems to result from a load of variable minor proteins. Importantly, principal component analysis (PCA) showed that replicate samples from each individual clustered together (Figure 19.4). PCA also indicated that there were significant differences between samples from different individuals (Figure 19.4). Thus, the results suggest that the surface metaproteomics method can detect individual differences in the dominant fraction of surface proteins in intestinal bacteria.

We developed the method by using fecal samples from healthy individuals. Future work is required to find out if the protocol can detect variation in bacterial surface proteomes between samples from healthy individuals and patients with intestinal inflammation. It is known that bacterial proteomes can be altered by contact with tumor epithelial cells in vitro [75] or in host intestine after experimental infection in vivo [76]. As surface proteins are critical for bacteria in functions such as host tissue adhesion and invasion, it may be that profiling of surface proteomes could give information on pathologic conditions in the intestine. Further development of the fecal

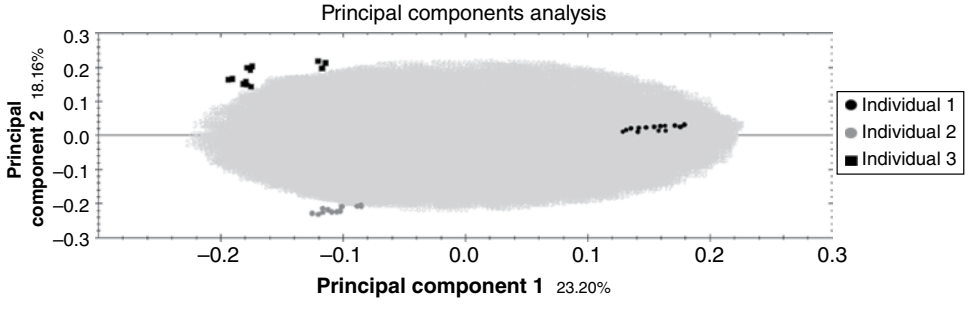


Figure 19.4 Principal component analysis of surface protein profiles of fecal bacteria. The bacteria were isolated from three or more replicate samples collected from three different individuals (1, black circles; 2, gray circles; 3, black squares). Biotin-enriched proteins were analyzed by RP-LC-MSMS as three technical replicates.

bacterial surface protein profiling would show how the protocol can detect possible proteomic signatures or changes characteristic of intestinal disorders.

19.8 Conclusions

Technological development in the field of MS has been dramatic in the past two decades. Included here is the ever-growing number of MS applications with biomolecules, mostly proteins. Life-science-associated MS has become more or less a field by itself in analytics, raising expectations regarding clinical implementations of protein biomarkers. Although not all of these expectations have been fully met, there is a wide range of studies showing that progress is in the offing.

The first widely deployed clinical application of protein MS occurred in the field of microbiology. MALDI-based microbial protein identification is a common practice in clinical microbiology laboratories all over the word, when a microbial species is to be identified. However, MALDI performance fails fast when the analyzable microbial sample contains multiple different species, not to mention extremely complex microbial ecosystems. Here, high-resolution and high-accuracy tandem MS has brought extremely usable tools to study complex microbiomes, such as the one colonized in human gut.

Analytical challenges in studying extremely complex set of microbes, the majority of which are probably still unknown, are immense. This explains the low number of GI proteomics studies reported so far, despite the large attention on the intestinal microbiota in human health and disease. The very first metaproteomic study on human intestinal microbiota focused on fecal samples from two infants, less than 10 years ago [12]. Similarly, the first human metagenomic data set [77] was utilized to mine the first adult intestinal metaproteome deriving from two Scandinavian individuals [53]. Here, more than 1000 different proteins could be identified, 30% of which were of human origin. To date, less than a dozen human intestinal metaproteomic studies have been published. We have presented here an overview of the possibilities and challenges when studying the metaproteome of the intestinal microbiota in its entire complexity as well as development of a technology for characterization of surface proteins in intestinal microbial populations, which extremely relevant when studying microbial–host interactions.

Even though the advancement of MS technology has been rapid and extremely promising, there are still many challenges to overcome in exclusive microbial ecosystem analytics. MS is only one part of the complex workflow, where many soft spots still exist. Sample preparation methodology is not straightforward, and special attention needs to be addressed here when analyzing the data acquired. Bioinformatic analysis tools need to be further developed, because human microbiome samples are of extreme complexity. Also, serious validation of the analytical reproducibility has not yet been widely addressed. The tandem MS methodology produces an enormous amount of data, and often only a limited set of information is highlighted, leaving most of the data unexamined. In order to process acquired data further and draw microbiological conclusions, both statistical and taxonomic tools need to be further improved.

Tandem MS enabling metaproteomics is likely to reinforce translational medicine associated with complex intestinal microbiota and its effect in human health.

High-resolution MS brings with it the secret promise of a comprehensive diagnostic tool in clinical microbiology, if not today, in the near future. Here, automated sample preparation, top-down proteomics, and advanced dynamic data analysis tools have the most potential for further development.

References

1 Whitman, W. B., Coleman, D. C., Wiebe, W. J. (1998). Procaryotes: The unseen majority. *Proc. Natl. Acad. Sci.* 95: 6578–6583.

2 Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science* 308: 1635–1638.

3 Gong, J., Yang, C. (2012). Advances in the methods for studying gut microbiota and their relevance to the research of dietary fiber functions. *Food Res. Int.* 48: 916–929.

4 Zoetendal, E. G., Rajilić-Stojanović, M., de Vos, W. M. (2008). High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut* 57: 1605–1615.

5 Rajilić-Stojanović, M, de Vos, W. M. (2014). The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol Rev.* 38: 996–1047.

6 Rooijers, K., Kolmeder, C., Juste, C., Dore, J., de Been, M., Boeren, S., Galan, P., Beauvallet, C., de Vos, W. M., Schaap, P. J. (2011). An iterative workflow for mining the human intestinal metaproteome. *BMC Genomics* 12: 6.

7 O'Hara, A. M., Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Rep.* 7(7): 688–693.

8 Kolmeder, C. A., de Vos, W. M. (2014). Metaproteomics of our microbiome – Developing insight in function and activity in man and model system. *J. Proteomics* 97: 3–16.

9 Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., and MetaHIT consortium (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464: 59–65.

10 Gosables, M. J., Durban, A., Pignatelli, M., Abellan, J. J., Jimenez-Hernandez, N., Perez-Cobas, A. E., Latorre, A., and Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS ONE* 6: e17447.

11 Jansson, J., Willing, B., Lucio, M., Fekete, A., Dicksved, J., Halfvarson, J., Tysk, C., and Schmitt-Kopplin, P. (2009). Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* 4(7): e6386.

12 Klaassens, E. S., de Vos, W. M., and Vaughan, E. E. (2007). Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl. Environ. Microbiol.* 73: 1388–1392.

13 VerBerkmoes, N. C., Denef, V. J., Hettich, R. L., and Banfield, J. F. (2009). Systems biology: Functional analysis of natural microbial consortia using community proteomics. *Nat. Rev. Microbiol.* 7: 196–205.

14 Kolmeder, C. A., de Been, M., Nikkilä, J., Ritamo, I., Mättö, J., Valmu, L., Salojärvi, J., Palva, A., Salonen, A., and de Vos, W. M. (2012). Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLoS ONE* 7: e29913.

15 Hettich, R. L., Pan, C., Chourey, K., and Giannone, R. J. (2013). Metaproteomics: Harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Anal. Chem.* 85(9): 4203–4214.

16 Sauer, S., and Kliem, M. (2010). Mass spectrometry tools for the classification and identification of bacteria. *Nature Rev. Microbiol.* 8: 74–82.

17 Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988). Protein and polymer analyses up to *m/z* 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2: 151–153.

18 Fenn, J. B. (2003). Electrospray wings for molecular elephants (Nobel lecture). *Angew. Chem. Int. Edn. Engl.* 42: 3871–3894.

19 Cox, J., and Mann, M. (2007). Is proteomics the new genomics? *Cell* 130(3): 395–398.

20 Lemoine, J., Fortin, T., Salvador, A., Jaffuel, A., Charrier, J.-P., and Choquet-Kastylevsky, G. (2012). The current status of clinical proteomics and the use of MRM and MRM^3 for biomarker validation. *Expert Rev. Mol. Diagn.* 12: 333–342.

21 Anhalt, J. P., and Fenselau, C. (1975). Identification of bacteria using mass spectrometry. *Anal. Chem.* 47: 219–225.

22 Emonet, S., Shah, H., Cherkaoui, A., and Schrenzel, J. (2010). Application and use of various mass spectrometry methods in clinical microbiology. *Clin. Microbiol. Infect.* 16: 1604–1613.

23 Fenselau, C. C. (2013). Rapid characterization of microorganisms by mass spectrometry—what can be learned and how? *J. Am. Soc. Mass Spectrom.* 24(8): 1161–1166.

24 Clark, A. E., Kaleta, E. J., Arora, A., and Wolk, D. M. (2013). Matrix-assisted laser desorption ionization-time of flight mass spectrometry: A fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* 26: 547–603.

25 Mellmann, A., Bimet, F., Bizet, C., Borovskaya, A. D., Drake, R. R., Eigner, U., Fahr, A. M., He, Y., Ilina, E. N., Kostrzewa, M., Maier, T., Mancinelli, L., Moussaoui, W., Prevost, G., Putignani, L., Seachord, C. L., Tang, Y. W., and Harmsen, D. (2009). High interlaboratory reproducibility of matrix-assisted laser desorption ionization-time of flight mass spectrometry-based species identification of nonfermenting bacteria. *J. Clin. Microbiol.* 47: 3732–3734.

26 Shah, H., and Gharbia, S. E. (2012). Using nano-LC-MS/MS to investigate the toxicity of outbreak *E. coli* O104:H4 strain. *Culture* 33(1): 1–4.

27 Ythier, M., Resch, G., Waridel, P., Panchaud, A., Gfeller, A., Majcherczyk, P., Quadroni, M., and Moreillon, P. (2012). Proteomic and transcriptomic profiling of *Staphylococcus aureus* surface LPXTG-proteins: Correlation with *agr* genotypes and adherence phenotypes. *Mol. Cell. Proteomics* 11: 1123–1139.

28 Garcia-del Portillo, F., Calvo, E., D'Orazio, V., and Pucciarelli, M. G. (2011). Association of ActA to peptidoglycan revealed by cell wall proteomics of intracellular *Listeria monocytogenes*. *J. Biol. Chem.* 286: 34675–34689.

29 Posch, G., Pabst, M., Brecker, L., Altman, F., Messner, P., and Schäffer, C. (2011). Characterization and Scope of S-layer Protein *O*-Glycosylation in *Tannerella forsythia*. *J. Biol. Chem.* 286: 38714–38724.

30 Christie-Olenza, J. A., Pina-Villalonga, J. M., Bosch, R., Nogales, B., and Armengaud, J. (2012). Comparative proteogenomics of twelve Roseobacter exoproteomes reveals different adaptive strategies among these marine bacteria. *Mol. Cell. Proteomics* 11(2): M111.013110.

31 Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* 72: 1156–1162.

32 Hardman, M., and Makarov, A. (2003). Interfacing the Orbitrap mass analyzer to an electrospray ion source. *Anal. Chem.* 75: 1699–1705.

33 Yates, J. R., Cociorva, D., Liao, J., and Zabrouskov, V. (2006). Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal. Chem.* 78: 493–500.

34 Macek, B., Waanders, L., Olsen, J. V., and Mann, M. (2006). Top-down protein sequencing and MS^3 on a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer. *Mol. Cell. Proteomics* 5: 949–958.

35 Scigelova, M., and Makarov, A. (2006). Orbitrap mass analyzer – overview and applications in proteomics. *Proteomics* 6 (Issue Supplement 1): 16–21.

36 Makarov, A., Denisov, E., Lange, O., and Horning, S. (2006). Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom.* 17: 977–982.

37 Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., and Horning, S. (2006). Performance evaluation of a hybrid linear ion trap/Orbitrap mass spectrometer. *Anal. Chem.* 78: 2113–2120.

38 Denef, V. J., Kalnejais, L. H., Mueller, R. S., Wilmes, P., Baker, B. J., Thomas, B. C., VerBerkmoes, N. C., Hettich, R. L., and Banfield, J. F. (2010). Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *PNAS* 107: 2383–2390.

39 Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., Wouters, E. R., Senko, M., Makarov, A., Mann, M., and Horning, S. (2009). A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* 8: 2759–2769.

40 Michalski, A., Damoc, E., Hauschild, J. P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011). Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* 10(9): M111.011015.

41 Zhao, Y., Sun, L., Champion, M. M., Knierman, M. D., and Dovichi, N. J. (2014). Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry for top-down characterization of the *Mycobacterium marinum* secretome. *Anal. Chem.* 86: 4873–4878.

42 Scheltema, R. A., Hauschild, J. P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., Kuehn, A., Makarov, A., and Mann, M. (2014). The Q Exactive HF, a benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell. Proteomics* 13: 3698–3708.

43 Catherman, A. D., Skinner, O. S., and Kelleher, N. L. (2014). Top down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun.* 445: 683–693.

44 Messana, I., Cabras, T., Iavarone, F., Vincenzoni, F., Urbani, A., and Castagnola, M. (2013). Unraveling the different proteomic platforms. *J. Sep. Sci.* 36: 128–139.

45 Wynne, C., Edwards, N. J., and Fenselau, C. C. (2010). Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using capLC-MS/MS on an Orbitrap. *Proteomics* 10: 3631–3643.

46 Damoc, E., Yip, P., Valmu, L., Delanghe, B., Denisov, E., Makarov, A., Gvozdyak, O., Cardasis, H., Neil, J., Cherkassky, A., and Stephenson, J. (2014). Performance evaluation of a benchtop quadrupole/high-field Orbitrap mass spectrometer for high-throughput top-down microbial proteomics. MSACL EU WedP32.

47 Lanucara, F., and Eyers, C. E. (2013). Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. *Mass Spectrom. Rev.* 32(1): 27–42.

48 Gault, J., Malosse, C., Duménil, G., and Chamot-Rooke, J. (2013). A combined mass spectrometry strategy for complete posttranslational modification mapping of *Neisseria meningitidis* major pilin. *J. Mass. Spectrom.* 48: 1199–1206.

49 Gault, J., Malosse, C., Machata, S., Millien, C., Podglajen, I., Ploy, M. C., Costello, C. E., Duménil, G., and Chamot-Rooke, J. (2014). Complete post-translational modification mapping of pathogenic *N. meningitidis* pilins requires top-down mass spectrometry. *Proteomics* 14: 1141–1151.

50 Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Muller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J.-J., Cox, J., Horning, S., Mann, M., and Makarov, A. (2012). Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol. Cell. Proteomics* 11(3): O111.013698.

51 Liu, X., Inbar, Y., Dorrestein, P. C., Wynne, C., Edwards, N., Souda, P., Whitelegge, J. P., Bafna, V., and Pevzner, P. A. (2010). Deconvolution and database search of complex tandem mass spectra of intact proteins. *Mol. Cell. Proteomics* 9: 2772–2782.

52 Zamdborg, L., LeDuc, R. D., Glowacz, K. J., Kim, Y. B., Viswanathan, V., Spaulding, I. T., Early, B. P., Bluhm, E. J., Babai, S., and Kelleher, N. L. (2007). ProSight PTM 2.0: Improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res*. 35: W701–6.

53 VerBerkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M. G., Apajalahti, J., Tysk, C., Hettich, R. L., and Jansson, J. K. (2009). Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* 3: 179–189.

54 Tanca, A., Palomba, A., Pisanu, S., Addis, M. F., and Uzzau, S. (2015). A human gut metaproteomic dataset from stool samples pretreated or not by differential centrifugation *Data Brief* 4: 559–562.

55 Lichtman, J. S., Marcobal, A., Sonnenburg J. L., and Elias J. E. (2013). Host-centric proteomics of stool: A Novel strategy focused on intestinal responses to the gut microbiota. Mol Cell Proteomics. 12: 3310–3318.

56 Kolmeder, C. A., Salojärvi, J., Ritari, J., de Been, M., Raes, J., Falony, G., Vieira-Silva, S., Kekkonen, R. A., Corthals, G. L., Palva, A., Salonen, A., and de Vos, W. M. (2016). Faecal metaproteomic analysis reveals a personalized and stable functional microbiome and limited effects of a probiotic intervention in adults. *PLoS ONE* 11: e0153294.

57 Ferrer, M., Ruiz, A., Lanza, F., Haange, S., Oberbach, A., Till, H., Bargiela, R., Campoy, C., Segura, M. T., and Richter, M. (2013). Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. *Environ. Microbiol.* 15: 211–226.

58 Hansen, S. H., Stensballe, A., Nielsen, P. H., and Herbst, F. A. (2014) Metaproteomics: Evaluation of protein extraction from activated sludge. *Proteomics* 14: 2535–2539.

59 Tanca, A., Palomba, A., Pisanu, S., Deligios, M., Fraumene, C., Manghina, V., Pagnozzi, D., Addis, M. F., and Uzzau, S. (2014). A straightforward and efficient analytical pipeline for metaproteome characterization. *Microbiome* 2: 49.

60 Kolmeder, C. A., Ritari, J., Verdam, F. J., Muth, T., Keskitalo, S., Varjosalo, M., Fuentes, S., Greve, J. W., Buurman, W. A., Reichl, U., Rapp, E., Martens, L., Palva, A., Salonen, A., Rensen, S. S., and de Vos, W. M. (2015). Colonic metaproteomic signatures of active bacteria and the host in obesity. *Proteomics* 20: 3544–3552.

61 Juste, C., Kreil, D. P., Beauvallet, C., Guillot, A., Vaca, S., Carapito, C., Mondot, S., Sykacek, P., Sokol, H., Blon, F., Lepercq, P., Levenez, F., Valot, B., Carre, W., Loux, V., Pons, N., David, O., Schaeffer, B., Lepage, P., Martin, P., Monnet, V., Seksik, P., Beaugerie, L., Ehrlich, S. D., Gibrat, J. F., Van Dorsselaer, A., and Dore, J. (2014). Bacterial protein signals are associated with Crohn's disease. *Gut* 63: 1566–1577.

62 Muth, T., Kolmeder, C. A., Salojärvi, J., Keskitalo, S., Varjosalo, M., Verdam, F. J., Rensen, S. S., Reichl, U., de Vos, W. M., Rapp, E., and Martens, L. (2015). Navigating through metaproteomics data: A logbook of database searching. *Proteomics* 15(20): 3439–3453.

63 Muth, T., Behne, A., Heyer, R., Kohrs, F., Benndorf, D., Hoffmann, M., Lehtevä, M., Reichl, U., Martens, L., and Rapp, E. (2015). The MetaProteomeAnalyzer: A powerful open-source software suite for metaproteomics data analysis and interpretation. *J. Proteome Res.* 14(3): 1557–1565.

64 Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J. R., Prifti, E., Nielsen, T., Juncker, A. S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X., Xiao, L., Liang, S., Zhang, D., Zhang, Z., Chen, W., Zhao, H., Al-Aama, J. Y., Edris, S., Yang, H., Wang, J., Hansen, T., Nielsen, H. B., Brunak, S., Kristiansen, K., Guarner, F., Pedersen, O., Dore, J., Ehrlich, S. D., Bork, P., and Wang, J. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nature Biotech.* 32: 834–841.

65 Jagtap, P., Goslinga, J., Kooren, J.A., McGowan, T., Wroblewski, M. S., Seymour, S. L., and Griffin, T. J. (2013). A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics* 13: 1352–1357.

66 Tanca, A., Palomba, A., Deligios, M., Cubeddu, T., Fraumene, C., Biosa, G., Pagnozzi, D., Addis, M. F., and Uzzau, S. (2013). Evaluating the impact of different sequence databases on metaproteome analysis: Insights from a lab-assembled microbial mixture. *PLoS ONE* 8: e82981.

67 Cantarel, B. L., Erickson, A. R., Verberkmoes, N. C., Erickson, B. K., Carey, P. A., Pan, C., Shah, M., Mongodin, E. F., Jansson, J. K., Fraser-Liggett, C. M., and Hettich, R. L. (2011). Strategies for metagenomic-guided whole-community proteomics of complex microbial environments. *PLoS ONE* 6: e27173.

68 Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40: D109–114.

69 Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43: D261–269.

70 Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldon, T., Rattei, T., Creevey, C., Kuhn, M., Jensen, L. J., von Mering, C., and Bork, P. (2014). eggNOG v4.0: nested orthology inference across 3686 organisms *Nucleic Acids Res*. 42: D231–239.

71 Mesuere, B., Devreese, B., Debyser, G., Aerts, M., Vandamme, P., and Dawyndt, P. (2012). Unipept: Tryptic peptide-based biodiversity analysis of metaproteome samples. *J. Prot. Res.* 11: 5773–5780.

72 Apajalahti, J. H., Särkilahti, L. K., Mäki, B. R., Heikkinen, J. P., Nurminen, P. H., and Holben, W. E. (1998). Effective recovery of bacterial DNA and percent-guanine-plus-cytosine-based analysis of community structure in the gastrointestinal tract of broiler chickens. *Appl. Environ. Microbiol.* 64(10): 4084–4088.

73 Bergonzelli, G. E., Granato, D., Pridmore, R. D., Marvin-Guy, L. F., Donnicola, D., and Corthésy-Theulaz, I. E. (2006). GroEL of *Lactobacillus johnsonii* La1 (NCC 533) is cell surface associated: Potential role in interactions with the host and the gastric pathogen *Helicobacter pylori*. *Infect. Immun.* 74(1): 425–434.

74 Kainulainen, V., and Korhonen, T. K. (2014). Dancing to another tune – adhesive moonlighting proteins in bacteria. *Biology* 3(1): 178–204.

75 Beck, H. C., Madsen, S. M., Glenting, J., Petersen, J., Israelsen, H., Nørrelykke, M. R., Antonsson, M., and Hansen, A. M. (2009). Proteomic analysis of cell surface-associated proteins from probiotic *Lactobacillus plantarum*. *FEMS Microbiol. Lett.* 297(1): 61–66.

76 Meyrand, M., Guillot, A., Goin, M., Furlan, S., Armalyte, J., Kulakauskas, S., Cortes-Perez, N. G., Thomas, G., Chat, S., Péchoux, C., Dupres, V., Hols, P., Dufrêne, Y. F., Trugnan, G., Chapot-Chartier, M.-P. (2013). Surface proteome analysis of a natural isolate of *Lactococcus lactis* reveals the presence of pili able to bind human intestinal epithelial cells. *Mol. Cell. Proteomics* 12: 3935–3947.

77 Bannoehr, J., Zakour, N. L. B., Reglinski, M., Inglis, N. F., Prabhakaran, S., Fossum, E., Smith, D. G., Wilson, G. J., Cartwright, R. A., Haas, J., Hook, M., van den Broek, A. H. M., Thoday, K. L., and Fitzgerald, J. R. (2011). Genomic and surface proteomic analysis of the canine pathogen *Staphylococcus pseudintermedius* reveals proteins that mediate adherence to the extracellular matrix. *Infect. Immun.* 79(8): 3074–3086.

78 Boleij, A., Dutilh, B. E., Kortman, G. A. M., Roelofs, R., Laarakkers, C. M., Engelke, U. F., and Tjalsma, H. (2012). Bacterial responses to a simulated colon tumor microenvironment. *Mol. Cell. Proteomics* 11(10): 851–862.

79 Pieper, R., Zhang, Q., Parmar, P. P., Huang, S.-T., Clark, D. J., Alami, H., Donohue-Rolfe, A., Fleischmann, R. D., Peterson, S. N., and TziporiPieper, S. (2009). The *Shigella dysenteriae* serotype 1 proteome, profiled in the host intestinal environment, reveals major metabolic modifications and increased expression of invasive proteins. *Proteomics* 9(22): 5029–5045.

20

Proteogenomics of Non-model Microorganisms

Jean Armengaud

CEA-Marcoule, DSV/IBITEC-S/SPI/Li2D, Laboratory "Innovative Technologies for Detection and Diagnostics", Bagnols-sur-Cèze, France

20.1 Introduction

Molecular characterization of pathogens and their detection has been drastically improved with the advent of next-generation sequencing of nucleic acids and comparative genomics, and later on, with the arrival of next-generation proteomics. High-throughput discovery proteomics is currently based on the shotgun bottom-up approach, where all the proteins of the sample are first proteolyzed with trypsin into peptides amenable to tandem mass spectrometry. These are then resolved by reverse-phase chromatography, their molecular weights analyzed by a first *m/z* ratio measurement by the mass spectrometer, fragmented, and then the MS/MS characteristics of the peptide fragments analyzed by means of a second *m/z* analysis. For interpreting the MS/MS spectrum of each peptide, the bioinformatics pipeline works best when a specific protein sequence database is available. Another strategy relies on the analysis of intact proteins by top-down proteomics such as what is done in whole-cell MALDI-TOF (Lavigne *et al.*, 2013), but to go further in their characterization, fragmentation of entire proteins should be performed. For this, novel tandem mass spectrometers and bioinformatics tools are required. Profiling of pathogens with such proteomics-based approaches has shown great potential for identifying numerous pathogens. The methodology is based on recording in the same experimental conditions the MS profile of proteins from thousands of microorganisms for compiling a comprehensive database and then, for a given pathogen, recording its specific profile and comparing it with the database. It is also possible to take advantage of the interpretation of peptide sequences generated by a bottom-up approach as demonstrated with the discrimination of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* strains (see Chapter 13; Dworzanski *et al.*, 2010). The application of this methodology to identification at the species level of non-genome sequenced bacteria shows great potential (Jabbour *et al.*, 2010; Karlsson *et al.*, 2012), opening a new era of accurate proteotyping in microbiology (Karlsson *et al.*, 2015). For characterizing the proteome of a given microorganism, sequencing of

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

the peptides is required, which depends on the availability of the appropriate protein sequence database; thus, proteomics relies heavily on genomics. The "proteogenomics" concept merges expertise and both pieces of information to gain higher insights into a given biological system. This chapter explains the concept and describes some of its applications that changed our views on what has been considered until now as non-model organisms.

20.2 The "Proteogenomics" Concept

The current avalanche of genome sequences is resulting in an exceptional amount of molecular information for numerous taxa. Because of the enormous number of genome sequences produced, automatic gene annotation with software applying consensual transcription and translation rules is mandatory. However, numerous errors can be generated during this annotation, especially when no already annotated closely phylogenetically related genome is available that could serve as template (Armengaud, 2009). Indeed, many genes are just missed if annotation parameters are too stringent, whereas open reading frames (ORFs) are frequently overcalled in poorly characterized phyla. It is noteworthy that identifying the correct translational initiation codons is far from trivial; in some bacteria, up to 20% of the polypeptides have been wrongly predicted in terms of N-terminus. For some organisms, the most used initiation codon is not ATG, leading to great difficulties in assessing the correct protein sequences. For example, the archaeon *Aeropyrum pernix K1* exhibits an atypical start codon preference with the codon TTG being the most preferred translational initiation codon far behind ATG and GTG (Yamazaki *et al.*, 2006). Experimental biochemical data on proteins acquired by shotgun proteomics can be exploited to better define the structural annotation of the corresponding genes. This new field of proteomics was called *proteogenomics* as the peptides certified by tandem mass spectrometry are mapped onto the nucleotide sequence before validating the corresponding gene structure. In this sense, proteomics expertise and genomics knowledge should be intimately linked. Figure 20.1 shows this

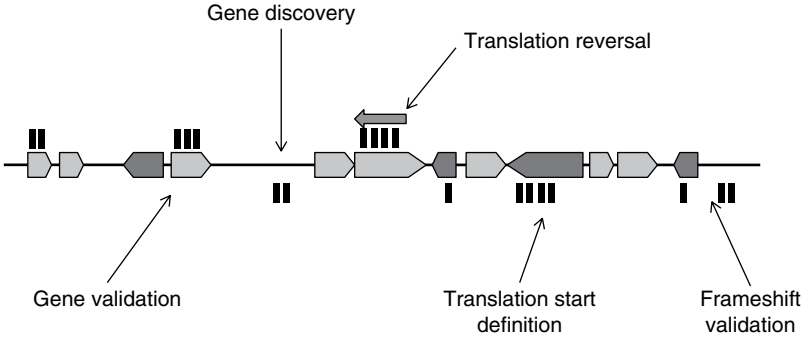


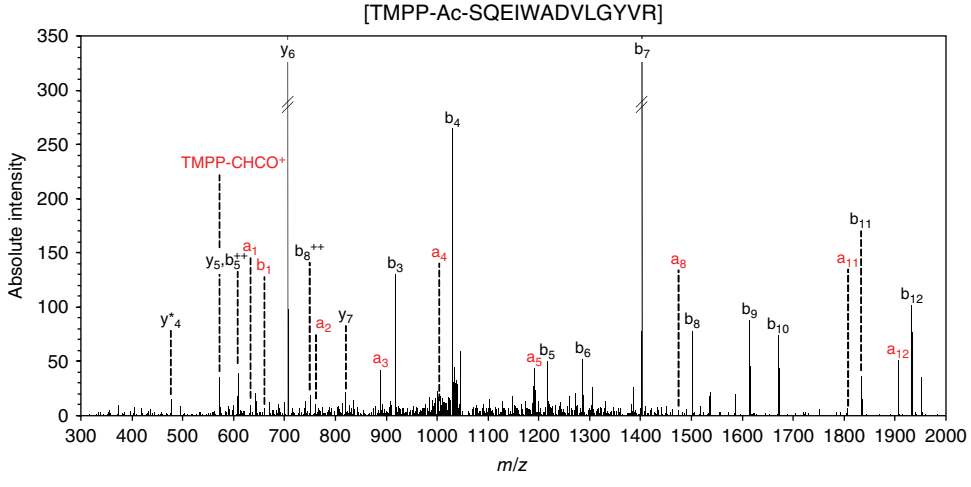
Figure 20.1 Proteogenomics map. The previously annotated genes are indicated with arrows, whereas the peptides established by tandem mass spectrometry are indicated with vertical bars. These peptides are mapped onto the nucleotide sequences. The different outcomes in terms of re-annotation of the genome are indicated.

proteogenomics mapping and the possible outcome in terms of genome re-annotation. Interpretation of MS/MS spectrum datasets is done against a database corresponding to a six-frame translation of the nucleotide sequences (Hernandez *et al.*, 2014). Peptide-to-spectrum assignment allows discriminating the correct ORFs from the whole list of all possible ORFs of this database.

Numerous studies have illustrated the power of proteogenomics for improving genome annotation as reported in recent reviews (Renuse *et al.*, 2011; Armengaud *et al.*, 2013; Kucharova and Wiker, 2014; Nesvizhskii *et al.*, 2014). However, a perfect annotation has still not been achieved, as proteomics is not yet comprehensive enough, especially for complex organisms such as eukaryotes for which the greater proteome dynamic range is the main difficulty (Armengaud, 2010). Many specific concepts and methodologies have been developed in the broad field of proteogenomics. For example, comparative proteogenomics relies on the use of proteomics data from several representatives of a given group of organisms in order to obtain a consensus for gene predictions of this specific group (Gupta *et al.*, 2008). In terms of methodology, labeling approaches allow tagging the N-termini of proteins and enriching them prior to mass spectral characterization (Bland *et al.*, 2014a; 2014b; Hartmann and Armengaud, 2014). With such a strategy, several studies conducted on the *Deinococcus deserti* bacterium led to unexpected results such as (1) the finding that reversal of gene sequences by automatic annotation pipelines is frequent, (2) the use of non-canonical start codons for translation is not only found in atypical viruses but also found in bacteria, and (3) a large amount of mRNAs in *Deinococcus* are leaderless (de Groot *et al.*, 2009, 2014; Baudet *et al.*, 2010; Bouthier de la Tour *et al.*, 2015). Figure 20.2 shows the specific MS/MS spectrum corresponding to the SQEIWADVLGYVR peptide corresponding to the N-terminus of the DnaA protein, involved in the initiation of chromosome replication. This peptide labeled with a specific reagent allows establishing that an ATC codon is the true translation initiation codon of the corresponding gene. This specific and unusual start is conserved in other *Deinococcus* genomes, namely, *Deinococcus radiodurans* and *Deinococcus geothermalis*.

20.3 Applications to Non-model Organisms: From Bacteria to Parasites

Besides the annotation or re-annotation of protein-coding nucleic acid sequences based on the empirical observation of their gene products, proteogenomics has expanded and now includes the use of draft genome sequences and RNA-seq nucleotide sequences to generate a six-frame database for proteomics data interpretation (Armengaud *et al.*, 2014). In this case, the objective per se is not genome annotation but rather proteomics interpretation to highlight the most interesting molecular features. These are then functionally annotated and subjected to further characterization or validation once discovered. For example, the newly isolated alpha-proteobacterium *Tistlia consotensis* has been subjected to a proteogenomics analysis in order to understand its halotolerance properties (Rubiano-Labrador *et al.*, 2014, 2015). Figure 20.3 shows the strategy to highlight the main molecular players by means of differential proteogenomics. First, its draft genome was established and obtained as a set of 2377 contigs totaling 5.7 Mb. These contigs were translated into all the possible ORFs for generating an extensive




```
>gi|226354810:30-1421 Deinococcus deserti VCD115, complete genome
AGTTCAAGGGAGGTGCGCTTTATCTCGCAGGAAATCTGGGCGGACGTGCTGGGTTACGTCCGCAAAAACATCTCGGAAGTG
1                M  R  F  I  S  Q  E  I  W  A  D  V  L  G  Y  V  R  K  N  I  S  E  V
2                   M  S  Q  E  I  W  A  D  V  L  G  Y  V  R  K  N  I  S  E  V

>gi|94984109:222-1634 Deinococcus geothermalis DSM11300, complete genome
ATGAAGGGGGAGGTGCGCGTTATCACGCAGGAAATCTGGGCGGACGTGCTGGGGTACGTCCGCAAAAATATCTCGGAAGTG
1  M  K  G  E  V  R  V  I  T  Q  E  I  W  A  D  V  L  G  Y  V  R  K  N  I  S  E  V
2                   M  S  Q  E  I  W  A  D  V  L  G  Y  V  R  K  N  I  S  E  V

>gi|11612676:3268-1904 Deinococcus radiodurans R1, complete genome
GAACGGGGGAGGGGGAGCACTATCTCACAGGAAATCTGGGCCGACGTCCTCGCTTACGTGCGCAAAAACGTCTCCGACTTG
1                                                   M  R  K  N  V  S  D  L
2                   M  S  Q  E  I  W  A  D  V  L  A  Y  V  R  K  N  V  S  D  L
```

Figure 20.2 Re-annotation of the translation initiation codon for the *dnaA* gene in *Deinococcus deserti* genome. The MS/MS spectrum corresponding to the SQEIWADVLGYVR peptide labeled with the TMPP reagent has been recorded with an LTQ-Orbitrap XL instrument (Thermo Scientific). This spectrum enables this peptide to be established as the N-terminus of the matured protein. It can be explained by a translation start at the ATC initiation codon highlighted in bold red in the genome sequence. The corresponding polypeptide sequence and the previous annotations are shown in red and black, respectively, for three different *Deinococcus* genomes.

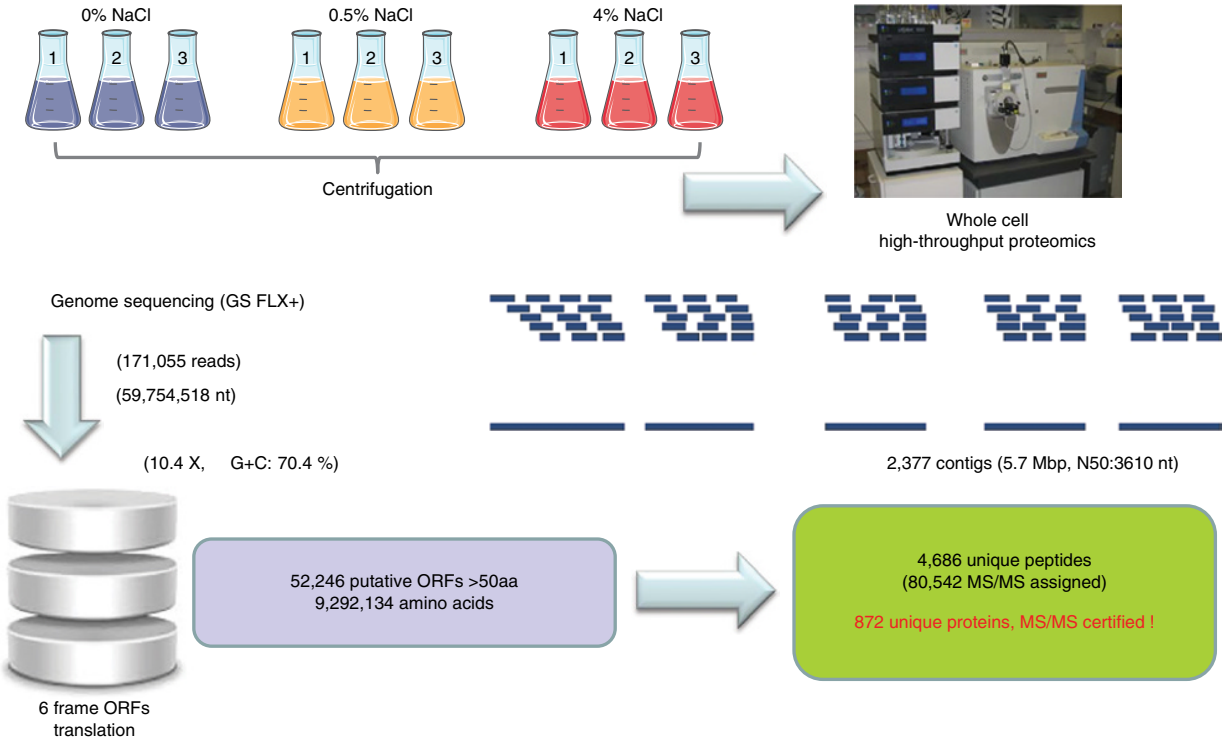


Figure 20.3 **Proteogenomics strategy using a draft genome sequence applied to the *Tistlia consotensis* bacterium.** *Tistlia consotensis* cells were grown in three different media differing in sodium chloride concentration. For each condition, three biological replicates were carried out. Cells were collected by centrifugation and subjected to whole-cell shotgun proteomics. To interpret the tandem mass spectrometry results, the genome was sequenced, resulting in a draft genome comprising 2377 contigs. These were translated in the six possible ways to establish a comprehensive database of 52,246 open reading frames (ORFs). The interpretation resulted in the identification of 4686 unique peptides and the abundance comparison of 872 proteins.

database, comprising 52,246 putative polypeptide sequences. Proteomics data were collected and then interpreted against this database, giving quantitative insights into 4686 unique peptide sequences and 872 MS/MS certified proteins. The comparison of protein quantities in different physiological conditions was straightforward with such a approach. The proteogenomics strategy was used to study the amphipod *Gammarus fossarum* and is illustrative of the application of proteogenomics on more complex non-model organisms (Trapp *et al.*, 2014, 2015c). This crustacean is used as the sentinel species to assess the quality of freshwater ecosystems, despite the absence of genome sequence data. A polypeptide sequence database was first generated on the basis of RNA-seq information. This allowed the interpretation of several proteomics datasets aimed at discovering the proteins involved in reproductive processes or proteins that could be interesting biomarkers of intoxication in the future. Further studies on related organisms took advantage of the same RNAseq-derived polypeptide sequence database (Trapp *et al.*, 2015a, 2015b). Molecular characterization of pathogens by means of proteogenomics strategies has been illustrated in several studies: for example, the enterobacteria, *Yersinia pestis* (Payne *et al.*, 2010; Schrimpe-Rutledge *et al.,* 2012) and *Shigella flexneri* (Zhao *et al.*, 2011), the firmicutes, *Bacillus anthracis* and *Streptococcus pyogenes* (Venter *et al.*, 2011), the epsilon proteobacterium, *Helicobacter pylori* (Muller *et al.*, 2013), the fungi *Cryptococcus neoformans* (Nagarajha *et al.*, 2014) and *Candida glabrata* (Prasad *et al.*, 2012), the complex protozoan parasites *Leishmania donovani* (Jamdhade *et al.*, 2015), *Toxoplasma gondii*, and *Neospora caninum* (Krishna *et al.*, 2015). This list continues to expand and shows the potential of proteogenomics in the field of infectious diseases.

20.4 Embracing Complexity with Metaproteogenomics

The proteogenomic methodology has been so refined in the last five years and tandem mass spectrometry throughput has increased so much with the advent of a new generation of mass analyzers adapted to biological molecules, that even complex mixtures of microorganisms can be analyzed (Armengaud, 2013). Indeed, meta-proteomics has already shown its great potential for analyzing human microbiomes and gaining new insights into their effects on Health (Jagtap *et al.*, 2015a; see Chapter 19, this book). For example, Young *et al.* (2015) have shown the different stages of microbial colonization of the gut of a preterm infant. They revealed the functional shifts between the observable proteins originating from the evolving microbiota and those from human epithelial cells. A work on oral biofilms from healthy and caries-bearing individuals has shown the abundance of actinobacteria (*Actinomyces*, *Corynebacterium*, *Rothia*) and firmicutes (*Streptococcus*) in such microbiomes, as well as the specificities of the protein repertoire of human dental plaque (Belda-Ferre *et al.*, 2015). A recent trend is emerging with metaproteogenomics carried out in order to get the most of both metagenomics and metaproteomics data acquired on the same sample. Although restricted until now to environmental samples, this methodology will have soon greatly impact the medical field. A sulfate-reducing enriched culture with m-xylene as the sole source of carbon and energy was subjected to metaproteogenomics in order to reveal the key enzymes of a *Desulfobacteraceae* representative involved in organic compound mineralization (Bozinovski *et al.*, 2014). The long-term adaptation of bacterial communities found in

metal-contaminated sediments from freshwater polluted sites revealed the prominent role of beta-proteobacteria (Gillan *et al.* 2015). Although efforts have been made to improve the bioinformatics tools for analyzing large metaproteogenomics datasets, annotating genomic sequences in terms of function with proteomics data and connecting these annotations to metabolic functions in microbiomes is still a difficult and time-consuming task (Seifert *et al.*, 2013; Jagtap, 2015b). These high-throughput approaches rely on genome information, protein sequence data, and taxonomic characteristics that are available from public repositories. Unfortunately, the quality of these databases is currently still not reliable (Pible *et al.*, 2014), with many errors and cross-contamination being evident. As recently discussed (Pible and Armengaud, 2015), their quality should be improved, which would leapfrog proteogenomics into the meta-omics 2.0 era and help move it to the frontline of diagnostics.

References

Armengaud, J. (2009). A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr. Opin. Microbiol.*, 12: 292–300.

Armengaud, J. (2010). Proteogenomics and systems biology: Quest for the ultimate missing parts. *Expert Rev. Proteomics*, 7: 65–77.

Armengaud, J. (2013). Microbiology and proteomics, getting the best of both worlds! *Environ. Microbiol.*, 15: 12–23.

Armengaud, J., Hartmann, E. M., and Bland, C. (2013). Proteogenomics for environmental microbiology. *Proteomics*, 13: 2731–2742.

Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., and Hartmann, E. M. (2014). Non-model organisms, a species endangered by proteogenomics. *J. Proteomics*, 105: 5–18.

Baudet, M., Ortet, P., Gaillard, J. C., Fernandez, B., Guerin, P., Enjalbal, C., Subra, G., de Groot, A., Barakat, M., Dedieu, A., and Armengaud, J. (2010). Proteomics-based refinement of *Deinococcus deserti* genome annotation reveals an unwonted use of non-canonical translation initiation codons. *Mol. Cell. Proteomics*, 9: 415–426.

Belda-Ferre, P., Williamson, J., Simon-Soro, A., Artacho, A., Jensen, O. N., and Mira, A. (2015). The human oral metaproteome reveals potential biomarkers for caries disease. *Proteomics*, 15: 3497–3507.

Bland, C., Bellanger, L., and Armengaud, J. (2014a). Magnetic immunoaffinity enrichment for selective capture and MS/MS analysis of N-terminal-TMPP-labeled peptides. *J. Proteome Res.*, 13: 668–680.

Bland, C., Hartmann, E. M., Christie-Oleza, J. A., Fernandez, B., and Armengaud, J. (2014b). N-Terminal-oriented proteogenomics of the marine bacterium *Roseobacter denitrificans* Och114 using N-Succinimidyloxycarbonylmethyl)tris(2,4,6-trimethoxyphenyl)phosphonium bromide (TMPP) labeling and diagonal chromatography. *Mol. Cell. Proteomics*, 13: 1369–1381.

Bouthier de la Tour, C., Blanchard, L., Dulermo, R., Ludanyi, M., Devigne, A., Armengaud, J., Sommer, S., and de Groot, A. (2015). The abundant and essential HU proteins in *Deinococcus deserti* and *Deinococcus radiodurans* are translated from leaderless mRNA. *Microbiology*, in press.

Bozinovski, D., Taubert, M., Kleinsteuber, S., Richnow, H.H., von Bergen, M., Vogt, C., and Seifert, J. (2014). Metaproteogenomic analysis of a sulfate-reducing enrichment culture

reveals genomic organization of key enzymes in the m-xylene degradation pathway and metabolic activity of proteobacteria. *Syst. Appl. Microbiol.*, 37: 488–501.

de Groot, A., Dulermo, R., Ortet, P., Blanchard, L., Guerin, P., Fernandez, B., Vacherie, B., Dossat, C., Jolivet, E., Siguier, P., Chandler, M., Barakat, M., Dedieu, A., Barbe, V., Heulin, T., Sommer, S., Achouak, W., and Armengaud, J. (2009). Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. *PLoS Genet.*, 5: e1000434.

de Groot, A., Roche, D., Fernandez, B., Ludanyi, M., Cruveiller, S., Pignol, D., Vallenet, D., Armengaud, J., and Blanchard, L. (2014). RNA sequencing and proteogenomics reveal the importance of leaderless mRNAs in the radiation-tolerant bacterium *Deinococcus deserti*. *Genome Biol. Evol.*, 6: 932–948.

Dworzanski, J. P., Dickinson, D. N., Deshpande, S. V., Snyder, A. P., and Eckenrode, B. A. (2010). Discrimination and phylogenomic classification of *Bacillus anthracis-cereus-thuringiensis* strains based on LC-MS/MS analysis of whole cell protein digests. *Anal. Chem.*, 82: 145–155.

Gillan, D. C., Roosa, S., Kunath, B., Billon, G., and Wattiez, R. (2015). The long-term adaptation of bacterial communities in metal-contaminated sediments: A metaproteogenomic study. *Environ. Microbiol.*, 17: 1991–2005.

Gupta, N., Benhamida, J., Bhargava, V., Goodman, D., Kain, E., Kerman, I. Nguyen, N., Ollikainen, N., Rodriguez, J., Wang, J., Lipton, M. S., Romine, M., Bafna, V., Smith, R. D., and Pevzner, P. A. (2008). Comparative proteogenomics: Combining mass spectrometry and comparative genomics to analyze multiple genomes. *Genome Res.*, 18: 1133–1142.

Hartmann, E. M., and Armengaud, J. (2014). N-terminomics and proteogenomics, getting off to a good start. *Proteomics*, 14: 2637–2646.

Hernandez, C., Waridel, P., and Quadroni, M. (2014). Database construction and peptide identification strategies for proteogenomic studies on sequenced genomes. *Curr. Top. Med. Chem.*, 14: 425–434.

Jabbour, R. E., Deshpande, S. V., Wade, M. M., Stanford, M. F., Wick, C. H., Zulich, A. W., Skowronski, E. W., and Snyder, A. P. (2010). Double-blind characterization of non-genome-sequenced bacteria by mass spectrometry-based proteomics. *Appl. Environ. Microbiol.*, 76: 3637–3644.

Jagtap, P., Griffin, T., and Armengaud, J. (2015a). Microbiomes – Embracing complexity. *Proteomics*, 15: 3405–3406.

Jagtap, P. D., Blakely, A., Murray, K., Stewart, S., Kooren, J., Johnson, J.E., Rhodus, N. L., Rudney, J., and Griffin, T. J. (2015b). Metaproteomic analysis using the Galaxy framework. *Proteomics*, 15: 3553–3565.

Jamdhade, M. D., Pawar, H., Chavan, S., Sathe, G., Umasankar, P. K., Mahale, K. N., Dixit, T., Madugundu, A. K., Prasad, T. S., Gowda, H., Pandey, A., and Patole, M. S. (2015). Comprehensive proteomics analysis of glycosomes from *Leishmania donovani*. *OMICS*, 19: 157–170.

Karlsson, R., Davidson, M., Svensson-Stadler, L., Karlsson, A., Olesen, K., Carlsohn, E., and Moore, E. R. (2012). Strain-level typing and identification of bacteria using mass spectrometry-based proteomics. *J. Proteome Res.*, 11: 2710–2720.

Karlsson, R., Gonzales-Siles, L., Boulund, F., Svensson-Stadler, L., Skovbjerg, S., Karlsson, A., Davidson, M., Hulth, S., Kristiansson, E., and Moore, E. R. (2015). Proteotyping: Proteomic characterization, classification and identification of microorganisms – A prospectus. *Syst. Appl. Microbiol.*, 38: 246–257.

Krishna, R., Xia, D., Sanderson, S., Shanmugasundram, A., Vermont, S., Bernal, A., Daniel-Naguib, G., Ghali, F., Brunk, B. P., Roos, D. S., Wastling, J. M., and Jones, A. R. (2015). A large-scale proteogenomics study of apicomplexan pathogens-*Toxoplasma gondii* and *Neospora caninum*. *Proteomics*, 15: 2618–2628.

Kucharova, V., and Wiker, H. G. (2014). Proteogenomics in microbiology: Taking the right turn at the junction of genomics and proteomics. *Proteomics*, 14: 2360–2675.

Lavigne, J. P., Espinal, P., Dunyach-Remy, C., Messad, N., Pantel, A., and Sotto, A. (2013). Mass spectrometry: A revolution in clinical microbiology? *Clin. Chem. Lab. Med.*, 51: 257–270.

Muller, S. A., Findeiss, S., Pernitzsch, S. R., Wissenbach, D. K., Stadler, P. F., Hofacker, I. L., von Bergen, M., and Kalkhof, S. (2013). Identification of new protein coding sequences and signal peptidase cleavage sites of *Helicobacter pylori* strain 26695 by proteogenomics. *J. Proteomics*, 86: 27–42.

Nagarajha Selvan, L. D., Kaviyil, J. E., Nirujogi, R. S., Muthusamy, B., Puttamallesh, V. N., Subbannayya, T., Syed, N., Radhakrishnan, A., Kelkar, D. S., Ahmad, S., Pinto, S. M., Kumar, P., Madugundu, A. K., Nair, B., Chatterjee, A., Pandey, A., Ravikumar, R., Gowda, H., and Prasad, T. S. (2014). Proteogenomic analysis of pathogenic yeast *Cryptococcus neoformans* using high resolution mass spectrometry. *Clin. Proteomics*, 11: 5.

Nesvizhskii, A. I. (2014). Proteogenomics: Concepts, applications and computational strategies. *Nat. Methods*, 11: 1114–1125.

Payne, S. H., Huang, S. T., and Pieper, R. (2010). A proteogenomic update to *Yersinia*: enhancing genome annotation. *BMC Genomics*, 11: 460.

Pible, O., and Armengaud, J. (2015). Improving the quality of genome, protein sequence, and taxonomy databases: A prerequisite for microbiome meta-omics 2.0. *Proteomics*, 15: 3418–3423.

Pible, O., Hartmann, E. M., Imbert, G., and Armengaud, J. (2014). The importance of recognizing and reporting sequence database contamination for proteomics. *EuPA Open Proteomics*, 3: 246–249.

Prasad, T. S., Harsha, H. C., Keerthikumar, S., Sekhar, N. R., Selvan, L. D., Kumar, P. Pinto, S. M., Muthusamy, B., Subbannayya, Y., Renuse, S., Chaerkady, R., Mathur, P. P., Ravikumar, R., and Pandey, A. (2012). Proteogenomic analysis of *Candida glabrata* using high resolution mass spectrometry. *J. Proteome Res.*, 11: 247–260.

Renuse, S., Chaerkady, R., and Pandey, A. (2011). Proteogenomics. *Proteomics*, 11: 620–630.

Rubiano-Labrador, C., Bland, C., Miotello, G., Armengaud, J., and Baena, S. (2015). Salt stress induced changes in the exoproteome of the halotolerant bacterium *Tistlia consotensis* deciphered by proteogenomics. *PLoS ONE*, 10: e0135065.

Rubiano-Labrador, C., Bland, C., Miotello, G., Guerin, P., Pible, O., Baena, S., and Armengaud, J. (2014). Proteogenomic insights into salt tolerance by a Halotolerant alpha-proteobacterium isolated from an Andean saline spring. *J. Proteomics*, 97: 36–47.

Schrimpe-Rutledge, A. C., Jones, M. B., Chauhan, S., Purvine, S. O., Sanford, J. A., Monroe, M. E., Brewer, H. M., Payne, S. H., Ansong, C., Frank, B. C., Smith, R. D., Peterson, S. N., Motin, V. L., and Adkins, J. N. (2012). Comparative omics-driven genome annotation refinement: application across *Yersiniae*. *PLoS ONE*, 7: e33903.

Seifert, J., Herbst, F. A., Halkjaer Nielsen, P., Planes, F. J., Jehmlich, N., Ferrer, M., and von Bergen, M. (2013). Bioinformatic progress and applications in metaproteogenomics for bridging the gap between genomic sequences and metabolic functions in microbial communities. *Proteomics*, 13: 2786–2804.

Trapp, J., Geffard, O., Imbert, G., Gaillard, J. C., Davin, A.H., Chaumot, A., and Armengaud, J. (2014). Proteogenomics of *Gammarus fossarum* to document the reproductive system of amphipods. *Mol. Cell. Proteomics*, 13: 3612–3625.

Trapp, J., Almunia, C., Gaillard, J. C., Pible, O., Chaumot, A., Geffard, O., and Armengaud, J. (2015a). Proteogenomic insights into the core-proteome of female reproductive tissues from crustacean amphipods. *J. Proteomics*, in press.

Trapp, J., Almunia, C., Gaillard, J. C., Pible, O., Chaumot, A., Geffard, O., and Armengaud, J. (2015b). Data for comparative proteomics of ovaries from five non-model, crustacean amphipods. *Data Brief*, 5: 1–6.

Trapp, J., Armengaud, J., Pible, O., Gaillard, J. C., Abbaci, K., Habtoul, Y., Chaumot, A., and Geffard, O. (2015c). Proteomic investigation of male *Gammarus fossarum*, a freshwater crustacean, in response to endocrine disruptors. *J. Proteome Res.*, 14: 292–303.

Venter, E., Smith, R. D., and Payne, S. H. (2011). Proteogenomic analysis of bacteria and archaea: a 46 organism case study. *PLoS ONE*, 6: e27587.

Yamazaki, S., Yamazaki, J., Nishijima, K., Otsuka, R., Mise, M., Ishikawa, H., Sasaki, K., Tago, S., and Isono, K. (2006). Proteome analysis of an aerobic hyperthermophilic crenarchaeon, *Aeropyrum pernix* K1. *Mol. Cell. Proteomics*, 5: 811–823.

Young, J. C., Pan, C., Adams, R. M., Brooks, B., Banfield, J. F., Morowitz, M. J., and Hettich, R. L. (2015). Metaproteomics reveals functional shifts in microbial and human proteins during a preterm infant gut colonization case. *Proteomics*, 15: 3463–3473.

Zhao, L., Liu, L., Leng, W., Wei, C., and Jin, Q. (2011). A proteogenomic analysis of *Shigella flexneri* using 2D LC-MALDI TOF/TOF. *BMC Genomics*, 12: 528.

21A

Analysis of MALDI-TOF MS Spectra using the BioNumerics Software

Katleen Vranckx, Katrien De Bruyne and Bruno Pot

Applied Maths NV, Sint-Martens Latem, Belgium

21A.1 Introduction

Typing methods aim at distinguishing isolates within the same species. The earliest typing systems were purely phenotypic in nature (serotypes, phagetypes, biochemical characteristics, antibiograms). However, with easier access to DNA technology, more and more methods have been developed on a genetic and molecular level. The first method to be used routinely on a larger scale for typing of isolates relevant to public health was pulsed field gel electrophoresis (PFGE), a technology used for the separation of long DNA fragments obtained after the digestion of high-molecular-weight DNA by rare cutting restriction enzymes. This technique was developed in 1984 by Schwartz and Cantor (Schwartz and Cantor, 1984). The comparison of profiles between different gels was difficult and very often subjective. In the early 1990s, the software GelCompar was developed and widely used to support a more objective computer-aided normalization and comparison of electrophoresis profiles, obtained from DNA or proteins (Vauterin and Vauterin, 1992). Later this software evolved to GelCompar II and BioNumerics. The increased level of standardization allowed the integration of PFGE technology in the public health arena and following the PulseNet initiative of the American CDC in 1996 (Swaminathan, Barrett, Hunter, Tauxe, and CDC Pulsenet Task Force, 2001), the technology started to be used almost worldwide (http://www.pulsenetinternational.org/protocols).

Many more typing methods have since been developed, including among others AFLP (amplified fragment length polymorphism), RAPD (rapid amplification of polymorphic DNA), ERIC or REP- PCR (enterobacterial repetitive intergenic consensus or repetitive extragenic palindromic sequence polymerase chain reaction), MLVA or VNTR (multilocus variant analysis or variable number tandem repeat analysis), MLST or MLSA (multilocus sequence typing or multilocus sequence analysis), and more recently whole genome sequencing (WGS). The choice of the technique to be used will depend on the desired speed, cost, reproducibility, available infrastructure, and

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

resolution. A detailed discussion of these techniques is beyond the scope of this chapter, though, and can be found in Sabat *et al.* (2013).

The above-mentioned software BioNumerics supports polyphasic analysis of all these typing methods. It consists of 10 compatible software modules, which each add specific functionality to the software suit. Five data modules deal with the import, processing, and storage of biological data types (one-dimensional fingerprints; binary, numerical or categorical data; Sanger or NGS sequences; trend curves, and whole genome restriction maps). An additional four modules offer analysis possibilities for the creation of trees and dendrograms, statistical outputs and visualizations, identification projects and WGS comparison, and annotation projects. There is also an audit trail module for recording and versioning of database activities, as required by the FDA 21 CFR Part 11 regulation. Individually installable plugins provide additional functionality to the software and assist in setting up the correct conditions for typing efforts that require consultation of for example, external database sources (spa-typing, MIRU-VNTR, MLST online).

Modules can be combined in order to allow for the import, storage, and analysis of experimental and metadata from different sources and experiments. The underlying relational database facilitates the discovery of correlations between metadata and experimental data or between phenotype and genotype. By using a well-selected set of data, new typing methods can easily be compared with golden standard methods for validation or other purposes.

21A.2 Typing with MALDI-TOF MS

The use of MALDI-TOF MS as a typing method is becoming more and more common in clinical microbiology. Currently, identification of an isolate to the species level is the most prevalent application in for example, a hospital or reference center, but there have been reports that discrimination at the subspecies level or below is also within reach (Spinali *et al.*, 2014).

The analysis of MALDI-TOF MS spectra has also been integrated in the BioNumerics software. Depending on the data output of the hardware used, BioNumerics accepts both raw and processed data. In this chapter, we will further discuss the different algorithms available for processing the raw spectra. We will discuss how to handle technical and biological replicates and illustrate the statistical analysis of the spectra using cluster analysis and peak mining approaches. The analysis therefore includes both the comparison of different profiles as well as the detection of biomarkers. It is not within the scope of this chapter to discuss in detail the mathematical background of the algorithms. The aim is to aid the biologist in understanding the concepts of the algorithms so that an informed decision on an analysis workflow can be made.

21A.3 Preprocessing of Raw MALDI-TOF MS Data

Preprocessing of raw data from a mass spectrometer is necessary in order to ensure that the profiles are comparable. Preprocessing aims at removing differences between profiles that are the result of technical variation, while preserving the true biological variation. For a more mathematical background on the methods described in this session, we refer the reader to Monchamp *et al.* (2007) and Yang *et al.* (2009).

21A.4 Downsampling

In general, the first step in preprocessing is the resampling (most often downsampling) of the spectra, making it possible to use more computationally intensive algorithms later on. Rescaling is also useful to create a constant m/z scale, which makes it possible to more easily compare MALDI-TOF MS profiles among one another. The profiles used for typing or identification purposes typically have a higher sampling rate than the peak resolution of the mass spectrometer. Therefore, a resampling is generally a downsampling, which will reduce the size of the profile, but will retain peak information (Figure 21A.1). Although the resampled profile is constructed with much fewer points than the original profile, both the position and the height of the peaks are maintained and are well within the normal variation on the measurement. Clearly, a too aggressive downsampling procedure will turn high-frequency signals into low-frequency events transforming multiple peaks into one single large peak after resampling (Figure 21A.2).

In BioNumerics, two resampling methods are available: *linear* and *fitted*. Using linear resampling, the user can define a fixed distance between the resampled points. The fitted resampling will automatically fit the resampled curve with the original curve, using a quadratic or linear function to determine the distance between the resampled points in such a way that the distance is as high as possible, while retaining the shape of the

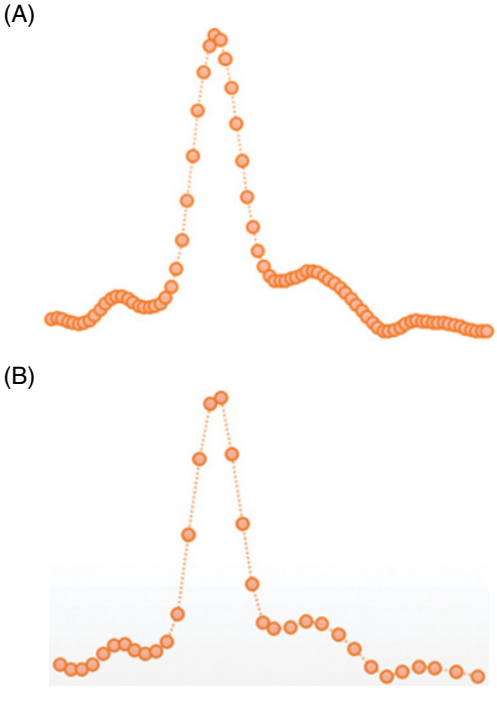


Figure 21A.1 Example of effect of downsampling on part of a spectral profile. The original profile with all sampled points is seen in (A). After downsampling, fewer points remain (B) while maintaining the shape of the peak.

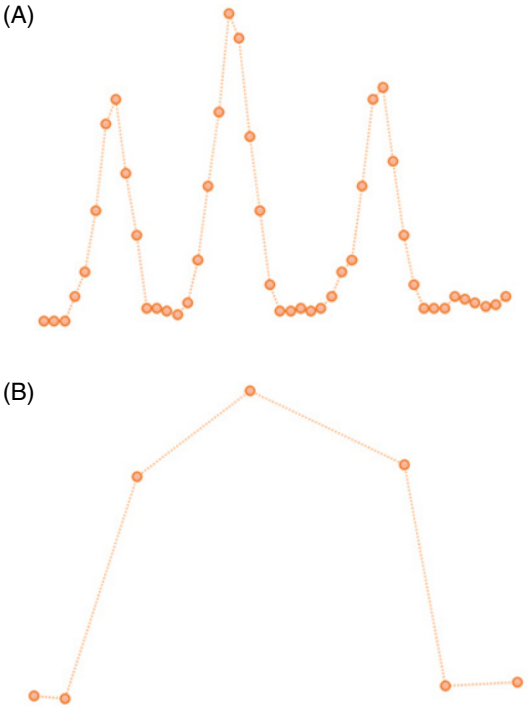


Figure 21A.2 Demonstration of too aggressive downsampling setting. The original profiles in (A) contains three peaks. After downsampling, the profile in (B) only contains one peak.

original curve. Because the m/z value is determined on the mass spectrometer as a quadratic function of the time of flight, the resolution of the mass spectrometer will decrease with increasing m/z. Therefore, the distance between resampled points can also be increased with increasing m/z. A quadratic increase is most suitable as it captures the quadratic nature of the m/z output of the time-of-flight detection. In the pipeline applied to the examples in this chapter, we chose to use a fitted resampling using a quadratic function.

21A.5 Baseline Subtraction

Once the data have been resampled, they are ready for the next step in preprocessing. Spectral profiles very often have a varying baseline, caused by technical variation during the experiment. Typical causes are differences in laser intensity and minor variations in the composition of the matrix. It is advisable to remove this background before analysis in order to ensure that the spectral profiles are quantitatively comparable, though it is not necessary for all analysis methods. During baseline processing, the baseline is determined and subtracted from the entire spectral profile. The baseline is considered to be the signal created by an empty sample (containing only the matrix). Unfortunately, this signal cannot be measured on the same sample spot, and the results obtained from a true empty sample cannot simply be extrapolated to the rest of the run. Therefore, a baseline needs to be estimated or calculated for each profile separately. Again, there are

different methods to calculate the most likely baseline. The results of these different methods are demonstrated in Figure 21A.3.

In the *Binned Baseline*, the m/z axis is divided into a number of bins. The distribution of the intensity is evaluated for each bin. If this does not meet certain statistical criteria (mean value, kurtosis, and skewness), the bin is discarded. Bins that fail these criteria are typically the bins that contain a peak or part of a peak. For bins that are retained, the minimum intensity value is placed at the center of the bin, and these points are then connected to give the complete baseline.

The *monotone minimum* baseline algorithm will consider the intensity at the lowest m/z point. It starts at the lowest m/z point; the intensity at this point is the first minimum. If the intensity at the next m/z point is lower, then at this point, the lower value is used. If it is higher, the previous intensity is considered the baseline. As a result, each time the intensity increases, the previous minimum intensity encountered in the spectra is used as the baseline.

The *moving bar* method is actually a moving local minimum method. It works like a horizontal bar that is moving up from the lower intensities of the spectrum, and moving as high as possible, until it touches the profile. The intensity at this position is considered to be the baseline at the middle of the bar.

The *rolling disk* algorithm works conceptually as though one is rolling a disk at the basis of the profile, retaining the intensity value at which the highest point of the disk touches the profile.

Although the first two methods are relatively simple, they do not always result in a satisfactory baseline subtraction. As seen with the examples in Figure 21A.3, the base of the peaks is still at a substantially higher intensity compared to the results of the moving bar or rolling disk methods. Furthermore, even though the binned baseline is a simple algorithm, there are quite some parameters that can be adapted and that are not always easy to interpret. The moving bar and rolling disk have the benefit of simplicity and are considerably better at pulling the base of the peak down to an intensity of zero. The only input they require is the size of the bar and the diameter of the disk, respectively. In most cases, the appropriate parameters can easily be deducted from the spectral profile itself. A good choice for a satisfactory baseline is to use twice the width of an average peak at its base as the width of the moving bar, or as the diagonal of the rolling disk.

21A.6 Curve Smoothing

Spectral profiles may contain many small spikes that do not correspond with a biological signal, but should be considered noise. A smoothing algorithm can reduce or eliminate this noise, making it easier to detect real peaks afterward. Examples of the effect of the different algorithms discussed here are shown in Figure 21A.4.

The *Savitzky–Golay* smoothing algorithm is based on a least square filtering. The smoothing is performed by fitting a local polynomial regression to a set of consecutive points of a user-defined sliding window, in order to determine the smoothed value at the center of the window. This smoothing preserves the shape and height of the peaks, but is not always efficient at removing all noise (Schafer, 2011). The size of the moving window is a critical parameter. When it is too small, not all noise may be removed, and when it is too big, neighboring peaks risk being merged.

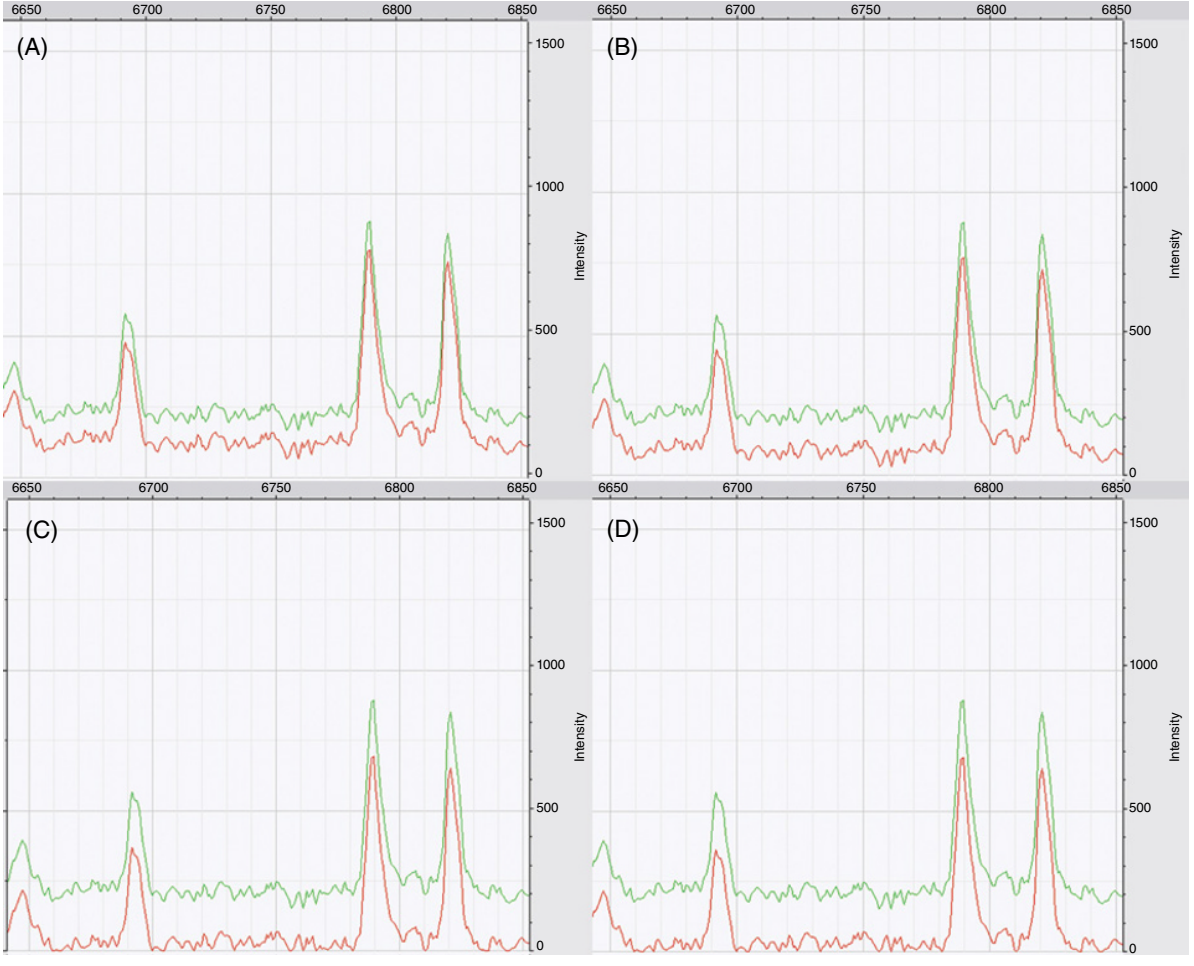


Figure 21A.3 The result of different baseline subtraction algorithms: (A) the binned baseline, (B) the monotone minimum, (C) the moving bar, and (D) the rolling disk. The top green curve is the raw profile, and the bottom red curve is after background subtraction.

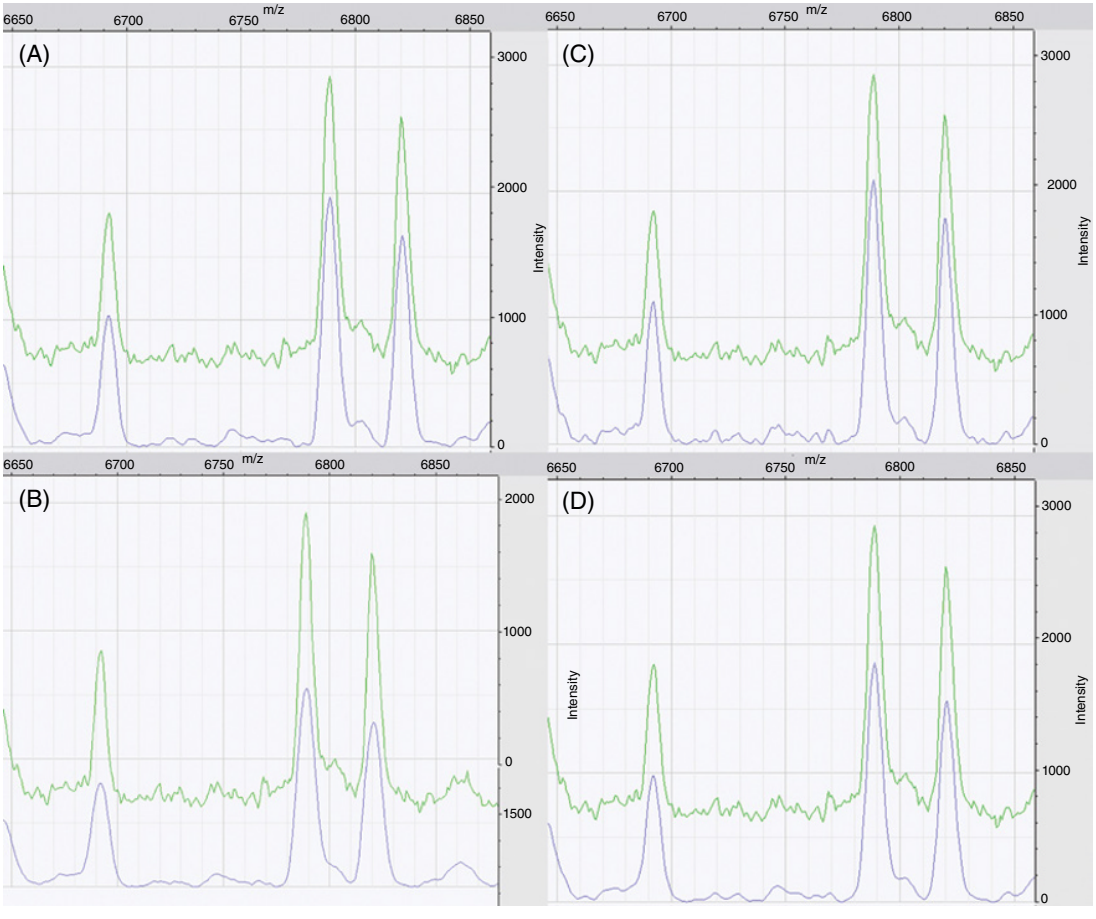


Figure 21A.4 Result of the different curve smoothing algorithms: (A) Savitsky–Golay, (B) Gaussian, (C) moving average, and (D) Kaiser. All algorithms were applied to the profile after background subtraction with a moving disk algorithm. The top green profile is the raw profile, and the bottom blue is the smoothed profile.

A very similar smoothing algorithm is the *moving average* algorithm. Here, each y-value is replaced by a local average value. The size of the window to be used for determining the average value can be adjusted. A larger window size results in more aggressive smoothing and vice versa. This smoothing does not entirely retain the shape of the peak.

Smoothing can also be performed by fitting a *Gaussian curve* through each point. This typically results in a more aggressive smoothing where the shape of the original peaks is not guaranteed.

A filter that is very popular in signal processing is the *Kaiser window*. It is a powerful smoothing algorithm that is very efficient with regard to calculation effort and smoothing result and retains the shape of the original peaks very well. Moreover, by adjusting the parameters, the aggressiveness of the smoothing can easily be varied (Kaiser, 1966).

With this wealth of algorithms the choice of the "best" algorithm to be used becomes difficult. When deciding, it might be a good idea to consider the algorithm for peak detection that will be used in the next step in the analysis of the MALDI-TOF MS spectra (see below). For some peak detection algorithms, the retention of the shape of the peak is important, whereas for others it is not. These considerations might clearly influence the decision of which smoothing algorithm to use. Also, it is good practice to visually inspect the impact of a chosen set of parameters and algorithms on the resulting spectra. The BioNumerics software allows you to change the analysis pipeline and visualize the result after each preprocessing step.

21A.7 Peak Detection

A very robust peak finding algorithm is the *local maxima* algorithm, which marks maxima in a given local window as peaks. When this algorithm is used, preliminary filtering is essential, as noise peaks also tend to give rise to local maxima, as well as imperfections in the shape of a peak. Smoothing algorithms that do not completely retain the shape of a peak are therefore more suitable to use before this peak finding algorithm is applied.

The *continuous wavelet transform* (CWT) is a complex algorithm that performs very well for the detection of relevant peaks. The input spectrum is convolved with a window function for a number of different window sizes. This means that we check for peaks of various widths, from small to large. Real peaks will typically fit both small and large peak templates. Only peaks that are present for a number of widths are retained, thus removing noise. A signal-to-noise threshold is also applied, using the CWT to calculate the noise before smoothing. If only one window size is used, with a signal-to-noise threshold, the algorithms reduce to a single CWT. This algorithm performs best with smoothed spectra for which the shape of the peaks is retained. For a more in-depth understanding of this algorithm, we refer to the specialized literature (Du, Kibbe, and Lin, 2006).

21A.8 Biological and Technical Replicates

There is a rich literature available on the mathematics of the above algorithms that goes into great detail on the different performances and specificities of the different approaches. In most cases, one is interested in detecting peaks that represent a true

biological signal, for example, a protein or peptide expressed by the isolate. A good way to check this is to run different technical replicates of the same strain. The term *technical replicate* is used here to signify spectra spotted on the same MALDI plate and originating from a single colony. The term *biological replicate* is used for repeated experiments in which the organism is grown repeatedly (or independently).

The set of technical replicates used for the comparison presented here, contained ten isolates from five different species (two gram-negative, two gram-positive, and one fungus); for each isolate, ten technical replicates were available. The spectra were obtained using complete protein extracts from colonies measured between 2,000 and 20,000 m/z. We choose two preprocessing workflows for comparison (see Table 21A.1 and Figure 21A.5).

A good preprocessing workflow should generate highly similar spectra for technical replicates, preferably with the same peaks detected in all spectra. We therefore compared several parameters of the technical replicates for both workflows 1 and 2. All parameters were compared between workflow 1 and 2 using a paired t-test with a significance level set at 0.05.

The overall similarity between spectra from technical replicates was determined on the complete spectrum after background subtraction and smoothing. For this purpose, an average spectrum was created, and the similarity between each replicate and the average was calculated using the Pearson correlation coefficient. The Pearson correlation of the replicates was significantly higher for workflow 1 compared to workflow 2. The standard deviation on this correlation was lower for workflow 1, albeit not significant.

The reproducibility of the peak detection was determined by calculating the peak detection rate (PDR) for each peak detected in one of the replicates, as well as the proportion of peaks with a PDR of 100%. Both these parameters were significantly higher for workflow 1.

We therefore conclude that the more advanced algorithms indeed give better results, and workflow 1 is more suitable than workflow 2 to eliminate technical variation in the profiles. Workflow 1 was therefore used throughout this chapter and has been implemented as default workflow in BioNumerics. For different types of spectra, other workflows may be more suitable.

Table 21A.1 Description of the two workflows used for comparative data analysis of the different preprocessing algorithms.

Preprocessing step	Workflow 1	Workflow 2
Background subtraction	Rolling disk	Rolling disk
Noise computation	CWT noise	MAD noise
Smoothing	Kaiser window	Moving average
Peak detection	CWT with S/N threshold 1	Local maxima
Peak filtering	Relative intensity 1% Double peaks	S/N threshold 1 Relative intensity 1% Double peaks

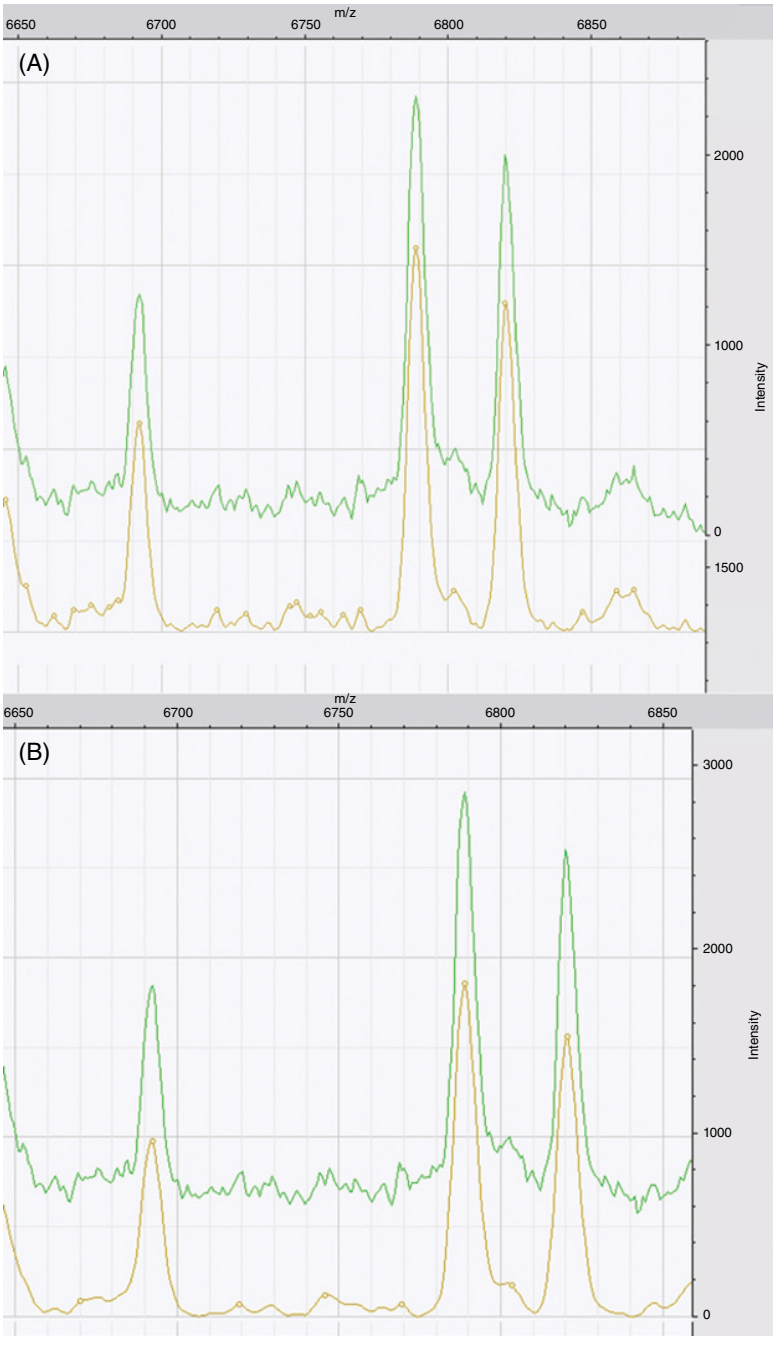


Figure 21A.5 The results of two different preprocessing workflows: (A) a moving average followed by peak detection with local maxima and (B) a Kaiser window smoothing followed by a CWT peak analysis. The top green profile is the original profile, and the bottom orange is the smoothed profile. Small circles on the bottom curve indicate the presence of a peak.

In order to further improve the quality of the analysis and gain confidence in the relevance of, for example, discriminating peaks, spectra can be determined repeatedly, including the process of repeated culturing. This may be important as MALDI-TOF MS analysis is a phenotypic technology that inevitably will yield results that depend on the interactions with the environment (growth medium, growth time, growth temperature, etc.).

21A.9 Averaging of Replicates

If replicates are available, technical or biological, they can be used to further reduce variation by creating an average spectrum of the replicates that will only retain the peaks with a sufficiently high peak detection ratio. Additional filters can thus be applied, depending on the number of replicates available.

From three or more replicates on, a similarity filter can be added that will compare the full spectral profile of the replicates with the calculated average profile, using a Pearson correlation coefficient. Preferably, the profiles should have correlations higher than 95% for technical replicates and 90 % for biological replicates. Lower numbers could indicate a problem with the profile, such as contamination of the matrix or the sample or differences in matrix composition or in sample preparation protocols. When applying a similarity filter, profiles that not reach the set threshold will be removed from the final average (Figure 21A.6).

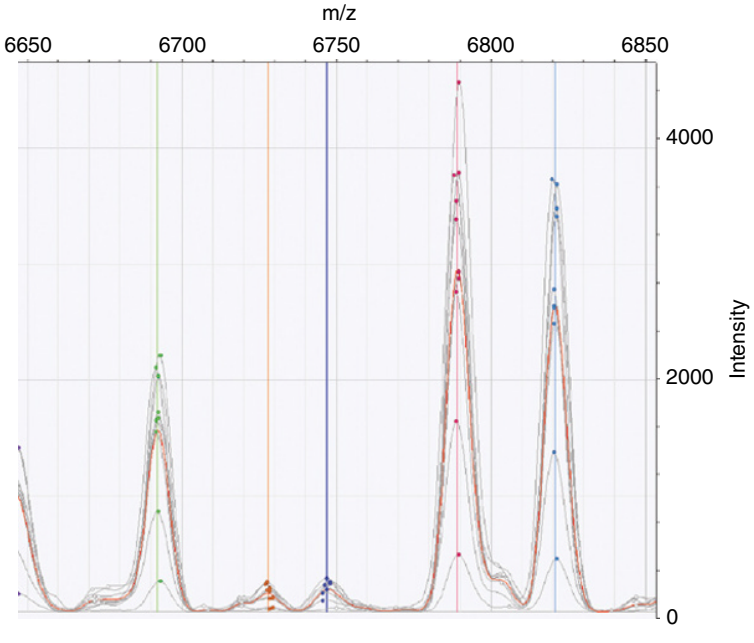


Figure 21A.6 Average spectral profile (red) derived from the member profiles (gray). Peaks are only retained if they are detected in the majority of members. Adjacent peaks are marked with different colored lines.

Other filters that can be applied are the minimal intensity filter that will remove profiles with an overall low intensity or filters on the minimum or maximum number of peaks detected. In a low-intensity profile, the signal-to-noise ratio of the peaks is lower than in comparable spectra with normal intensity. As a result, many of the smaller peaks will not reach the signal-to-noise threshold, even though they are reliably found in biological replicates of higher intensity and thus can be considered a true biological signal.

21A.10 Spectrum Analysis

Spectral profiles can yield very complicated signals, typically with 100–300 peaks per isolate. This can complicate the comparison of profiles, especially as many, also non-relevant factors can and will influence the absence/presence of a peak and its amplitude. Depending on the underlying question, unsupervised and supervised learning methods can be used to analyze MALDI-TOF MS data. Unsupervised methods, such as a hierarchical clustering or principal components analysis (PCA), require no prior knowledge on the dataset used. Supervised methods, such as linear discriminant analysis (LDA) and classifying algorithms, are based on prior knowledge of some characteristics of the dataset.

21A.11 Hierarchical Clustering

A cluster analysis will group highly similar profiles together, making it easier to detect and delineate clusters of highly related strains. In a first step, a similarity coefficient is calculated for each pair of spectra, depicted in a similarity matrix. The latter, although it contains very valuable information, is not organized, and clusters cannot be easily detected. Therefore, in the second step, the pairwise similarity matrix will be submitted to a clustering algorithm, resulting in reorganization of the similarity matrix and the presentation of a dendrogram that will group highly related profiles. As both steps are independent of each other, we will discuss them separately. In this chapter, we discuss only the characteristics of the coefficients and the possible consequences this can have on an analysis. For the exact formulas of the coefficients, we refer the reader to the specialized literature (Press, Teukolsky, Vetterling, and Flannery, 2007).

To calculate a pairwise similarity of MALDI-TOF MS profiles, there are two major approaches: curve-based and peak-based. The latter approach only takes into account the presence or absence of a peak detected during preprocessing. In a pairwise comparison, typically X out of Y peaks will match, resulting in a lower or higher similarity level. With a curve-based approach, the complete spectral profile of the isolates will be used for similarity estimation.

The curve-based similarity coefficients implemented in BioNumerics are the Pearson coefficient, the cosine coefficient, and the ranked Pearson coefficient (Vauterin and Vauterin, 2006). The Pearson and ranked Pearson have characteristics that make them very suitable for typing purposes. As the complete profile is compared, the intensity of the peaks is used to obtain the similarity between the two profiles. As the Pearson coefficient contains an average intensity correction, it is insensitive to global differences in background and intensity (see Figure 21A.7). All three profiles in Figure 21A.7(A) have

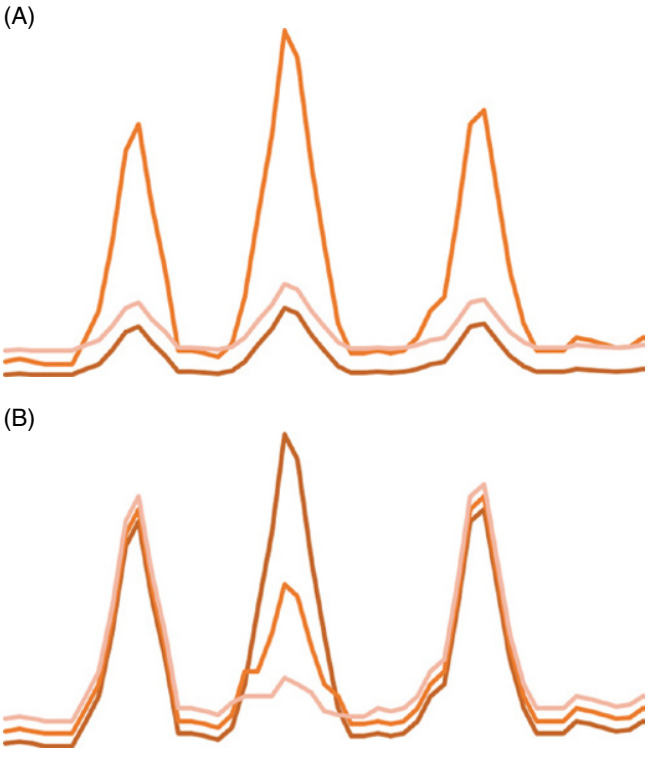


Figure 21A.7 Examples of profiles yielding (A) high and (B) very low pairwise similarity values using the Pearson correlation coefficient while containing the same peaks.

very great similarities with the Pearson correlation coefficient. This means it is unnecessary to perform an intensity-based normalization before calculating the coefficient. Pearson correlation is, however, very sensitive to local differences in intensity. The profiles in Figure 21A.7(B) contain the same peaks, but the middle peak is present at different intensities. This has a considerable effect on the similarity calculated with the Pearson correlation, and these three profiles will share a much lower similarity than the profiles in Figure 21A.7(A).

The cosine correlation coefficient is sensitive to differences in background and overall intensity and is therefore less suitable for our purposes. It should only be used on background-subtracted and intensity-normalized spectra. On these spectra, it has the same characteristics as the Pearson correlation coefficient.

The ranked Pearson correlation coefficient is calculated using the same formula as the Pearson correlation coefficient, but it does not use the absolute intensity value, but the rank of the intensity. As a result, the high-intensity peaks have a lower impact when using a ranked Pearson coefficient as compared to a regular Pearson correlation, whereas the low peaks have a higher impact. It might therefore be an interesting coefficient to use on profiles of the same species, as the high-intensity peaks correspond mainly to ribosomal proteins, which show little variation within a species. By increasing the importance of the lower peaks, we therefore increase the signals that differentiate strains within a species.

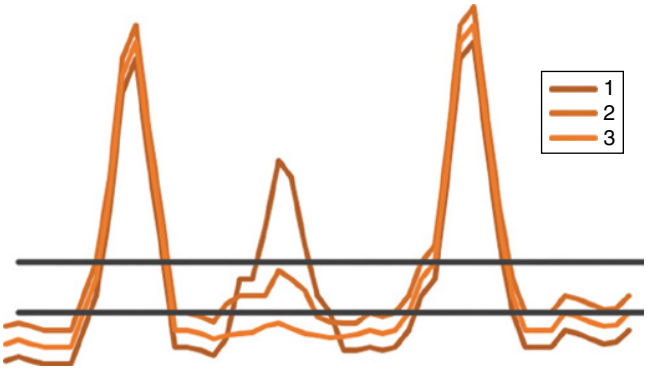


Figure 21A.8 Examples of profiles which are affected by a different choice of intensity threshold for peak detection when using a binary coefficient and which also score differently with binary versus curve-based coefficients.

In contrast to the above-discussed correlation coefficients, peak-based coefficients are typically binary coefficients that only take into account the absence or presence of a peak. The most used binary coefficients are the Jaccard and Dice coefficients, which count the number of peaks in common, relevant to the total number of peaks present in both profiles. In Figure 21A.8, all three profiles have the rightmost and leftmost peak in common. The middle peak has a high intensity in profile 1, a low intensity in 2, and is absent in profile 3. According to Dice and Jaccard profiles, 1 and 2 are considered a 100% match, whereas profile 3 for that part of the profile would account for 66% similarity only. However, using a Pearson correlation coefficient, profiles 2 and 3 would be considered more similar, as the intensity difference at the position of the middle peak is smaller between these two.

It is obvious from the above that the intensity or signal-to-noise threshold that was used to detect the peaks can have a considerable impact on the similarity calculation results using the Jaccard or Dice coefficient. In Figure 21A.8, the two horizontal lines represent two different thresholds. With the lower threshold, profiles 1 and 2 make a 100% match, but considering the higher threshold, it is profiles 2 and 3 that match at 100%.

Binary coefficients have another disadvantage. To determine whether two profiles have a peak in common, we need to consider a peak position tolerance. If a peak is present at 4508.2 Dalton on one profile, and at 4508.1 Dalton on another profile, it is clear that the position difference falls within the technical variation of the MALDI-TOF MS measurement, and both peaks should therefore be considered matching. In BioNumerics, the position tolerance is defined as a linear function of m/z, thus taking into account the higher variation at the higher-mass region. The constant and linear variation factors of the function are user defined. Typical values for the linear factor are between 300 and 500 ppm, and the constant factor is generally between 0.5 and 1.5 m/z, though the exact values depend on the resolution of the mass spectrometer used. If two peaks are within the position tolerance of each other, they are considered to be in common (matching) between both profiles. In Figure 21A.9, whether the middle peak is considered to be in common between both profiles depends on the settings of the position tolerance.

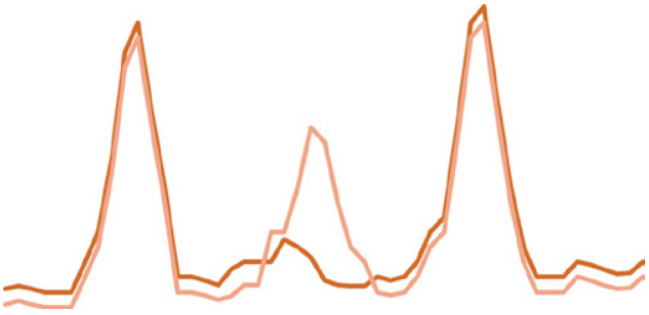


Figure 21A.9 Illustration of the importance of the position tolerance settings when using a binary similarity measure. The middle peaks will be considered as matching only when the position tolerance will be substantially increased.

From the above examples, it is clear that both peak detection and position tolerance settings have a high impact on the similarity calculated with a binary similarity coefficient. Choosing inappropriate parameters could potentially result in incorrect estimation of the relatedness of strains. It is therefore advisable to take care when defining the parameters and preferably validate the parameters on a set of biological replicates before performing this analysis on a set of strains with unknown relationships.

A last coefficient to be discussed uses peaks as the basis for similarity estimation, but it is not a binary coefficient. As some profiles can have much higher noise levels than others, this might affect the Pearson correlation coefficient results obtained between these profiles. The solution to this is to use a Pearson correlation coefficient that is applied not to the raw curve, but instead to a synthetic curve. The synthetic curve is constructed by fitting a Gaussian curve at all positions where a peak was detected with the height of the Gaussian curve corresponding to the original intensity. As the peak detection procedure considers the signal-to-noise ratio, this means that the synthetic curve is much cleaner than the raw curve. In return, results will also largely depend on the choice of appropriate thresholds for the peak detection.

Once the similarity has been determined between all profiles in our sample set, the obtained similarity matrix should be rearranged in order to group the more similar samples together. Therefore, a cluster algorithm is generally used, resulting in a dendrogram accompanied by a structured similarity matrix. Cluster algorithms will locate the most similar pair in the similarity matrix and merge them together. The similarity matrix is therefore reduced by one element, and the similarity values of the remaining samples with the new pair needs to be recalculated. How the matrix is updated varies between different clustering algorithms. In the UPGMA algorithm (unweighted pair group method using averages), the *average* of both similarities is calculated; in a single linkage or neighbor joining approach, the maximum similarity will be used; and in a complete linkage or furthest neighbor clustering, the minimum similarity is taken (Figure 21A.10).

These steps are then repeated in the updated matrix until the complete dataset is ordered.

A clustering algorithm is always true to its name: it will always cluster the data, even random data. However, because of the random way in which high similarity values are picked up, the resulting dendrogram is often only one of the many possible dendrograms that can be constructed on the basis of a given similarity matrix (Kettenring,

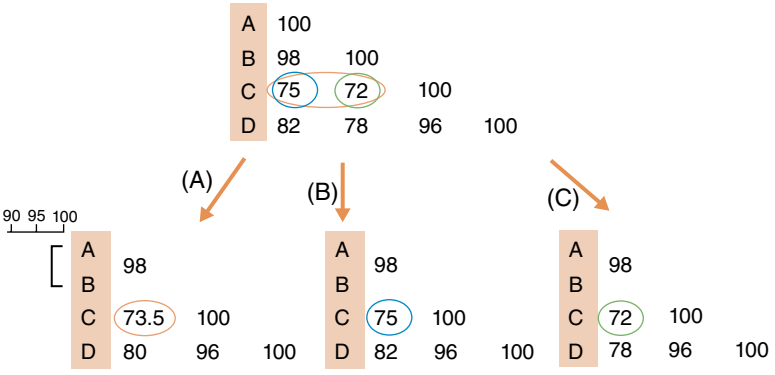


Figure 21A.10 An illustration of the differences of the different cluster algorithms: on top, the original similarity matrix; (A) the averaged UPGMA reduction; (B) the single linkage approach, and (C) the complete linkage method. All methods yield a slightly different merged similarity matrix. The cells on which the merged matrix is based are marked with the same color in the original and merged matrix for entry C.

2006). The main cause is the appearance of identical similarity values in a given similarity matrix. The choice of the two samples to be merged first is then arbitrarily made by the algorithm, giving rise to degeneracies in the resulting dendrogram.

In practice, because UPGMA, single linkage, and complete linkage are *pair group* methods, the clustering algorithm will simply pick the first pair it encounters. In the resulting dendrogram, the complexity of the relationships is therefore not fully inferred. There are several strategies to assess degeneracy or to indicate the significance or reliability of a branch in a dendrogram. An error flag can be calculated that is based on the standard deviation of the similarity values that contribute to the similarity that represents a branch. Another method, the cophenetic correlation, compares the reconstituted similarity matrix derived from the similarity levels in the resulting dendrogram with the original matrix using a Pearson correlation. Finally, one can change the order in which data are considered by the algorithm, either by varying the order of the strains (degeneracy handling), or by permuting the peak classes (bootstrap analysis). In the latter case, a peak matching table is required, in which absence and presence of peaks are scored for each peak class found in the total set of profiles included in the cluster analysis. The different dendrograms obtained from these variations are compared, and for each branching point, the percentage of dendrograms that contains the same branching members is calculated.

21A.12 Alternatives to Cluster Analysis

Although a dendrogram will generally yield a very good picture of the relatedness of complete profiles, one may be interested in specific parts of the profile, for example, subsets of peak changes introduced by differences in culture conditions, variations in growth media or temperature, or differences in sample preparation protocols. Although it is easy to visualize the presence of the different subgroups on a dendrogram (e.g., by using different colors), the underlying peaks that are responsible for this grouping cannot easily be discovered with a dendrogram.

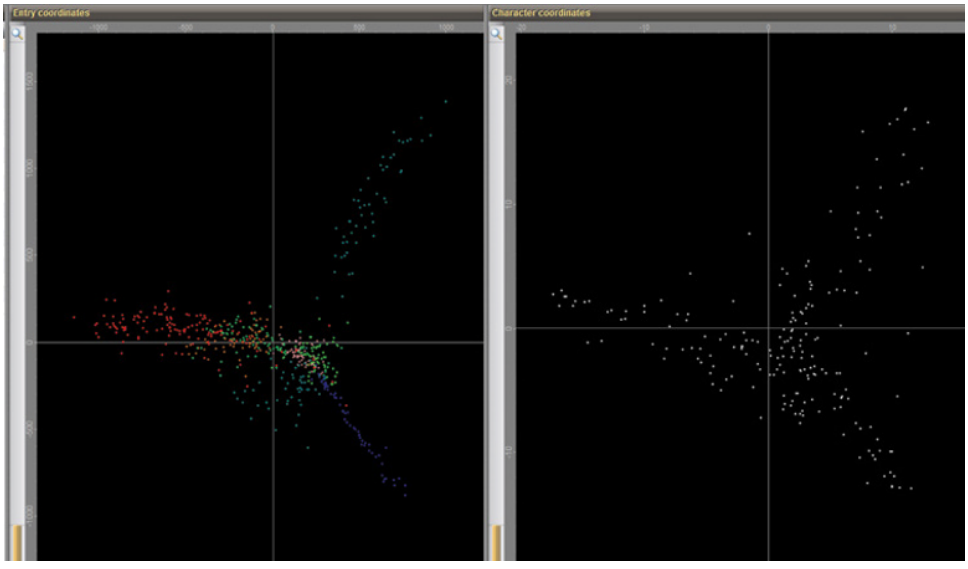


Figure 21A.11 Dual two-dimensional plot of a PCA analysis. Left: a plot of the samples, and right: a plot of the corresponding peak classes (see text for further explanation). The color of the samples is based on the genus.

More advanced techniques are available for this, including PCA and LDA. These algorithms will reduce the complexity of the data to make visual interpretation easier. The visualization generally is composed of two plots that represent on the one hand the position of the samples and on the other hand the position of the peaks. Both windows are linked, as similar positions in the plots support positive links (Figure 21A.11).

Both PCA and LDA start from a closed set of numerical or binary data. To obtain this data from spectral profiles, a peak matching needs to be performed. As explained above, peak matching analysis will determine the absence or presence (and optionally amplitude) for each peak class detected in the total set of profiles included in the cluster analysis. Peak classes will be searched for using a user-defined position tolerance and will be positioned at the average position of all peaks belonging to the class. The result of this analysis can be a binary peak matching table, with absence or presence information for all peak classes, or can be a peak table with numerical values that represent the intensity of the peaks. An example of a peak matching analysis and the resulting binary peak matching table can be found in Figure 21A.12.

The number of dimensions in our dataset corresponds to the number of peak classes, and so lies typically between 100 and 300 dimensions. PCA and LDA reduce the number of dimensions in the dataset, and if the dimensions are reduced to 2 or 3, this allows for visualizations that can be interpreted by eye.

To illustrate the principle of PCA, a typical example is to take a picture of a teapot (Li, 2009). The process of taking a picture reduces the dimensions of an object from 3 to 2. In order to describe the teapot reliably when starting from the picture, the picture needs to be taken at an angle that optimally shows the different elements of the teapot. In all probability, this will not be a picture from the front, where you can only see the spout. It will not be from below or above, where only the bottom or lid will be visible. It will also not be from the back, as this will only show a projection of the handle. Instead, the

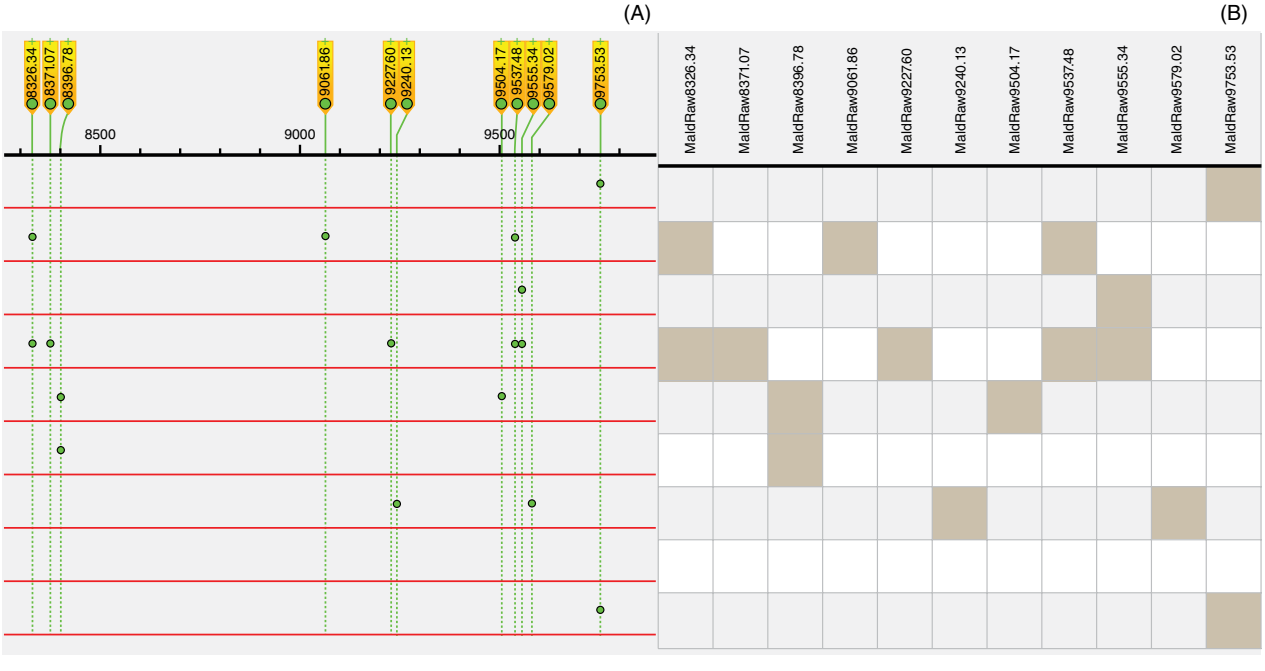


Figure 21A.12 (A) Details of a peak matching analysis with (B) the resulting peak matching table.

side of the teapot is most appropriate, as this captures all elements – lid, spout, and handle – quite well. A similar process happens with PCA, where we will construct new components to describe the dataset that are linear recombinations of the existing dimensions. The first component is constructed in such a way that it will capture as much of the variation in the dataset as possible (the highest standard deviation). The second component will capture the second highest variation, and so on, until the dataset is completely separated or the maximum number of components specified by the user has been reached. The data can now be visualized in a two- or three-dimensional space, using the first component as the X-axis, the second component as the Y-axis, and optionally the third component as the Z-axis. If the cumulative variation of the first two or three components is sufficiently high, the resulting image can be used as a reliable representation of the groups of profiles that are most closely linked and thus can be considered related (Figure 21A.13).

As mentioned above, PCA-like approaches do not allow only visualization of the samples included in the analysis, but the peak classes can also be shown in the same space. Interestingly, the positions of the peak classes in the plot can be compared to the positions of the samples in the neighboring plot (Figure 21A.11). Peak classes appearing in the same location of the plot as a cluster of samples will have a higher chance of being present (or present at a higher intensity) in this group of samples. In contrast to cluster analysis, PCA-like analyses can therefore be used to identify potential biomarkers.

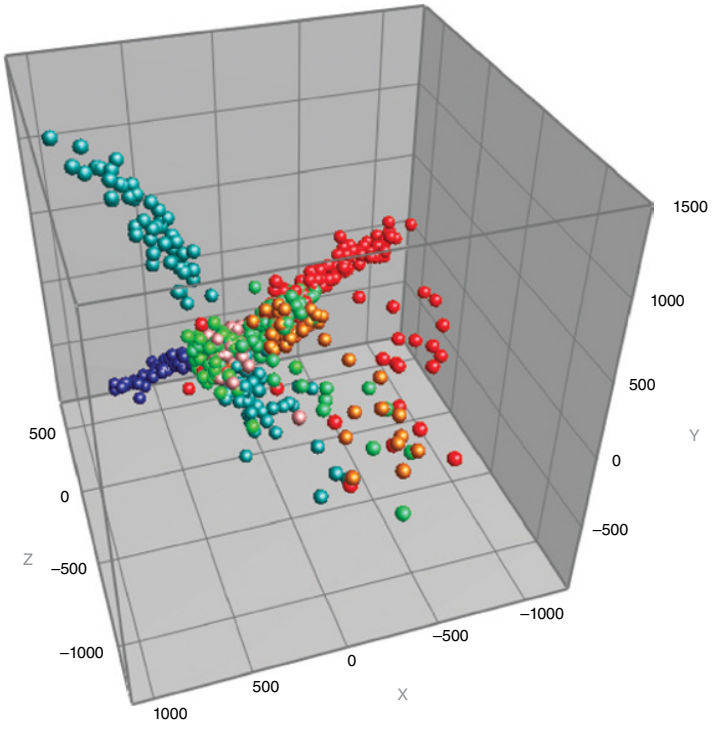


Figure 21A.13 An example of a three-dimensional PCA plot. Samples are colored as in Figure 12.

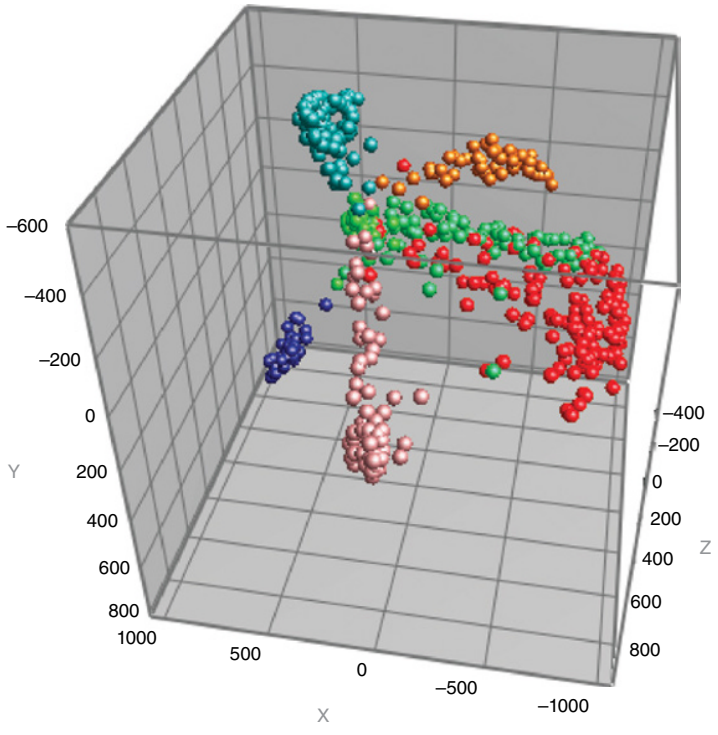


Figure 21A.14 Result of an LDA analysis, presented as a three-dimensional plot. Colors represent groups that were the bases of the LDA analysis (same as Figure 12). Here, the groups were quite well separated.

In order now to illustrate the difference between a PCA and an LDA analysis, we can return to the example of the teapot. Having several teapots that need to be classified, it is likely that the shape of the spout, handle, and lid will help determine which teapots are most similar. Performing a PCA analysis will allow one to create groups of teapots with similar spouts, handles, and lids. However, consider the case of a group of teapots that are highly similar, but half of them are expensive antique teapots while the other half are cheap reproductions of the same antiques. In this case, a PCA will not really be helpful in distinguishing fake from real. To help us find the components that distinguish the real from the fake teapots, an LDA will be very useful. An LDA analysis is similar to PCA, but instead of looking for components that capture the maximum variation of the complete dataset, it constructs components that maximize the difference between predefined groups in the dataset, while minimizing the difference within these groups. In the example of the teapots, this could result in a picture of the bottom of the teapots, where there is a mark distinguishing a fake teapot from a real teapot. By just looking at pictures from the side of the teapots, we never would have been able to distinguish a real teapot from a fake one.

Similar to PCA, we can visualize the dataset and the peak classes with the first, second, and, optionally, the third component as axes (Figure 21A.14). Again, if a peak class is in a similar location as a group of profiles, it is a likely candidate biomarker for this group. If the LDA of the profiles results in a cloud of points with no clear separation, this means that there is no information in our profiles that is able to clearly distinguish between the defined groups (Figure 21A.15).

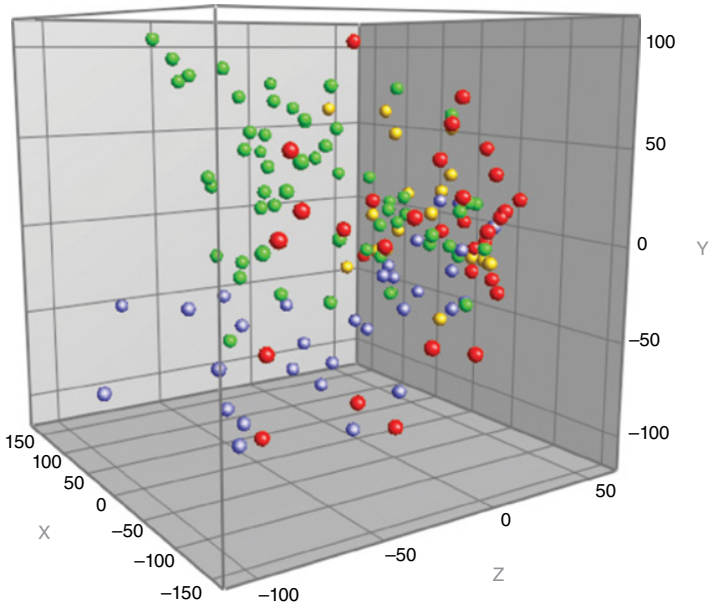


Figure 21A.15 Result of an LDA analysis, presented as a three-dimensional plot. Colors represent groups that were the bases of the LDA analysis. Here, the groups were difficult to separate on the basis of the underlying dataset.

21A.13 Classifying Algorithms

The objective of an LDA is to identify potential biomarkers and determine whether predefined groups can be distinguished using the available data. This requires prior knowledge of the groups the user wishes to distinguish. With classifying algorithms, we can go even further and use the dataset of which we have this prior knowledge as a reference set to then identify unknown samples. There are several algorithms to do this, such as naïve Bayesian statistics or support vector machines. Other techniques are based on similarities (Ressom *et al.*, 2007; De Bruyne *et al.*, 2011).

Similarity-based algorithms work by assigning the unknown profile to the group it has the highest similarity to. The similarity can be calculated with the same coefficients as is used for hierarchical clustering. To assign an unknown profile to a group, the average, maximum, or minimum similarity can be used, though more advanced similarity algorithms also exist, such as balanced and weighted similarity. A balanced similarity takes both the maximum and minimum similarity into account, whereas the weighted similarity increases the weight of highly similar profiles in the reference set while decreasing the weight for profiles further away.

The naïve Bayesian and SVM algorithms work on the peak matching table. The full peak matching table can be used, as well as subsets. This way, the identification can be based on just those peaks that were identified as potential biomarkers using LDA or any other statistical analysis (e.g., MANOVA), generally speeding up the analysis and in some cases avoiding an overfitting effect. This occurs if the number of data points used for identification (peak classes in this case) is much larger than the number of profiles in each group. More advanced algorithms will very often find specific rules to assign

this small reference set in the correct class, but these rules are often based more on chance and are not applicable to larger sets. The SVM is most prone to overfitting and should therefore not be used if the sample sets are very small. The naïve Bayesian approach, however, makes the assumption that the different peak classes are independent of each other, which seems an oversimplification in the case of MALDI-TOF MS profiles.

Again, the proof of the pudding is in the eating. BioNumerics allows one to perform an internal validation of the identification performance, based on the reference set. Each profile in the reference set is removed one by one, treated as unknown, and identified using a chosen algorithm. The results are then compared to the true group to get an idea of the amount of correct and false classifications. This allows one to compare different algorithms. However, this approach is influenced by the reference set used and is only reliable if performed on a reference set that is representative of the population of unknown samples to be studied.

For all datasets used further in this chapter, we constructed identification projects and performed the classification based on maximum, average, and balanced similarity, using a Pearson correlation coefficient to calculate the similarity, as well as a naïve Bayesian and support vector machine approach. The precision and recall was determined for each classification. For the dataset used in Part B (dataset B), the species assignment was used to construct groups. Any species with fewer than four reference profiles was excluded from the analysis. This dataset was considered "easy" to identify as these species are known to be readily distinguishable with MALDI-TOF. The dataset used in Part C (dataset C) was identified to a VNTR clonal complex. As mentioned in more detail in Part C, no correlation was seen between the MALDI-TOF MS profiles and the VNTR type. This dataset was therefore considered difficult (or even impossible) to identify. Two further datasets were added that contain spectra of species difficult to separate with MALDI-TOF: one dataset containing several *Listeria* species (dataset 1) and another containing *E. coli* and several closely related *Shigella* species (dataset 2). These two datasets were also considered difficult to identify (Mahé *et al.*, 2013). The detailed results can be found in Table 21A.2.

For dataset B, all algorithms performed very well and were able to identify the species correctly in the internal validation for the majority of profiles. There was a slightly better performance for the naïve Bayesian and SVM, but the differences were marginal.

Table 21A.2 Results (P for precision, R for recall) of the comparison of five different classifying algorithms on four different datasets.

	Dataset B		Dataset C		Dataset 1		Dataset 2	
	P	R	P	R	P	R	P	R
Maximum similarity	98.95%	98.89%	17.25%	10.87%	70.49%	70.00%	62.86%	61.70%
Average similarity	97.16%	96.31%	8.50%	15.22%	70.16%	70.00%	61.29%	42.55%
Balanced similarity	99.64%	90.41%	11.66%	17.39%	71.98%	71.67%	66.00%	62.41%
Naïve Bayesian	98.95%	98.89%	21.15%	15.22%	60.00%	60.00%	59.45%	44.68%
SVM	98.92%	98.89%	39.49%	30.43%	86.86%	86.67%	70.84%	70.21%

For dataset C, none of the algorithms performed well, though the SVM did outperform the other algorithms, though not to a degree that we could consider as a reliable identification. For Datasets 1 and 2, the SVM outperforms all other algorithms, leading to a decent number of correct identifications. We therefore consider the SVM to be the preferred algorithm for the identification of this type of data.

21A.14 Conclusion

MALDI-TOF MS profiles generated from whole proteomes of bacterial isolates can be readily used for species identification. However, the use for subtyping is not as straightforward and requires more optimization and validation of the complete protocol, from sample preparation to computational analysis. It is not feasible to use MALDI-TOF MS to subtype each and every dataset to the degree that matches other typing technologies, but the possibility is often worth investigating. The methods discussed here can help make the call regarding whether subtyping can be performed to a satisfactory degree. In many cases, the ability to perform a subtyping with MALDI-TOF, even if it is not perfect, can vastly decrease the time to make a (preliminary) clinical decision.

References

De Bruyne, K., Slabbinck, B., Waegeman, W., Vauterin, P., De Baets, B., & Vandamme, P. (2011). Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Syst Appl Microbiol*, 20–29.

Du, P., Kibbe, W., & Lin, S. (2006). Improved peak detection in mass spectrum by incorparating continuous wavelet transform-based pattern matching. *Bioinformatics*, 2059–2065.

Kaiser, J. (1966). Digital filters. In F. Kuo, & J. Kaiser, *System Analysis by Digital Computer* (p. Chap. 7). New York: Wiley.

Kettenring, J. (2006). The practice of cluster analysis. *Journal of Classification*, 3–30.

Li, J. (2009, September 23). A layman's introduction to principal component analysis.

Mahé, P., Arsac, M., Chatellier, S., Monnin, V., Perrot, N., Maillier, S., Girard, V., Ranjeet, M., Surre, J., Lacroix, B., van Belkum, A., & Veyrieras, J. (2013). Automatic identification of mixed bacterial species fingerprints in a MALDI-TOF mass-spectrum. *Bioinformatics*, 4–23.

Monchamp, P., Andrade-Cetto, L., Zhang, J., & Henson, R. (2007). Signal processing methods for mass spectrometry. In G. Alterovitz, & M. Ramoni, *Systems Bioinformatics: An Engineering Case-Based Approach* (pp. 101–124). Norwood: Artech House Publishers.

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (2007). Hierarchical clustering by phylogenetic trees. In *Numerical Recipes: The Art of Scientific Computing (3rd ed.)* (pp. 868–882). New York: Cambridge University Press.

Ressom, H., Varghese, R., Drake, S., Hortin, G., Abdel-Hamid, M., Loffredo, C., & Goldman, R. (2007). Peak selection from MALDI-TOF mass spectra using ant colony optimization. *Bioinformatics*, 619–626.

Sabat, A., Budimir, A., Nashev, D., Sá-Leão, R., Van Dijl, J., Laurent, F., Grundmann, H., & Friedrich, A. (2013). Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Eurosurveillance*, 20380.

Schafer, R. (2011). What is a Savitsky-Golay filter. *IEEE Signal Processing Magazine*, 111–117.

Schwartz, D., & Cantor, C. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 67–75.

Spinali, S., van Belkum, A., Goering, R., Girard, V., Welker, M., Van Nuenen, M., Pincus, D. H., Arsac, M., & Durand, G. (2014). Microbial typing by MALDI-TOF MS: Do we need guidance for data interpretation? *J Clin Microbiol*, 53, 760–765.

Swaminathan, B., Barrett, T., Hunter, S., Tauxe, R., & CDC Pulsenet Task Force. (2001). PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerging Infectious Diseases*, 382–389.

Vauterin, L., & Vauterin, P. (1992). Computer-aided objective comparison of electrophoretic patterns for grouping and identification of microorganisms. *European Microbiology*, 37–41.

Vauterin, L., & Vauterin, P. (2006). Integrated databasing and analysis. In E. Stackebrandt, *Molecular Identification, Systematics, and Population Structure of Prokaryotes* (pp. 141–217). Berlin: Springer-Berlin Heidelberg.

Yang, C., He, Z., & Yu, W. (2009). Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics*, 4.

21B

Subtyping of *Staphylococcus* spp. Based upon MALDI-TOF MS Data Analysis

Zhen Xu,[1] Ali Olkun,[2] Katleen Vranckx,[3] Hermine V. Mkrtchyan,[1] Ajit J. Shah,[2] Bruno Pot,[3] Ronald R. Cutler[1] and Haroun N. Shah[2,4]

[1] *School of Biological and Chemical Sciences, Queen Mary University of London, London, UK*
[2] *Department of Natural Sciences, Middlesex University, London, UK*
[3] *Applied Maths NV, Sint-Martens Latem, Belgium*
[4] *Proteomics Research, Public Health England, London, UK*

21B.1 Introduction

Mass spectral data of microorganisms derived from a linear MALDI-TOF instrument are subject to variation in resolution depending on the method of sample preparation and the parameters used for analysis. For clinical applications, highly consistent and reproducible spectra between different analytical runs and between various instruments are essential and, over the years, staphylococci have been used widely as test organisms to help optimize sample analysis. For example, fast atomic bombardment mass spectrometry was used initially to study the taxonomy of staphylococci based on their polar lipid profiles (Drucker and Abdullah, 1995). With the arrival MALDI-TOF MS, attention turned towards the analysis of proteins, and unique m/z values were reported for staphylococci derived from their cell wall moieties (Claydon *et al.*, 1996). Methods at this time varied between laboratories, and recognizing the potential of the technique, attempts were made to develop a universal method in which *S. aureus* was again one of the reference species (Shah *et al.*, 2000). Subsequently, this method was used to assemble the first microbial MALDI-TOF MS database in which members of the genus *Staphylococcus* were a fundamental component (Keys *et al.*, 2004).

To field-test the method, a random selection of 95 hospital isolates and 39 from a Staphylococcal Reference Unit (PHE, London) were analyzed by MALDI-TOF MS using the first dedicated microbiology bench-top instrument, a Micromass/Waters M@ LDI-MS with the company's bespoke MicrobeLynx software to facilitate the identification of species. The identity of each isolate was confirmed separately using 16S rRNA sequence analysis and showed excellent concordance. These and other studies (Carbonnelle *et al.*, 2007; Etienne *et al.*, 2011; Stoltenberg and Nag, 2010; Seng *et al.*, 2010) provided proof of concept that MALDI-TOF MS could be used for the routine

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

identification of members of the genus *Staphylococcus*. Attention then turned to investigating the potential of MALDI-TOF MS to subtype strains, particularly for *S. aureus*, where genetic typing methods such as pulsed-field gel electrophoresis (PFGE), *spa* typing, or multilocus sequence typing (MLST) were revealing immense intra-species diversity (Monecke *et al.*, 2014). These methods are, however, time consuming, costly and cumbersome; consequently, MALDI-TOF MS, by virtue of its simplicity and cost-effectiveness, prompted intensive investigations.

The impetus for these studies was largely driven by concerns over the emergence of methicillin resistance (MRSA) and the high rates of mortality and morbidity occurring globally (Oliveira and Tomasz, 2002). To implement infection control measures, epidemiological relatedness of clinical isolates is essential. Subtle variation in mass intensity and m/z values (most likely due to post-translational modifications) give rise to a plethora of mass spectral profiles that reflect the diversity within a species. Although the 'core' component of the spectrum could be exploited to derive reliable species identification, investigators began exploiting minor variation in m/z values to develop methods to use as epidemiological tools. The latter is largely dependent on complex data analysis, and several methods such as principal component analysis, artificial neural networks, or software packages such as ClinProTools have been used (Lancashire *et al.*, 2005; Shah *et al.*, 2011; see Chapter 6). Here we assess the potential of BioNumerics (version, 7.5' Applied Maths, Belgium), as a hierarchical clustering tool to analyze MALDI-TOF MS data from staphylococcal isolates from the environment to help trace possible transmission between sites and to discern intra- and inter-species relationships.

21B.2 Sample Collection

In a recent study conducted between 2013 and 2015, 411 isolates of staphylococci recovered from the hands of volunteers (n = 107), handbags (n = 10), different sites in hotel rooms (n = 53) and from air (n = 5), libraries (n = 25), restaurants (n = 67), supermarkets (n = 59), baby care facilities (n = 33) and public transport systems (n = 52) in London, United Kingdom, were collected. Staphylococci were provisionally identified using selective media (Mannitol Salt Agar, Oxoid Ltd, Basingstoke, UK) and were additionally characterized using the Prolex™ Staph Xtra Latex Kit (Prolab Diagnostics, Neston, South Wirral, UK).

21B.3 MALDI-TOF Mass Spectrometry

Staphylococci samples were prepared as described previously (Sogawa *et al.*, 2012) and identified using matrix-assisted laser desorption/ionization time-of-flight mass spectroscopy (Bruker Microflex LT, MALDI-TOF-MS; Public Health England, London) in positive linear mode (2,000 to 20,000 m/z range). The resulting spectra for each culture was analyzed by MALDI Biotyper 3.1 software (Bruker Daltonics, Coventry, UK). *E. coli* DH5 α (Bruker Daltonics, Coventry, UK) was used as a standard for calibration and quality control (Samb-Ba *et al.*, 2014).

21B.4 Cluster Analysis of Environmental Staphylococci

BioNumerics 7.5 (Applied Maths, Belgium) was employed for cluster analysis based on the mass spectral data.

MALDI-TOF MS spectra of isolates recovered from different sites were analyzed using the BioNumeric software package (BioNumeric Version 7.5) described above (see Section 21A). In order to differentiate each sample's collection site, different colours were used to indicate each site. Furthermore, the spectrum of different isolates was compared using the Pearson correlation coefficient, after background subtraction and noise removal. The interrelationship between strains was derived using the UPGMA method, and the 3D images were constructed using the multidimensional scaling method described in the BioNumerics manual. All selected isolates were additionally grouped based on their antibiotic susceptibility profiles. In these groups, the red colour was selected to demonstrate the presence of multiple-resistant staphylococci, and the green colour was used to indicate susceptible staphylococci isolates.

21B.5 Antibiotic Susceptibility Test

Isolates were screened for their susceptibility against 12 antibiotics using the disc diffusion method (BSAC disc diffusion method) (Andrews and Howe, 2011). The antibiotics tested included oxacillin (1 μg); vancomycin (5 μg) (VAN); gentamicin (10 μg); mupirocin (20 μg) (MUP); amoxicillin (10 μg); erythromycin (15 μg) (ERY); tetracycline (10 μg) (TET); streptomycin (10 μg); cefepime (30 μg) (CEP); fusidic acid (10 μg) (FC); penicillin (1 unit) and chloramphenicol (30 μg) (CHL) (Andrews and Howe, 2011). The categories susceptible, intermediate or resistant were assigned on the basis of the breakpoints recommended by the BSAC standardized disc susceptibility test method (version 10) (Andrews and Howe, 2011). The MICs for oxacillin were additionally evaluated using M.I.C. evaluators, antimicrobial gradient strips designed for accurate minimum inhibitory concentration (MIC) values (Oxoid Ltd., Basingstoke, UK) (Andrews and Howe, 2011).

A total of 411 isolates recovered from the general public and different environmental sites were analyzed using MALDI-TOF MS. For simplicity, these are abbreviated as follows: different sites in hotels (DSH); hotel air samples (HAS); human hands (HH); different sites in a supermarket (DSS); handbags (HB); baby care facilities (BCF); different sites in library (DSL); different sites in transportation (DST) and different sites in restaurant (DSR).

Nineteen species of staphylococci were identified, most of which were coagulase-negative staphylococci. Eleven out of all isolates identified were *Staphylococcus aureus*. Six out of 19 species identified were not systematically analyzed in BioNumerics, as the identification scores generated by the Biotyper 3.0 software were below 2.000. Using MALDI-TOF MS, coagulase-negative staphylococci were classified into 12 species, including *S. epidermidis* (n = 123), *S. hominis* (n = 111), *S. warneri* (n = 35), *S. capitis* (n = 50), *S. haemolyticus* (n = 42), *S. pasteuri* (n = 21), *S. saprophyticus* (n = 9), *S. simiae* (n = 4), *S. cohnii* (n = 2), *S. caprae* (n = 1), *S. lugdunensis* (n = 1) and *S. simulans* (n = 1).

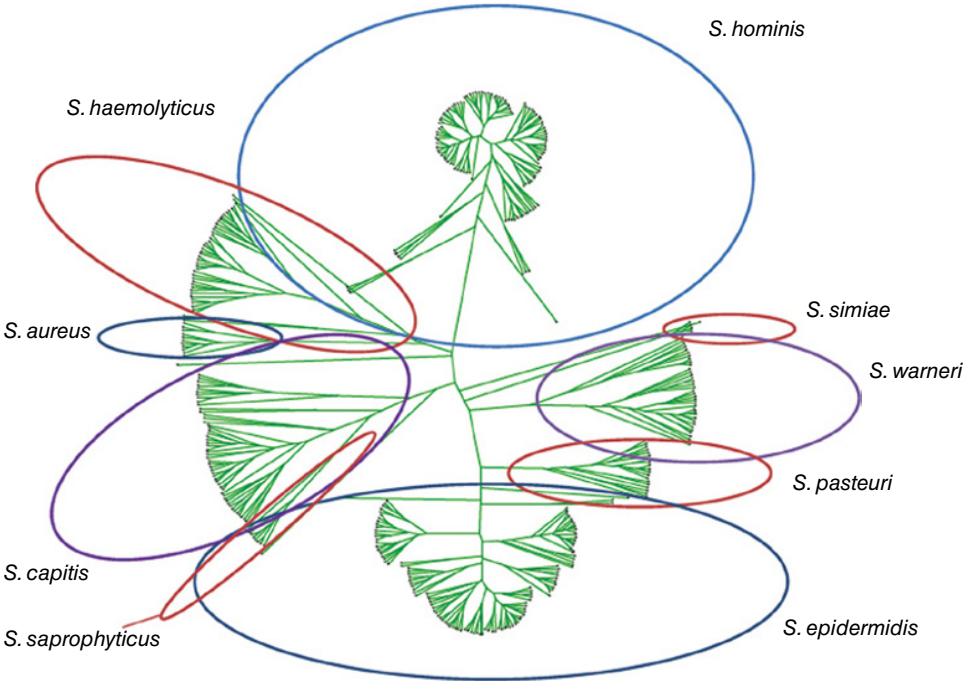


Figure 21B.1 Unrooted cluster analysis of staphylococci species using MALDI-TOF-MS. The intra-species diversity of each taxon (within each circle) is clearly apparent.

Analogous to phylogenetic analysis, nine major clusters, of staphylococci were identified, namely, *S. hominis*, *S. haemolyticus*, *S. epidermidis*, *S. pasteuri*, *S. warneri*, *S. aureus*, *S. saprophyticus*, *S. capitis* and *S. simiae*. The hierarchical interrelationships derived using BioNumerics 7.5 are shown in Figure 21B.1.

21B.6 Cluster Analysis of *Staphylococcus* spp. Recovered from Different Sites

The presence of staphylococcal species differed between sites. The most common species isolated from DSL were *S. epidermidis* and *S. haemolyticus*, whereas *S. epidermidis* was predominant among the isolates recovered from DST and HB. Moreover, *S. epidermidis* together with *S. hominis* were predominant among the isolates recovered from the HH, DSS and DSR. *S. haemolyticus* and *S. hominis* were predominant among the isolates recovered from the DSH. In addition, *S. haemolyticus* was predominant among the isolates recovered from HAS. The common species isolated from BCF included *S. hominis*, *S. epidermidis* and *S. saprophyticus*. Interestingly, *S. aureus* was not the predominant species in any site tested. Instead, *S. epidermidis* was the major component of the flora in all samples except those associated with hotels. It is interesting to note that Harris *et al.* (2010) utilized the same system, namely, Bruker Daltonics Microflex MALDI-TOF/MS with MALDI Biotyper software to identify 158

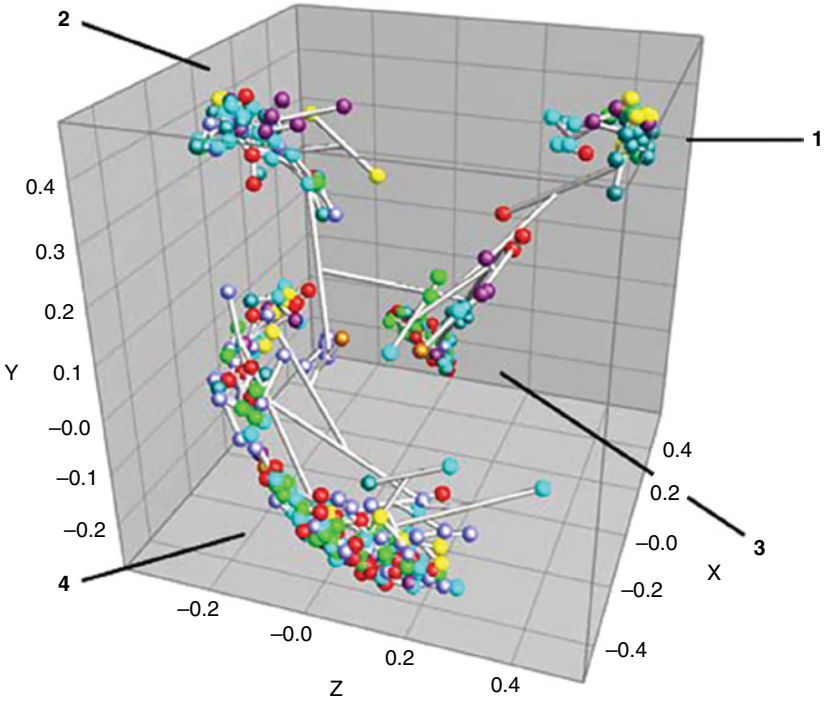


Figure 21B.2 Three-dimensional (3D) scatter plot of 411 staphylococci recovered from nine sites. DSS ; DSR ; DST ; DSL ; HH ; DSH ; HB ; BCF ; HAS . X: –0.4 to 0.4; Y: –0.2 to 0.4; Z: –0.2 to 0.4.

characterized staphylococcal isolates from prosthetic joint infections, but did not achieve the same level of confidence of species identification. Although they confidently separated *Staphylococcus aureus* from coagulase-negative staphylococci, nearly 25% *Staphylococcus epidermidis* achieved very low confidence scores and may reflect the great strides made in the last five years in increasing the level of confidence of identification to the species level.

Three-dimensional scaling was performed to demonstrate the overall relationship between the 411 staphylococcal isolates (Figure 21B.2). On the basis of the MALDI-TOF MS data, all isolates were distributed into four groups. Groups 1, 2 and 3 lacked extensive diversity compared to the fourth group (Figure 21B.2).

21B.7 Correlation of Staphylococci Recovered from Different Sites

Staphylococcal isolates recovered from different sites were closely related. Interestingly, *Staphylococcus* spp. recovered from HH were related to those isolated from DSS, DSR, HH, DSH, HB, BCF, DSL and HAS (Figure 21B.3A), whereas staphylococci isolates recovered from BCF were related to those recovered from DSS, DST, DSL, HH, DSH

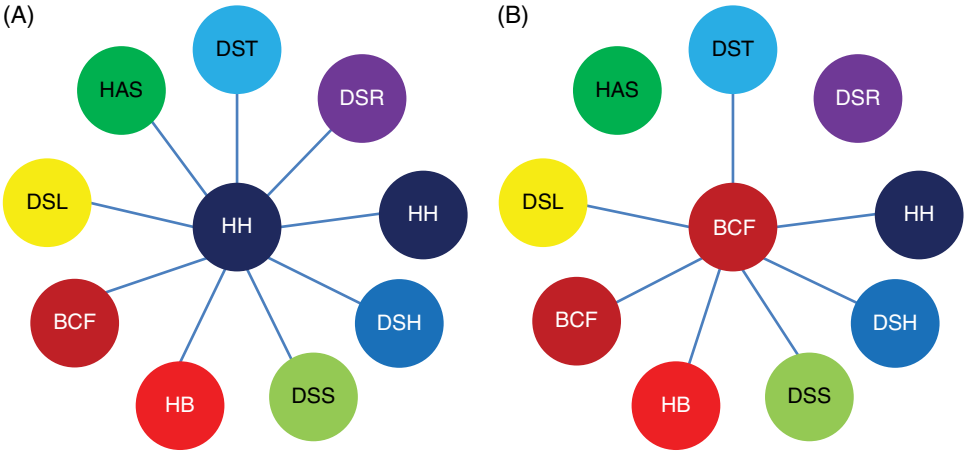


Figure 21B.3A and 21B.3B Staphylococcal species recovered from different sites. Abbreviations: DSH = different sites in hotels; HAS = hotel air samples; HH = human hands, DSS = different sites in a supermarket; HB = handbags; BCF = baby care facilities; DSL = different sites from library; DST = different sites in public transportation, DSR = different sites in a restaurant.

and BCF (Figure 21B.3B). In addition, staphylococci isolated from DSS were related to staphylococci isolated from DSR, DST, HH, DSH, BCF and DSS itself. Staphylococci recovered from DST were related to staphylococci recovered from DSS, DSR, HH, BCF and DST. Staphylococci recovered from DSH were related to those recovered from DSS, HH, BCF, HAS and DSH. Whereas staphylococci isolated from DSR were related to isolates recovered from DST and HH, the staphylococci recovered from DSL were related to those isolated from HH and BCF. In addition, staphylococci recovered from HB were related to HH and BCF.

In addition to analyzing all staphylococcal species, cluster analysis was applied to each staphylococcal spp. Here, *S. epidermidis* and *S. aureus* were selected to explore their correlation with different sites.

21B.8 Cluster Analysis of *S. epidermidis* Isolated from Different Sites

Isolates of *S. epidermidis* recovered from DSH, BCF, DSR, DSS, HH, HB, DSL and DST were shown to be closely related. In addition, these isolates were organized into nine large clusters (Figure 21B.1). *S. epidermidis* recovered from HH, DSR, DSS, DSL and DST were in the same cluster as *S. epidermidis* recovered from HH, DSR, DSS, DSL and DST. In addition, these results showed that *S. epidermidis* isolates recovered from DST were located in the same cluster as these isolated from BCF, indicating their close relationship. In addition, *S. epidermidis* isolates recovered from HH and DSH, HH and DSR, HH and DST, HH and DSS, HH and BCF, HH and HB, DSS and DSL, DSS and BCF, DSS and DSR, DSS and DSH, HB and DSR, HB and BCF, and DSR and DST were also located in same cluster (Figure 21B.4).

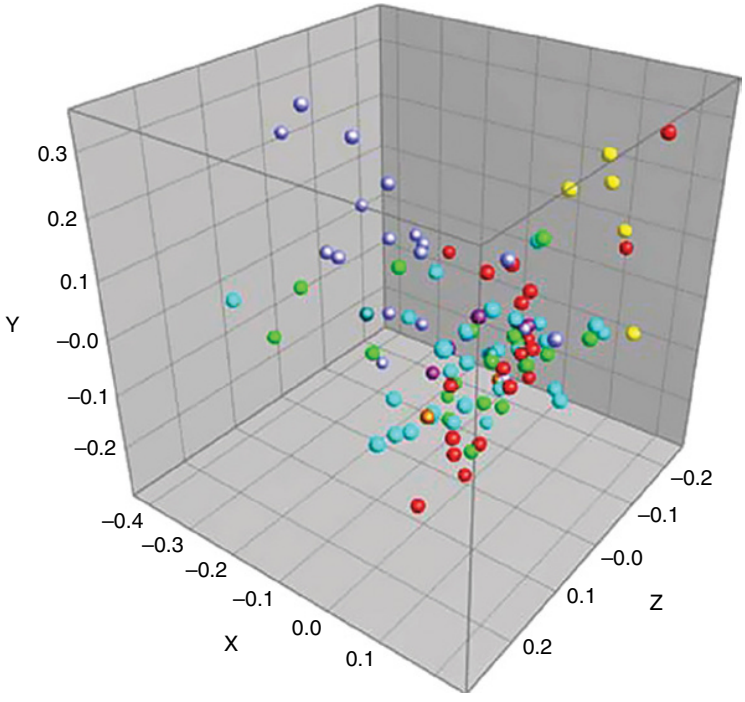


Figure 21B.4 Three-dimensional (3D) scatter plot of *S. epidermidis* isolated from DSS, DSR, DST, DSL, HH, DSH, HB and BCF. DSS ; DSR ; DST ; DSL ; HH ; DSH ; HB ; BCF . X: –0.4 to 0.1; Y: –0.2 to 0.3; Z: –0.2 to 0.2.

21B.9 Cluster Analysis of *S. aureus* Isolated from Different Sites

Eleven *S. aureus* isolates were analyzed in this study. Isolates that were recovered from six different sites (DSS, DSR, HB, HH, DSH and BCF) formed two major clades. Two *S. aureus* isolates recovered from DSR were found in the same clade. Two *S. aureus* isolates recovered from DSR were closely related to those recovered from DSS. Apart from this, two *S. aureus* isolates (one recovered from DSH and the other from HH) were found to be located in same cluster, indicating a close relationship (Figure 21B.5).

21B.10 Cluster Analysis of *Staphylococcus* spp. Combined with Antibiotic Susceptibility

An antibiotic susceptibility test was performed for all isolates, and 336 (82%) out of 411 staphylococci were resistant against two or more antibiotics. Others varied in their resistance profiles. Of all isolates tested, 22 were susceptible to all antibiotics tested.

Staphylococci resistant to more than two antibiotics were considered to be multiple antibiotic resistant, whereas susceptible isolates were those that demonstrated resistance to one antibiotic or none. The distribution of antibiotic resistance patterns in

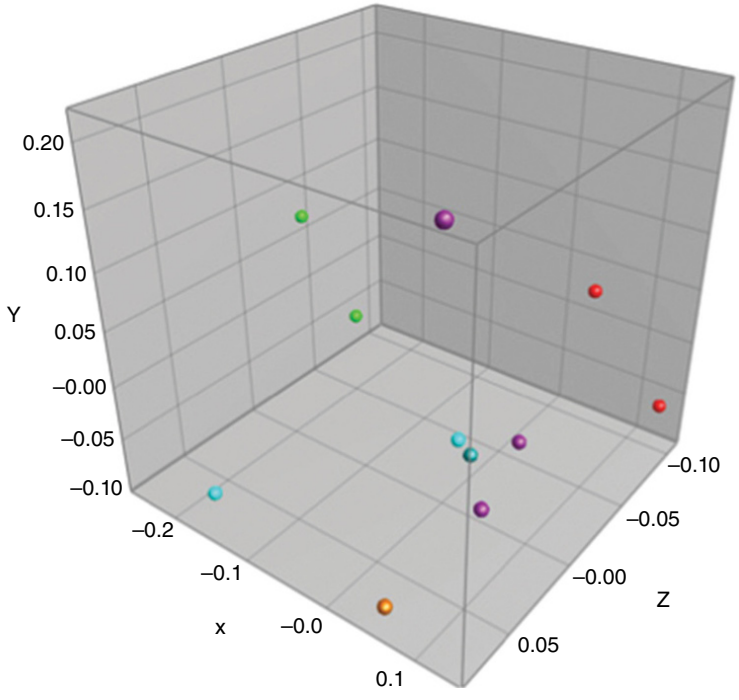


Figure 21B.5 Three-dimensional (3D) scatter plot of *S. aureus* isolated from DSS, DSR, HH, DSH, HB and BCF. DSS ●; DSR ●; HH ●; DSH ●; HB ●; BCF ●. X: –0.2 to 0.1; Y: –0.1 to 0.2; Z: –0.1 to 0.05.

staphylococcal isolates was analyzed using BioNumerics. It was found that susceptible and multiple-resistant isolates were related (Figure 21B.6), and 30 multidrug resistant isolates were closely related to 30 susceptible isolates, respectively, indicating that these might belong to the same genotype of the parent strain.

21B.11 Antibiotic Resistance Patterns of Closely Related *S. epidermidis*

Eighty-five (69%) multiple-resistant and 38 (31%) susceptible *S. epidermidis* were analyzed. The pattern was again similar, with 14 multiple-antibiotic-resistant *S. epidermidis* being closely related to 14 susceptible *S. epidermidis* (Figure 21B.7).

21B.12 Antibiotic Resistance Patterns of Closely Related *S. aureus*

In this study, there were 10 multiple-resistant and 1 susceptible *S. aureus* strains. The multiple resistant *S. aureus* were not related to this susceptible *S. aureus*, but the number of strains analyzed was too small to (Figure 21B.8).

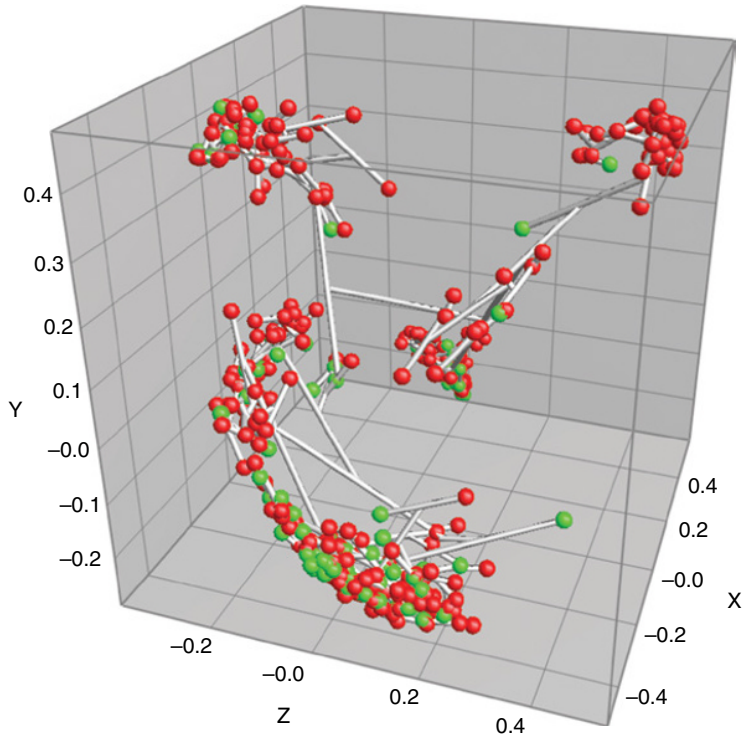


Figure 21B.6 Three-dimensional (3D) scatter plot of staphylococcal multiple-antibiotic-resistant ● and susceptible ● isolates. X: –0.4 to 0.4; Y: –0.2 to 0.4; Z: –0.2 to 0.4.

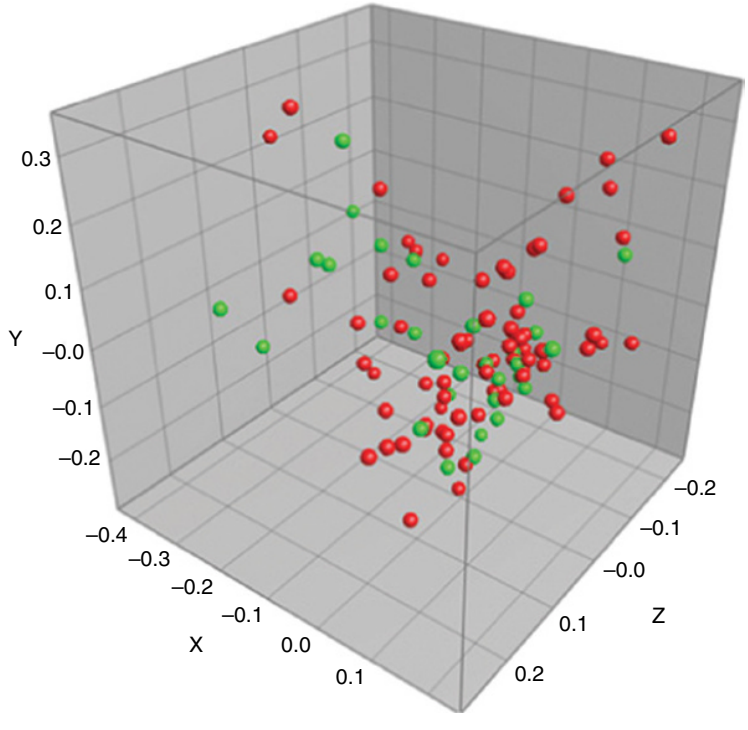


Figure 21B.7 Three-dimensional (3D) scatter plot of multiple-resistant ● and susceptible ● isolates of *S. epidermidis*. X: –0.4 to 0.1; Y: –0.2 to 0.3; Z: –0.2 to 0.2.

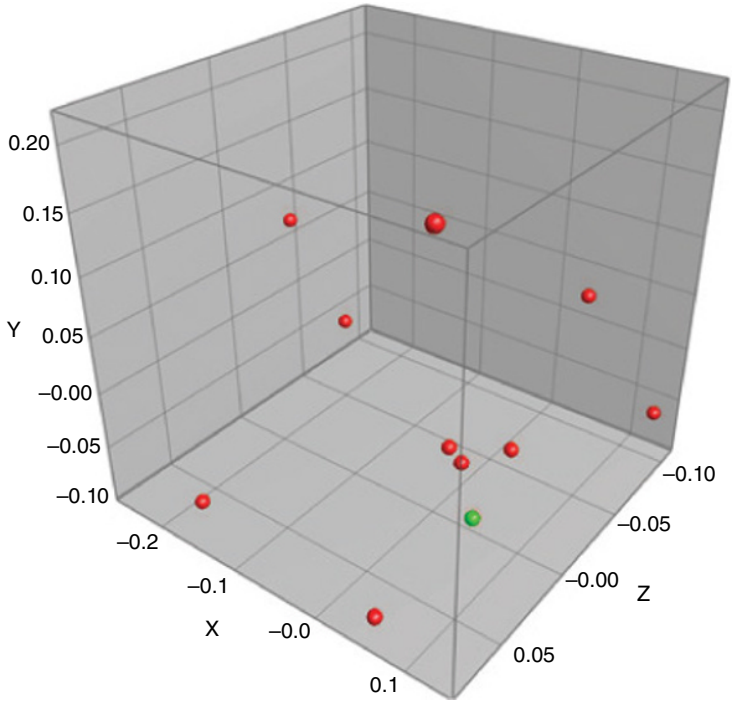


Figure 21B.8 Three-dimensional (3D) scatter plot of multiple-resistant ● and susceptible ● *S. aureus.* X: –0.2 to 0.1; Y: –0.1 to 0.2; Z: –0.1 to 0.05.

21B.13 Variations of Antibiotic Susceptibility of Closely Related *S. epidermidis*

In order to identify antibiotic susceptibility variations of closely related *S. epidermidis*, two representative isolates recovered from different sites as well as from the same sites were selected from each cluster.

For *S. epidermidis,* 34 closely related clusters were identified. Twenty-three of these clusters were formed by isolates recovered from different sites. No antibiotic sensitivity difference was detected in any one cluster but instead were widely distributed. In 2 out of 11 clusters, there was no antibiotic susceptibility variation, but other clusters contained a mixture of susceptibility patterns.

21B.14 Percentage of Multiple-Resistant Staphylococci Recovered from Each Site

A total of 325 (80%) of all staphylococci isolates were multiple resistant, and 86 were susceptible. All five staphylococci isolates recovered from HAS were multiple antibiotic resistant. Thirty-one (94%) of all isolates recovered from BCF were multiple-resistant staphylococci, and 2 (6%) were susceptible. More than 80% of staphylococci isolates recovered from DSH (46), DSS (51), HH (86) and DSL (22) were also multiple resistant.

Forty-eight (72%) of 67 isolates recovered from DSR were multiple resistant. In addition, 6 (60%) and 30 (58%) of staphylococci isolates recovered from HB and DST, respectively, were also multiple resistant.

21B.15 Conclusion

MALDI-TOF MS was used to examine the distribution of members of the genus *Staphylococcus* in multiple sites within one city and utilized to investigate its capacity to identify such diverse environmental isolates. The data was subjected to analysis using BioNumerics to study species diversity and antibiotic sensitivity patterns of different staphylococci species. It has long been suggested that the human hand is a reservoir of one of the major causes of cross-contamination and antibiotic resistance transmission in health care facilities (Pratt *et al.*, 2001), and this is supported by the above study, in which isolates recovered from hands were related to DSH, DSS, DSR, DSL, DST, HB, BCF, HAS and HH itself. Thus, the majority of isolates here were recovered from inanimate objects regularly touched by hands. This finding also aids our understanding of antibiotic resistance transmission and dissemination, which has also been discussed by others (Simões *et al.*, 2011). De Neeling *et al.* (2007) reported the isolation of MRSA from slaughterhouse air samples, and demonstrated the transmission of MRSA via aerosol. Here, air isolates were recovered from hotel environments, and there was good correlation between isolates recovered from air and different sites in the same hotels.

The relationship between antibiotic susceptibility and genotype has been widely reported (Thouverez *et al.*, 2003). Here, MALDI-TOF MS analysis showed that closely related antibiotic-resistant staphylococci were recovered from different sites, indicating their wide dissemination in the environment. Up to nine antibiotic susceptibility variations were observed in two closely related staphylococci, which were recovered from same site, and up to eight antibiotic susceptibility variations were found in two other related staphylococci recovered from different sites.

Though staphylococci are residents of human skin, *S. aureus* and many coagulase-negative staphylococci may act as potential human pathogens and cause life-threatening infections (Schafer, 1979; Otto, 2009; Meers *et al.*, 1975; Lowy, 1998; Ma *et al.*, 2005; Diekema *et al.*, 2001). Twelve out of 13 species identified were coagulase-negative staphylococci. *S. aureus* isolates were recovered from all sites but not from public transportation. Other authors reported different staphylococci spp. isolated from soil, water and food (Kamal *et al.*, 2013; Normanno *et al.*, 2007). Interestingly, *S. simiae* (Pantucek, 2005) has been shown to be associated with monkeys in South Africa.

Sexton *et al.* (2006) reported that the environment may play an important role in the dissemination of antibiotic resistance. The results obtained here are consistent with these findings as multiple-antibiotic-resistant staphylococci were isolated from different sites, and 325 of 411 (80%) staphylococcal isolates were resistant to two or more antibiotics. Coagulase-negative staphylococci are considered to be less virulent than *S. aureus* (Livermore, 2000); however, the isolation of a wide range of multiple-antibiotic-resistant coagulase-negative staphylococci in the current study gives cause for concern. There was no correlation between individual species and antibiotic resistance patterns; however, it was noticeable that more than 60% of staphylococci isolates showed resistance to penicillin and fusidic acid. In addition, 30 multiple-antibiotic-resistant

staphylococci were closely related to 30 susceptible staphylococci, respectively, indicating that these might belong to the same clone. Kraemer and Iandolo (1990) reported the transfer of antibiotic-resistant genes between species or inter-species, which may be a contributory factor in the development of different antibiotic resistance patterns in closely related isolates.

The studies reported here add credibility to the capacity of MALDI-TOF MS to delineate staphylococcal isolates derived from a wide range of environmental sites to the species level. Library- and bioinformatics-based approaches have been developed for MALDI-TOF MS microbial profiling. MASCOT and MATLAB are among the early developments that have been used to aid protein identification by MALDI-TOF (Lasch *et al.*, 2010; Fagerquist *et al.*, 2006). Access to various algorithms and data analysis software has enabled more in-depth analysis of MALDI-TOF data of various staphylococcal subpopulations, particularly in relation to methicillin resistance, multiple antibiotic resistance or toxin-producing strains. Artificial neural network (ANN) analysis has been used to trace the transition of methicillin resistance in subpopulation of *S. aureus* using a variation of MALDI-TOF MS involving pre-capture of proteins (Shah *et al.*, 2011).

Unlike species identification, where the entire mass spectrum may be used to characterize a particular taxon, methods aimed at distinguishing specific genotypes often rely on one or two low-abundance mass ions that may vary between laboratories and may therefore produce conflicting outcomes. The ability to distinguish MRSA from MSSA using MALDI-TOF MS, combined with simple statistical analysis has been widely reported but still remain divisive (Edwards-Jones *et al.*, 2000; Du *et al.*, 2002). The latter worked on the premise that the MALDI-TOF spectral profile between MRSA and MSSA differed from each other. Du *et al.* (2002) developed a database comprising the combined spectra of each test group, using the nuc-based PCR test to validate the method. Despite some discrepancies, the authors concluded that MALDI-TOF MS spectral profiles of each provide a simple and rapid method for identification and antibiotic susceptibility analysis of *S. aureus.* At a more subtle level, isogenic strains of *Staphylococcus aureus* differing in their expression of resistance to methicillin or teicoplanin were analyzed by MALDI-TOF MS. The study reported greater differences among strains differing in methicillin than in teicoplanin resistance (Majcherczyk *et al.*, 2006).

Data analysis software such as BioNumerics enables deeper insight to be obtained into species diversity and its potential to follow the transmission of isolates (Stephanie *et al.*, 2012). Similarly, Bruker's ClinProTools is being used more frequently now for studying diversity within species and typing (see Chapter 6). Wang *et al.* (2013) set out to establish a MALDI-TOF MS method to differentiate MRSA from MSSA in a manner similar to the those reported previously (Edwards-Jones *et al.*, 2000; Du *et al.*, 2002) but using ClinProTools to analyze the data. Although the majority of strains were separated into two groups, there was an overlap of strains between the groups. In a later study, Wolters *et al.* (2011) utilized *Staphylococcus aureus* strain type USA300 to create a model using ClinProTools software based on three mass ions 5932, 6423 and 6592. The model was found to be discriminatory between MRSA and MSSA strains and, although perfect differentiation was not achieved, it showed a high degree of specificity.

It is clear that the differences between the mass spectra of MRSA and MSSA are so small that when studies are expanded to include large numbers of clinical isolates, the

results are never clear-cut. In an attempt to resolve this, Josten *et al.* (2013) utilized MALDI-TOF MS to investigate the epidemic lineages of *S. aureus.* It was shown elegantly that unique mass signals were derived mainly from stress and ribosomal proteins. Peak shifts that differentiated the main *S. aureus* clonal complexes CC5, CC22, CC8, CC45, CC30 and CC1 correlated with point mutations in the relevant genes. It was therefore possible to differentiate unrelated MSSA, MRSA and borderline resistant *S. aureus* strains isolated from health care workers.

Despite the resourcefulness and refinement of current software in discriminating minor differences in mass ions of different spectra, results to date would suggest that linear MALDI-TOF MS does not have sufficient resolution to provide comprehensive intra-species maps of staphylococci and at best would be better served by addressing very specific problems. For example, Fadi *et al.* (2009) investigated the detection of Panton–Valentine leukocidin (PVL) versus non-PVL-producing *S. aureus* using MALDI-TOF MS. The study identified a unique mass ion at 4448 Da and used this to differentiate between toxin and non-toxin producers of *S. aureus*. The method was cross-validated with a success of 77% and enabled cases of PVL-producing strains to be detected within a few minutes. The authors were able to perform this in real time using ClinProTools™ 2.0 and were sufficiently confident to offer this as a point-of-care method. Similarly, Decristophoris *et al.* (2011) focussed specifically on resolving the *Staphylococcus intermedius* group that includes *S. intermedius*, *S. pseudintermedius* and *S. delphini*, coagulase-positive bacteria commonly isolated from animals and for which members of the group are identified using molecular methods such as *hsp60* gene sequencing. MALDI-TOF MS clearly differentiated members of the group and showed good congruence with *hsp60* gene sequencing. By contrast, Lasch *et al.* (2014) utilized manual peak inspection with pseudo-gel views, unsupervised hierarchical cluster analysis and supervised ANN analysis and were unable to unequivocally differentiate phylogenetic lineages, clonal complexes or sequence types for *S. aureus* and attributed it to 'insufficient discriminatory power of MALDI-TOF mass spectrometry'. New developments in MALDI-TOF MS, TOF/TOF and others such as top-down proteomics are on the horizon, and perhaps *S. aureus* may again be used to help develop new mass spectral approaches that can be unequivocally applied for typing strains.

References

Andrews, J. M., and Howe, R. A. (2011). BSAC standardized disc susceptibility testing method (version 10). *J. Antimicrob. Chemother.* 66, 2726–2757.

Carbonnelle, E., Beretti, J. L., Cottyn, S., Quesne, G., Berche, P., Nassif, X., and Ferroni, A. (2007). Rapid identification of Staphylococci isolated in clinical microbiology laboratories by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 45, 2156–2161.

Claydon, M. A., Davey, S. N., Edwards-Jones, V., and Gordon, D. B. (1996). The rapid identification of intact microorganisms using mass spectrometry. *Nat. Biotechnol.* 14, 1584–1586.

Decristophoris, P., Fasola, A., Benagli, C., Tonolla, M., and Petrini, O. (2011). Identification of *Staphylococcus intermedius* Group by MALDI-TOF MS ☆. *Syst. Appl. Microbiol.* 34, 45–51.

Diekema, D. J., Pfaller, M. A., Schmitz, F. J., Smayevsky, J., Bell, J., Jones, R. N., and Beach, M. (2001). Survey of infections due to *Staphylococcus* species: Frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific Region for the SENTRY Antimicrobial Surveillance. *Clin. Infect. Dis.* 32, S114–S132.

Drucker, D. B., and Abdullah, N. (1995). Polar lipids of *Staphylococcus* strains analysed by fast atom bombardment mass spectrometry. *J. Appl. Microbiol.* 79, 219–224.

Du, Z., Yang, R., Guo, Z., Song, Y., and Wang, J. (2002). Identification of *Staphylococcus aureus* and determination of its methicillin resistance by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.* 74, 5487–5491.

Edwards-Jones, V., Claydon, M. A., Evason, D. J., Walker, J., Fox, A. J., and Gordon, D. B. (2000). Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. *J. Med. Microbiol.* 49, 295–300.

Etienne, C., Cécile, M., Emmanuelle, B., Nesrine, D., Brunhilde, D., Jean-Luc, B., Agnès, F., Laurent, G., and Xavier, N. (2011). MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clin. Biochem.* 44, 104–109.

Fadi, B., Zoulikha, O., Farida, S., Didier, R., and Jean-Marc, R. (2009). MALDI-TOF-MS for rapid detection of staphylococcal Panton–Valentine leukocidin. *Int. J. Antimicrob. Agents* 34, 467–470.

Fagerquist, C. K., Bates, A. H., Heath, S., King, B. C., Garbus, B. R., Harden, L. A., and Miller, W. G. (2006). Sub-speciating *Campylobacter jejuni* by proteomic analysis of its protein biomarkers and their post-translational modifications. *J. Proteome Res.* 5, 2527–38.

Harris L. G., El-Bouri, K., Johnston, S., Rees, E., Frommelt, L., Siemssen, N., Christner, M., Davies, A. P., Rohde, H., and Mack, D. (2010). Rapid identification of staphylococci from prosthetic joint infections using MALDI-TOF mass-spectrometry. *Int. J. Artif. Organs*. 33, 568–574.

Josten, M., Reif, M., Szekat, C., Al-Sabti, N., Roemer, T., Sparbier, K., Kostrzewa, M., Rohde, H., Sahl, H. G., and Bierbaum, G. (2013). Analysis of the MALDI-TOF mass spectrum of *Staphylococcus aureus* identifies mutations which allow differentiation of the main clonal lineages. *J. Clin. Microbiol.* 51, 1809–1817.

Kamal, R. M., Bayoumi, M. A., and Abd El Aal, S. F. A. (2013). MRSA detection in raw milk, some dairy products and hands of dairy workers in Egypt, a mini-survey. *Food Control.* 33, 49–53.

Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., Mckenna, T., Mcdowall, M., and Shah, H. N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 4, 221–242.

Kraemer, G. R., and Iandolo, J. J. (1990). High-frequency transformation of*Staphylococcus aureus* by electroporation. *Curr. Microbiol.* 21, 373–376.

Lancashire, L., Schmid, H. O., and Ball, G. (2005). Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. *Bioinformatics* 21, 2191–2199.

Lasch, P., Drevinek, M., Nattermann, H., Grunow, R., Stämmler, M., Dieckmann, R., Schwecke, T., and Naumann, D. (2010). Characterization of Yersinia using MALDI-TOF mass spectrometry and chemometrics. *Anal. Chem.* 82, 8464–8475.

Lasch, P., Fleige, C., Stämmler, M., Nübel, U., Witte, W., and Werner, G. (2014). Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of *Enterococcus faecium* and *Staphylococcus aureus* isolates. *J. Microbiol. Methods* 100, 58–69.

Livermore, D. M. (2000). Antibiotic resistance in staphylococci. *Int. J. Antimicrob. Agents* 16 Suppl 1, S3–10.

Lowy, F. (1998). *Staphylococcus aureus* infections. *N. Engl. J. Med.* 339, 520–532.

Ma, E. S. K., Wong, C. L. P., Lai, K. T. W., Chan, E. C. H., Yam, W. C., and Chan, A. C. W. (2005). Kocuria kristinae infection associated with acute cholecystitis. *BMC Infect. Dis.* 5, 60.

Majcherczyk, P. A., McKenna, T., Moreillon, P., and Vaudaux, P. (2006). The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant *Staphylococcus aureus*. *FEMS Microbiol. Lett.* 255, 233–239.

Meers, P., Whyte, W., and Sandys, G. (1975). Coagulase-negative staphylococci and micrococci in urinary tract infections. *J. Clin. Pathol.* 28, 270–273.

Monecke, S., Müller, E., Dorneanu, O. S., Vremeră, T., and Ehricht, R. (2014). Molecular typing of MRSA and of clinical *Staphylococcus aureus* isolates from Iasi, Romania. *PLoS ONE.* 9, e97833.

De Neeling, A. J., van den Broek, M. J. M., Spalburg, E. C., van Santen-Verheuvel, M. G., Dam-Deisz, W. D. C., Boshuizen, H. C., van de Giessen, A. W., van Duijkeren, E., and Huijsdens, X. W. (2007). High prevalence of methicillin resistant *Staphylococcus aureus* in pigs. *Vet. Microbiol.* 122, 366–372.

Normanno, G., Corrente, M., La Salandra, G., Dambrosio, A., Quaglia, N. C., Parisi, A., Greco, G., Bellacicco, A. L., Virgilio, S., and Celano, G. V (2007). Methicillin-resistant *Staphylococcus aureus* (MRSA) in foods of animal origin product in Italy. *Int. J. Food Microbiol.* 117, 219–222.

Oliveira, D. C., and Tomasz, A. L. H. (2002). Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant *Staphylococcus aureus*. *Lancet Infect. Dis.* 2, 180–189.

Otto, M. (2009). *Staphylococcus epidermidis* – the "accidental" pathogen. *Nat. Rev. Microbiol.* 7, 555–567.

Pantucek, R. (2005). *Staphylococcus simiae* sp. nov., isolated from South American squirrel monkeys. *Int. J. Syst. Evol. Microbiol.* 55, 1953–1958.

Pratt, R. J., Pellowe, C., Loveday, H. P., Robinson, N., Smith, G. W., Barrett, S., Davey, P., Harper, P., Loveday, C., McDougall, C. *et al.* (2001). The epic project: Developing national evidence-based guidelines for preventing healthcare associated infections. Phase I: Guidelines for preventing hospital-acquired infections. Department of Health (England). *J. Hosp. Infect.* 47 Suppl, S3–82.

Samb-Ba, B., Mazenot, C., Gassama-Sow, A., Dubourg, G., Richet, H., Hugon, P., Lagier, J.-C., Raoult, D., and Fenollar, F. (2014). MALDI-TOF identification of the human Gut microbiome in people with and without diarrhea in Senegal. *PLoS ONE.* 9, e87419.

SCHAFER, F. (1979). Infectious endocarditis caused by *Rothia dentocariosa*. *Ann. Intern.* 91, 747–748.

Seng, P., Rolain, J. M., Fournier, P. E., Scola, B. La, Drancourt, M., and Raoult, D. (2010). MALDI-TOF-mass spectrometry applications in clinical microbiology. *Future Microbiol.* 5, 1733–1754.

Sexton, T., Clarke, P., O'Neill, E., Dillane, T., and Humphreys, H. (2006). Environmental reservoirs of methicillin-resistant *Staphylococcus aureus* in isolation rooms: Correlation with patient isolates and implications for hospital hygiene. *J. Hosp. Infect.* 62, 187–194.

Shah, H. N., Keys, C. J., Gharbia, S. E., Ralphson, K., Trundle, F., Brookhouse, I., and Claydon, M. A. (2000). The application of matrix-assisted laser desorption/ionisation time of flight mass spectrometry to profile the surface of intact bacterial cells. *Microb. Ecol. Heal. Dis.* 12.

Shah, H. N., Rajakaruna, L., Ball, G., Misra, R., Al-Shahib, A., Min, F., and Gharbia, S. E. (2011). Tracing the transition of methicillin resistance in sub-populations of *Staphylococcus aureus,* using SELDI-TOF Mass Spectrometry and Artificial Neural Network Analysis. *Syst. Appl. Microbiol.* 34, 81–86.

Simões, R. R., Aires-de-Sousa, M., Conceição, T., Antunes, F., da Costa, P. M., and de Lencastre, H. (2011). High prevalence of EMRSA-15 in Portuguese public buses: a worrisome finding. *PLoS ONE.* 6, e17630.

Sogawa, K., Watanabe, M., Sato, K., Segawa, S., Miyabe, A., Murata, S., Saito, T., and Nomura, F. (2012). Rapid identification of microorganisms by mass spectrometry: Improved performance by incorporation of in-house spectral data into a commercial database. *Anal. Bioanal. Chem.* 403, 1811–1822.

Stephanie, S., Borror, C. M., and Sandrin, T. R. (2012). Automating data acquisition affects mass spectrum quality and reproducibility during bacterial profiling using an intact cell sample preparation method with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 26, 243–253.

Stoltenberg, S. F., and Nag, P. (2010). Identification of a variety of *Staphylococcus* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 48, 941–945.

Thouverez, M., Muller, A., Hocquet, D., Talon, D., and Bertrand, X. (2003). Relationship between molecular epidemiology and antibiotic susceptibility of methicillin-resistant *Staphylococcus aureus* (MRSA) in a French teaching hospital. *J. Med. Microbiol.* 52, 801–806.

Wang, Y. R., Chen, Q., Cui, S. H., and Feng Qin, L. I. (2013). Characterization of *Staphylococcus aureus* isolated from clinical specimens by matrix assisted laser desorption/ionization time-of-flight mass spectrometry. *Biomed. Environ. Sci.* 26, 430–436.

Wolters, M., Rohde, H., Maier, T., Belmar-Campos, C., Franke, G., Scherpe, S., Aepfelbacher, M., and Christner, M. (2011). MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. *Int. J. Med. Microbiol.* 301, 64–68.

21C

Elucidating the Intra-Species Proteotypes of *Pseudomonas aeruginosa* from Cystic Fibrosis

Ali Olkun, Ajit J. Shah and Haroun N. Shah

21C.1 The Emergence of *Pseudomonas aeruginosa* as Key Component of the Cystic Fibrosis Lung Flora

Pseudomonas aeruginosa is not a normal component of the human microflora but instead is ubiquitously found in the environment. Perhaps because of accessibility and predisposition for the CF lung, *P. aeruginosa* has emerged today as the primary CF pathogen among the polymicrobial flora and affects up to 75% of adults (see Chapter 17; Lipuma *et al.*, 2012; Lynch and Bruce, 2013; Rogers *et al.*, 2010). A recent study by Fodor *et al.* (2012) revealed that samples from patients who had Pseudomonas, Burkholderia or neither as the primary pathogen could be categorized into three distinct groups. However, when the primary pathogen was removed from the analysis, the samples merged into a single group. Indirectly, this has been attributed to the dependence of the underlying microbial community on the primary pathogen to shape the microbiome of an individual. A search for inflammatory markers to the CF microbiome profiles of 21 patients showed significant correlation with lower FEV_1, serum C-reactive protein and neutrophil elastase in sputum, again supporting the view for a primary role of *P. aeruginosa* in CF airways (Zemanick *et al.*, 2013). In a series of elegant experiments, culture-enriched molecular profiling assessed low-abundant species for their virulence and antibiotic susceptibilities, microbial communities were assembled in vitro or in animal models, and measurement of polymicrobial interactions was performed. The results showed that many species that would normally be considered benign had the capacity to heighten the pathogenicity of *P. aeruginosa* (Sibley *et al.*, 2011). In patients where there was no change in the bacterial load of *P. aeruginosa*, such polymicrobial interactions may explain the observed exacerbation of disease. Such indirect but compelling evidence suggests a prominent role for *P. aeruginosa* in the aetiology of cystic fibrosis.

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

21C.2 Diversity and Rational for Proteotyping

Found naturally in soil and aquatic environments, *P. aeruginosa* is also an important opportunistic pathogen for humans, animals and plants (Spiers *et al.*, 2000). It can produce severe infections in immunocompromised hosts and is the major factor for morbidity and mortality in cystic fibrosis patients (Govan and Deretic, 1996). Its remarkable metabolic versatility has been attributed to its ecological diversity. It is believed that the major pathogenic determinants of *P. aeruginosa* stems from the presence of several cell-associated and secreted virulence factors, such as elastase, exotoxin A, phospholipase and alkaline protease, among others (Döring *et al.*, 1987). A key step in the cytotoxic and invasion processes is its use of a type III secretion system to directly deliver several effector proteins into the cytoplasm of the host cell (Holder *et al.*, 2001; Yahr *et al.*, 1997). Interestingly, the genes encoding its virulence factors are dispersed on the bacterial chromosome unlike most species, in which they are clustered in pathogenicity islands (Stover *et al.*, 2000).

Among many opportunistic pathogens, particular subtypes often correspond to specific disease sites and ecological niches within the human body. Clonal types exhibit a range of traits that are monitored in epidemiological studies to identify particular virulent clones. Extensive studies on *P. aeruginosa* over the last 15 years, using some of the most powerful molecular techniques, have consistently failed to demonstrate any specific ecotypes that are preferentially transmitted from the environment. Thus, strains isolated from the environment are indistinguishable from clinical isolates in terms of several genotypic, taxonomic or metabolic properties (Alonso *et al.*, 1999; Cabrol *et al.*, 2003; Foght, *et al.*, 1996). The complete genome sequence of *P. aeruginosa* PA01 was reported in 2000, and a number of unique properties were revealed (Stover *et al.*, 2000). Analysis of further strain types have revealed a high degree of gene conservation even among its most pronounced virulence factors. Thus, strain-specific genes within a set of 18 strains isolated from clinical and nonclinical habitats revealed no correlation between genome content and infection type (Wolfgang *et al.*, 2003). However, once within the host, clonal types are selected, and these can be traced though epidemiological studies (Armstrong *et al.*, 2003).

To date, diversity has been studied mostly by DNA-based approaches. In a species such as *P. aeruginosa* that thrives in such varied ecosystems, protein expression is likely to play a key role as it traverses new habitats. This may be reflected in the selection of new proteotypes. The present study is part of larger study on the proteome of a broad selection of isolates of *P. aeruginosa* and seeks to shed light on the basis of pathogenicity of this species.

21C.3 Selecting Representative Strains for Profiling

The ubiquitous and resilient nature of *P. aeruginosa*, together with its huge intra-species diversity, may contribute to the selection of specific variants in various nosocomial infections. Accurate typing of the organism is critical for identification and monitoring sources of infection, chains of transmission as well as reservoirs for proliferation (Grundmann *et al.*, 1995) The gradual decline in lung function described above results in patients with CF becoming colonized by a variety *P. aeruginosa* strain types.

Most worthy of note are the Liverpool, Midlands, Manchester and Clone C strains – all associated with increased resistance and cross-infection (Fothergill *et al.*, 2012). The term *transmissible* has been applied to strains with increased prevalence among CF patients (Armstrong *et al.*, 2003). The ability of transmissible strains to survive outside the CF patient reservoir intensifies the need for informative and practical typing methods. Any such typing scheme must allow for large number of samples to be analyzed at a low cost, while accommodating inter-laboratory reporting and communication (Armstrong *et al.*, 2003; Fothergill *et al.*, 2012).

21C.4 Selection of Strains against a Background of Their Variable Number Tandem Repeat (VNTR) Designation

Bacterial genomes contain DNA sequence regions that vary in size, location, complexity and repeat mode that may be used for bacterial typing (Rogers and Döring, 2015). One such repeat unit, often from non-coding regions, contains enough discriminatory information to differentiate between isolates (Roring *et al.*, 2002). These are referred to as VNTRs. The preference for typing *P. aeruginosa* at the Antimicrobial Resistance and Healthcare Associated Infections (AMRHAI) Labs, PHE, is VNTR analysis (Turton *et al.,* 2010). The scheme characterizes isolates by the number of repeats at each locus and is a widely used typing method for a variety of pathogenic bacteria (Vu-Thien *et al.*, 2007). Over the last decade, the scheme has been reviewed and fine-tuned to deliver robust, reliable results (Lindstedt *et al.*, 2012). VNTR assays are dependent on PCR amplification for a number of loci, a process termed multiple-locus VNTR analysis (MVLA). The results for each strain are described by their confirmation for a number of repeats at each selected locus, which allows for further downstream analysis and corroboration (Vu-Thien *et al.*, 2007). The initial setup of the scheme at AMRHAI was constructed on 12 VNTR loci (ms61, ms77, ms142, ms172, ms207, ms209, ms211, ms212, ms213, ms214, ms217 and ms222) as described by Onteniente *et al.* (2003) and Vu-Thien *et al.* (2007), but reduced to 8 by Turton *et al.* (2010) with no loss of discrimination. Consequently, a scheme involving these loci (ms61, ms172, ms207, ms209, ms211, ms214, ms217 and ms222) was adopted, with the inclusion of a single extra locus (ms213). The ability of VNTR to discriminate between *P. aeruginosa* isolates that have identical PFGE profiles (Turton *et al.*, 2010) demonstrates the rationale for implementing the scheme for subtyping: suitability for online sharing and storage of data. The system is less cumbersome, more rapid and reproducible and has therefore superseded PFGE (Turton *et al.*, 2010). However, despite these advantages, the scheme is nevertheless complex, mostly restricted to the specialist's reference laboratory and would benefit from a more simple, accessible method of subtyping.

21C.5 Potential to Type *P. aeruginosa* using MALDI-TOF MS

Although MALDI-TOF MS is now a well-established method for microbial species identification, its use as a typing method has been restricted to few taxa. It has been investigated as a typing method for several species such as *Listeria* spp., enterococci,

Acinetobacter baumannii, *Klebsiella pneumonia* (see, e.g., Barbuddhe *et al.*, 2008; Berrazeg *et al.*, 2013; Griffin *et al.*, 2012) and *Escherichia coli* clonal groups (Matsumura *et al.*, 2014). In general, spectral profiles, using, for example, Bruker's Biotyper, are often represented by spatial diagrams derived by principal component analysis or hierarchical clustering based on dendrograms. Here, we assess the potential of the technique to type strains of *Pseudomonas aeruginosa* that were previously subtyped using VNTRs. The sample set of 53 clinical isolates included transmissible strain sets, as well as those with resistance to antibiotics. The VNTR analysis was performed and provided by AMRHAI according to the scheme reported by Turton *et al.* (2009) and Jane Turton (pers. comm.). The selection represented geographical distribution, as well as being both CF and non-CF in origin. Each sample was unique to a person, with patient age varying across all age groups, including infants. For MALDI-TOF MS (Bruker), all samples were grown overnight at 37 °C on Muller-Hinton agar and spectra generated on the Bruker Autoflex MALDI-TOF MS instrument (Bruker Daltonics), and recorded in the linear mode. The laser frequency was 20 Hz, and mass ions between 2,000 and 20,000 Da were collected. Calibration was performed using *Escherichia coli* DH5 alpha standard peaks in the form of a bacterial test standard. Only the peaks with a signal/noise ratio ≥ 2.0 were accepted and converted to a log (score) from 0 (no spectra) to 3 (absolute match). The log (score) values were acquired by comparing to the matches in the database entries on the MALDI Biotyper Library (MBL). The cut-off scores were at 2.0 and above to qualify a sample as a positive species-level identification. The consensus for a dendrogram generation revolves around peak profiling, and for this, the window for spectra analysis was reduced to m/z 3,000–15,000 Da in order to exclude abundant, non-discriminatory peaks outside those values.

21C.6 Data Processing: Analyzing Data using BioNumerics 7

After executing the automated Biotyping process, raw data was exported from the Autoflex instrument as text files. These files retained the signal intensity for each 0.5 m/z value. A Data Import Script (Applied Maths, Belgium) was used to import peak lists of each of the samples from one spot set, to separate text files, which were used as input files for the BioNumerics 7 software package (Applied Maths, Belgium; see Section 21A above). Post import, the data pre-processing involved an initial baseline subtraction, consecutive continuous wavelet transform (CWT) noise estimation and baseline subtraction using a rolling-disc algorithm (with the size of 200 points). Each peak with a signal-to-noise (S/N) ratio (minimum 10) and absolute intensity was annotated. The data patterns were saved in the form of (characteristic patterns) peak list and sizes as well as signal-to-noise ratios. For each replicate, both technical and biological replicates were combined into a single summary spectral profile. A peak-matching analysis was conducted with constant and linearly varying tolerance values of 1 m/z and 800 ppm, respectively. The minimum peak detection rate was set at 100%, meaning that each summary peak occurred in each individual spectrum of the technical and biological replicates. The use of the 100% peak detection rate during this summarizing procedure excluded any technical or biological variation from the analysis. The signal intensity for each data point in the Super Spectrum was calculated by averaging the respective signal intensities in the technical and biological replicates. The species

identification by MALDI-TOF MS performed by the extraction method (as opposed to the direct colony approach) was significantly higher (data not shown).

Dendrograms of MALDI-TOF MS and VNTR are shown in Figures 21C.1 and 21C.2, and their discordance is shown in Figure 21C.3. Assuming a cut-off value at 50% for comparing the clustering on the congruence chart, it is evident that there is no equivalence between geographical groups (represented using different colours for each hospital sample) as well as the overall distribution of each strain type. The Spearman rank correlation coefficient was calculated at 0.135 ($P < 0.001$) between VNTR and the MALDI-TOF MS dendrogram, indicating a very weak correlation. A measurement of closeness was not performed because no real significance between both methods could be demonstrated, nullifying the need to further query the depth of an absent relationship.

21C.7 Discussion and Data Interpretation

The rationale for exploring MALDI-TOF MS as a tool for typing is logical because it seeks to take advantage of data already generated from microbial identification of a strain. Various studies have justified comparative approaches with varying degrees of success (Ayyadurai *et al.*, 2010; Stephan *et al.*, 2011; Shitikov *et al.*, 2012; Bader, 2013; Berrazeg *et al.*, 2013; Matsumura *et al.*, 2014), and several pipelines have been reported to interrogate MALDI-TOF MS data using, for example, Clin ProTools (Bruker – see Chapter 6) or MS/MS data (see, e.g., Al-Shahib *et al.*, 2010). Here the primary aim was to incorporate the BioNumerics software into existing workflows. It was anticipated that the high resolution of DNA-based methods such as VNTRs would not corroborate data derived from MALDI-TOF MS because the former depends on variation of loci, whereas MALDI-TOF MS relies on the conserved, high-abundance ribosomal proteins. However, the comparison was made to access the relative intra-species diversity that would be revealed by MALDI-TOF MS against that of VNTR as a reference. Figure 21C.1 shows that at an arbitrarily taken level of 50%, eight clusters were derived that increased threefold at a 90% similarity level, indicating that this method is about 30% as efficient as VNTR analysis in terms of revealing intra-species diversity.

Figure 21C.4, demonstrates the steps involved in the MALDI-TOF MS workflow as a hierarchical-free approach. Instead of building up a set of heavily queried reference spectra, acquired from in-house commands or manual acquisition, the approach here was to maintain the integrity of recommended MSP acquisition (Bruker MSP generation) to ensure that the translational value of the findings would be maintained, and Figure 21C.1 shows that this could be reliably achieved and integrated into the current workflow. In similar studies, Berrazeg *et al.* (2013) succeeded in phenotypically differentiating the geographical distribution of isolates as opposed to assessing the credibility of MALDI-TOF MS typing against other sequencing techniques, while Matsumura *et al.* (2014) distinguished the clonality of *Escherichia coli* strains associated with increased virulence using multi locus sequences typing (MLST) as a reference.

The major obstacle to incorporating spectra from the Biotyper output relates to the spectral intensity (Gekenidis *et al.*, 2014). By design, a spectral matching software (Biotyper) is built around the concept of turnaround times; once sufficient spectral integrity is accumulated, the instrument moves onto the next target because an adequate

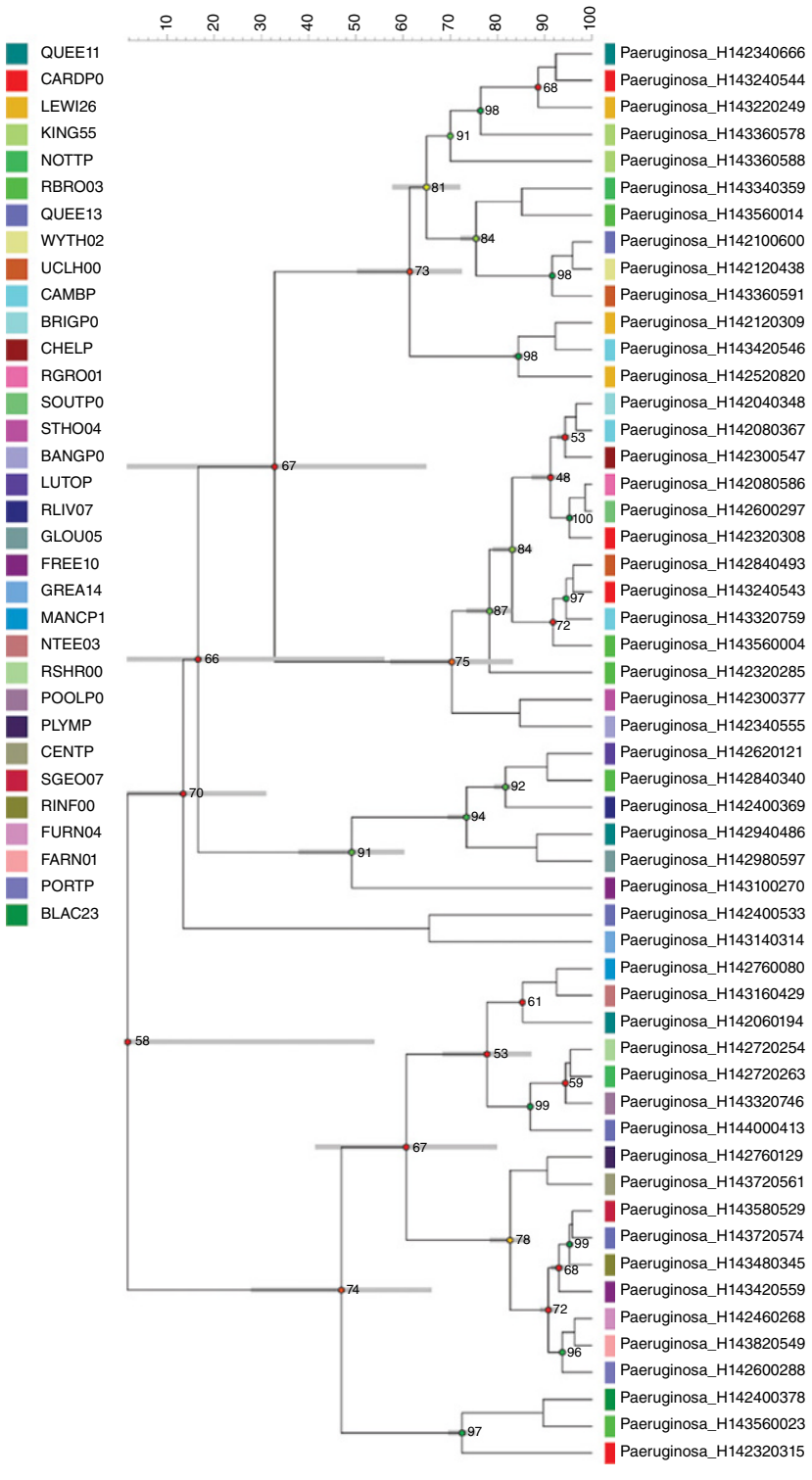


Figure 21C.1 MALDI-TOF MS dendrogram showing the similarity and relationship between 53 strains of *P. aeruginosa* from both CF and non-CF patients. The 33 coloured nodes are representative of geographical distribution from which the samples were sourced. With a cut-off value of 50%, the results show eight clusters with no geographical relationship between isolate groupings.

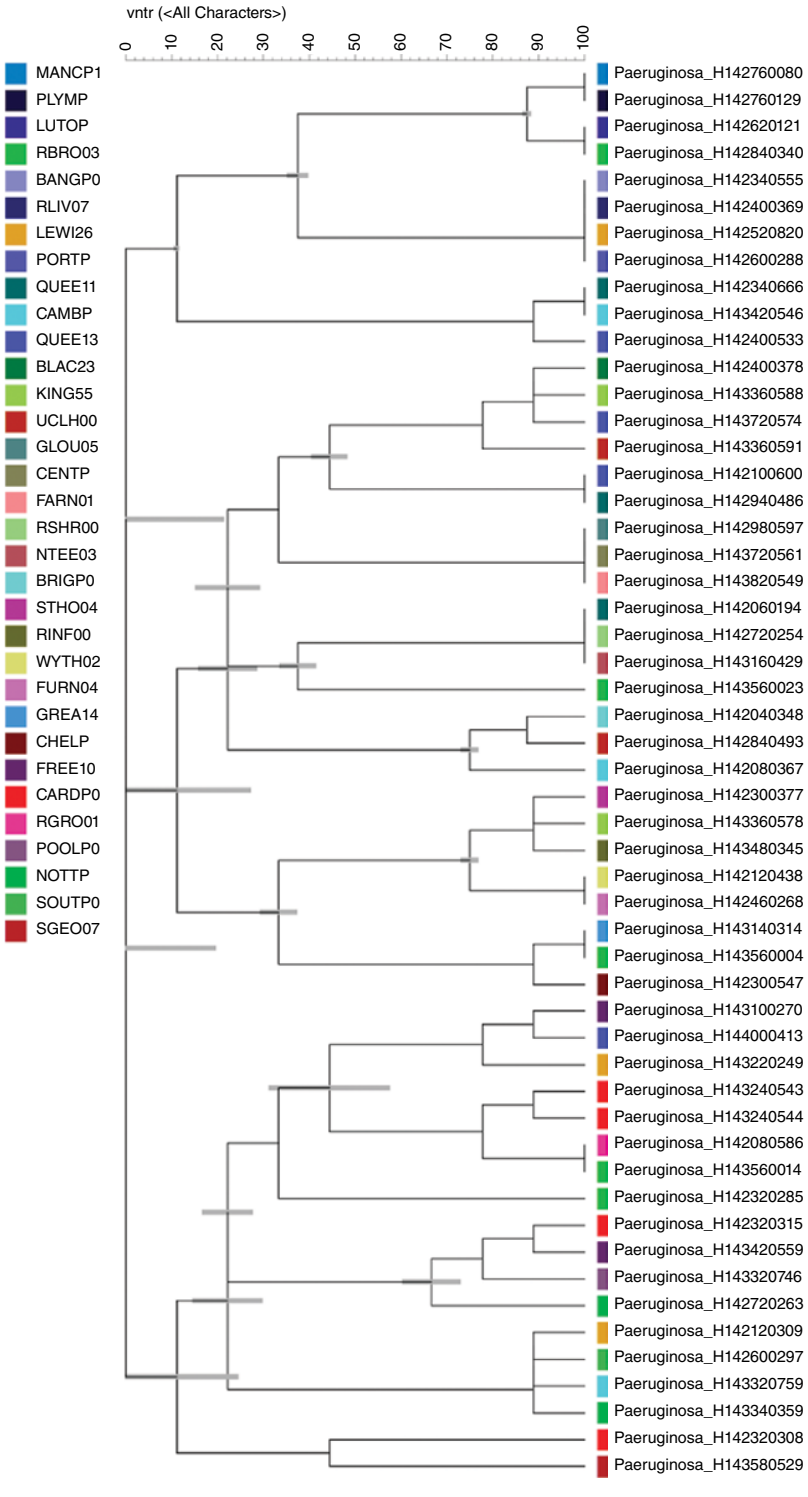


Figure 21C.2 VNTR dendrogram based on nine loci which also assess the similarity and distancing between 53 CF and non-CF strains of *P. eruginosa*. As in Figure 21C.1, the 33 coloured nodes represent geographical distribution from which the samples were sourced. The results show 18 clusters also cut-off at 50%, with no geographical relationship between isolate groupings.

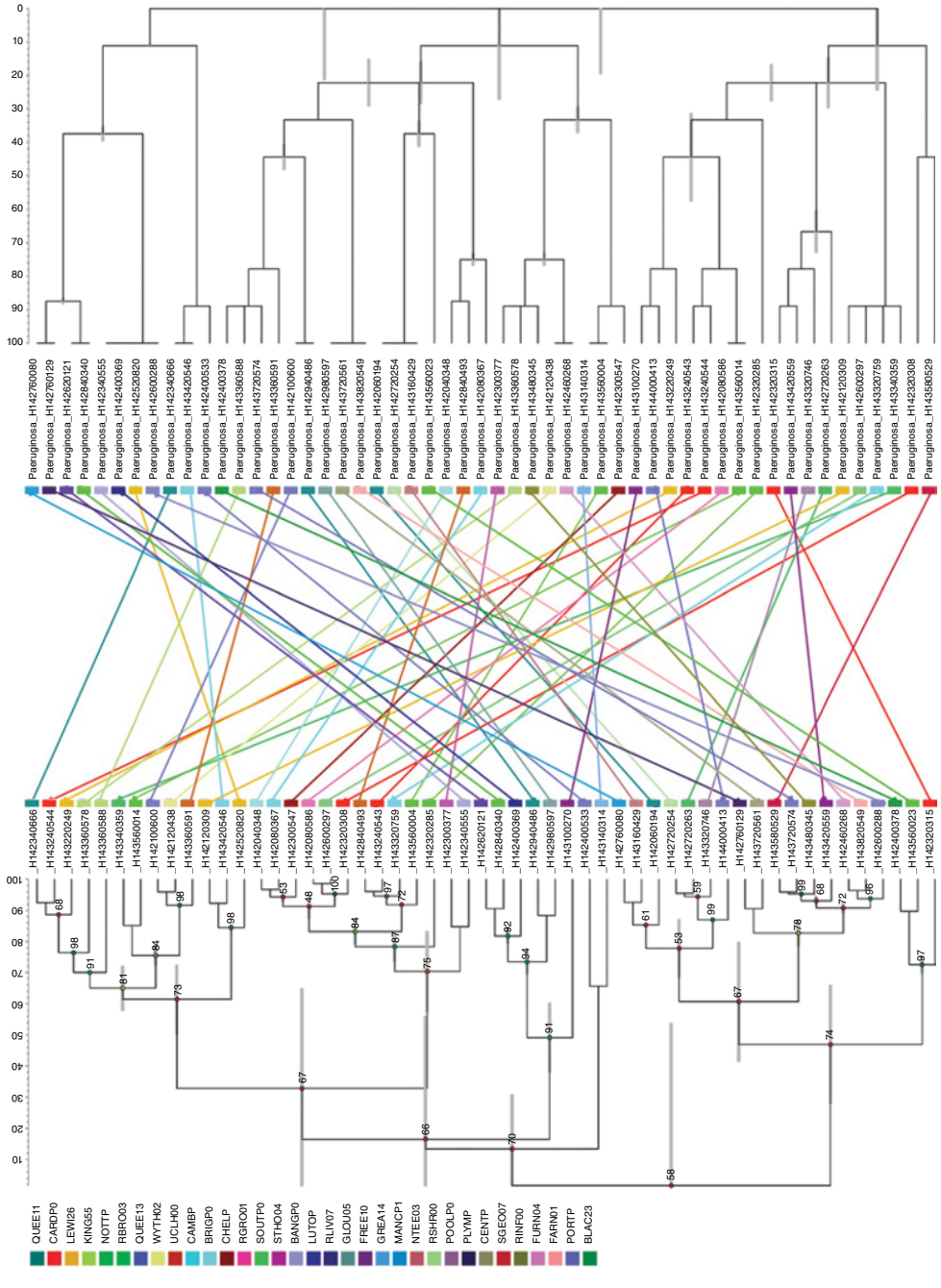


Figure 21C.3 Congruence chart. Comparison between the MALDI-TOF MS dendrogram (left) and VNTR (right). Coloured lines that match geographical groups are used to assess the equivalence between nodes. This chart is combines both MALDI-TOF MS (Figure 21C.1) and VNTR dendrograms (Figure 21C.2). The cut-off value of 50% demonstrates no equivalence. The Spearman rank correlation coefficient between the two approaches was calculated at 0.135 ($P<0.001$). The congruence between clade distributions also fails to demonstrate any meaningful relationship within the overall grouping of samples.

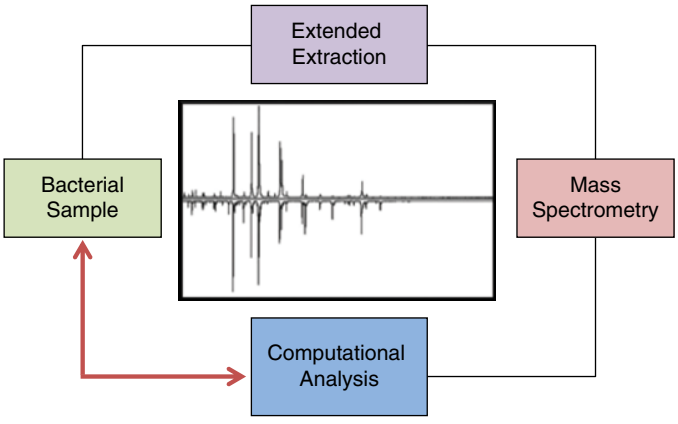


Figure 21C.4 Outline of MALDI-TOF MS workflow. The bacterial sample is prepared for analysis using a protein extraction method prior to analysis on the mass spectrometry platform (Bruker autoflex). Post analysis, the spectra generated is compared to a database (Biotyper), where pattern matching allows for rapid species-level identification.

match criterion has been satisfied. For a phyloproteomic approach as reported by Galzitskaya and Lobanov (2015), the conceptual methodology is at variance with this because spectra need to be accumulated with significant intensity as well as the number of acquisitions in order to fully explore minor differences between closely related species (Starostin *et al.*, 2015). The ability to distinguish closely related species is tied to dependency upon the physiological status of the cell such as environmental stress, growth phase, nutrients, and so on, because such parameters alter protein levels and post-translational modifications (Keys *et al.*, 2004; Chierico *et al.*, 2014). Protocols therefore need to be rigorously standardized to minimize noise introduced into the data, suggesting that samples processed for typing require more rigor than those used for identification.

21C.8 Going Forward – Reproducibility the Salient Determinant

Studies are now under way to undertake 'blind analysis' to confirm that the minor components of MALDI-TOF MS spectral data are robust and reproducible for typing. The overall approach is to assess the potential of BioNumerics to flag errors and establish confidence levels. Instead of peak list, raw data is being used to ensure there is no bias in the procedure. In the first instance, unsupervised analysis is being used across a range of datasets. If this can be established, it will provide more flexibility to the system.

In most published work, supervised learning techniques have been used to analyze MALDI-TOF MS data based on known phenotypes in which specific biomarkers can be discerned. These are then used to train a classifying algorithm, after which unknown isolates can be interrogated. This has been highly successful for SELDI-TOF MS because a larger range of biomarkers is available for analysis (see, e.g., Lancashire *et al.*, 2005; Schmid *et al.*, 2005; Lundquist *et al.*, 2005; Lasch *et al.*, 2009; Shah *et al.*, 2011). Any

attempt to develop a phyloproteomic database for typing organisms needs to provide confident (Christner *et al.*, 2014), translational data where another laboratory running the same set of strains demonstrates identical typing (Del Chierico *et al.*, 2014). Singhal *et al.* (2015) summarized numerous studies where bacterial typing has been reported. The ability to reproducibly challenge phenotypic patterns will be the essential requirement on any proposed typing system. Starostin *et al.* (2015) applied a computational analysis model to identify *Bacillus pumilus* from their MALDI spectra and showed how geometric distancing can be used to create mathematical algorithms to define profiles and calculate spectral similarity. The authors propose an algorithm for representing mass spectra as vectors in a multidimensional Euclidean space. The issue of reproducibility is resolved by constructing an open access database. By profiling proteins on an open platform, new data can be added; to overcome the limitation of lockdown databases managed by instrument manufacturers (Christner *et al.*, 2014; Gekenidis *et al.*, 2014). The translational maxim of these algorithms to profile mutable organisms such as *Pseudomonas aeruginosa* which has the ability to switch proteomes (seasile state to biolfilm), if achieved, will have a major impact on MALDI-TOF-based typing. Zautner *et al.* (2015) demonstrated the accomplishment of mass-spectrometry-based PhyloProteomics (MSPP) by calculating allelic isoforms from genome sequences of reference spectra. The authors identified masses of proteins to be used as biomarkers by calculating amino acid sequence list of allelic isoforms caused by non-synonymous mutations in biomarker genes to detect these as mass shifts in an overlay of calibrated MALDI-TOF spectra. This approach is of particular interest because the translational issues are overcome by anchoring biomarkers from a DNA-based technique. The limitation of dealing with samples without genome sequences is clearly an issue, but with rapid whole genome sequencing being carried out by a variety of laboratories, this may not be a problem in the future (Land *et al.*, 2015).

The full prowess of MS tied to computational biology is beginning to be realized. The case for using MALDI-TOF MS in-house databases, translated to open access platforms, warrants further study but may require higher-resolution mass spectrometers. The mass spectral profiles produced by a linear MALDI-TOF MS are very limited for typing and require such a high degree of rigour that this would be difficult to correlate between laboratories. MALDI-TOF MS is very much operator dependent, and significant differences have been reported even among just a few laboratories performing ring tests (see Chapter 5). It is our view that MALDI-TOF MS typing may be useful for some taxa where the diversity is low and there are discernible biomarkers. We have shown that it is possible to mirror phyletic lineages among *Propionibacterium acnes* with such confidence that MALDI-TOF MS data was used recently for the first time to propose new subspecies (Dekio *et al.*, 2015; Chapter 5). However, for the majority of species, there is considerable evidence to show that this may not be possible (see review by Singhal *et al.*, 2015) and MS/MS-based approaches will be necessary.

References

Alonso, A., Rojo, F., and Martínez, J. L. (1999). Environmental and clinical isolates of *Pseudomonas aeruginosa* show pathogenic and biodegradative properties irrespective of their origin. *Environ. Microbiol.* 1: 421–430.

Al-Shahib, A., Misra, R., Ahmod, N., Fang, M., Shah, H. N., and Gharbia, S. E. (2010). Coherent pipeline for biomarker discovery using mass spectrometry and bioinformatics. *BMC Bioinformatics*, 11: 437.

Armstrong, D., Bell, S., Robinson, M., Bye, P., Rose, B., Harbour, C., Lee, C., Service, H., Nissen, M., Syrmis, M., and Wainwright, C. (2003). Evidence for spread of a clonal strain of *Pseudomonas aeruginosa* among cystic fibrosis clinics. *J. Clin. Microbiol.* 41: 2266–2267.

Ayyadurai, S., Flaudrops, C., Raoult, D., and Drancourt, M. (2010). Rapid identification and typing of *Yersinia pestis* and other *Yersinia* species by matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry. *BMC Microbiol.* 10: 285 10.1186/1471-2180-10-285

Bader, O. (2013). MALDI-TOF-MS-based species identification and typing approaches in medical mycology. *Proteomics* 13: 788–799.

Barbuddhe, S. B., Maier, T., and Schwarz, G. (2008). Rapid identification and typing of Listeria species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* 74: 5402–5407.

Berrazeg, M., Diene, S. M., Drissi, M., Kempf, M., Richet, H., Landraud, L. *et al*. (2013). Biotyping of multidrug-resistant *Klebsiella pneumoniae* clinical isolates from France and Algeria using MALDI-TOF MS. *PLoS ONE.* 8(4): e61428. doi:10.1371/journal. pone.0061428

Cabrol, S., Olliver, A., Pier, G. B., Andremont, A., and Ruimy, R. (2003). Transcription of quorum-sensing system genes in clinical and environmental isolates of *Pseudomonas aeruginosa*. *J. Bacteriol.* 185: 7222–7230.

Chierico, F., Petrucca, A., Vernocchi, P., Bracaglia, G., Fiscarelli, E., Bernashi, P., Muraca, M., Urbani,A., and Putignani, L. (2014). Proteomics boosts translational and clinical microbiology. *J. Proteomics.* 97: 69–87.

Christner, M., Trusch, M., Rohde, H., Kwiatkowski, M., Schlüter, H., Wolters, M., Aepfelbacher, M., and Hentschke M. (2014). Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-toxigenic *Escherichia coli*. *PLoS ONE* 9(7): e101924.

Dekio, I., Culak, R. A., Misra, R., Gaulton, T., Fang, M., Sakamoto, M., Oshima, K., Hattori, M., Klenk, H.-K., Rajendram, D., Gharbia, S. E., and Shah, H. N. (2015). Dissecting the taxonomic heterogeneity within *Propionibacterium acnes*: Proposal for *Propionibacterium acnes* subsp. *acnes* subsp. nov. and *Propionibacterium acnes* subsp. *elongatum* subsp. nov. *Int. J. System. Evol. Microbiol.* 65: 4776–4787.

Del Chierico, F., Petrucca, A., Vernocchi, P., Bracaglia, G., Fiscarelli, E., Bernaschi, P., Muraca, M., Urbani, A., and Putignani, L. (2014). Proteomics boosts translational and clinical microbiology. *J. Proteomics.* 97: 69–87.

Döring, G., Maier, M., Muller, E., Bibi, Z., Tümmler, B., and Kharazmi, A. (1987). Virulence factors of *Pseudomonas aeruginosa*. *Antibiot. Chemother.* 39: 136–148.

Fodor, A. A., Klem, E. R., Gilpin, D. F, Elborn, J. S., Boucher, R. C., Tunney, M. M., and Wolfgang, M. C. (2012). The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations. *PLoS ONE.* 7: e45001

Foght, J. M., Westlake, D. W., Johnson, W. M., and Ridgway, H. F. (1996). Environmental gasoline-utilizing isolates and clinical isolates of *Pseudomonas aeruginosa* are taxonomically indistinguishable by chemotaxonomic and molecular techniques. *Microbiology.* 142: 2333–2340.

Fothergill, J. L., Walshaw, M. J., and Winstanley, C. (2012). Transmissible strains of *Pseudomonas aeruginosa* in cystic fibrosis lung infections. *Eur. Respir. J.* 40: 227–238.

Galzitskaya, O. V., and Lobanov, M. Y. (2015) "Phyloproteomic analysis of 11780 six-residue-long motifs occurrences," *BioMed Res. Int*., vol. 2015, Article ID 208346.

Gekenidis, M. T., Studer, P., Wüthrich, S., Brunisholz, R., and Drissner, D. (2014). Beyond the matrix-assisted laser desorption ionization (MALDI) biotyping workflow: in search of microorganism-specific tryptic peptides enabling discrimination of subspecies. *Appl. Environ. Microbiol.* 80: 4234–4241.

Govan, J. R., and Deretic, V. (1996). Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol. Rev.* 60: 539–574.

Griffin, P. M., Price, G. R., and Schooneveldt, J. M. (2012). Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak." *J. Clin. Microbiol.* 50: 2918–2931.

Grundmann, H., Schneider, C., Hartung, D., Daschner, F. D., and Pitt, T. L. (1995). Discriminatory power of three DNA-based typing techniques for *Pseudomonas aeruginosa*. *J. Clin. Microbiol.* 33: 528–534.

Holder, I. A., Neely, A. N., and Frank. D. W. (2001). Type III secretion/intoxication system important in virulence of *Pseudomonas aeruginosa* infections in burns. *Burns.* 27:129–130.

Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M., and Shah, H. N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect. Gen. Evol.* 4: 221–242.

Lancashire, L., Schmid, O., Shah, H. N., and Ball, G. (2005). Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. *Bioinformatics.* 21: 2191–2199.

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., and Ussery, D. W. (2015). Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics.* 15(2), 141–161.

Lasch, P., Beyer, W., Nattermann, H., Stämmler, M., Siegbrecht, E., Grunow, R., and Naumann D. (2009). Identification of *Bacillus anthracis* by using matrix-assisted laser desorption ionization-time of flight mass spectrometry and artificial neural networks. *Appl. Environ. Microbiol.* 75: 7229–7242.

Lindstedt, B. A., Torpdahl, M., Vergnaud,G., S Le Hello, S., Weill, F. X. *et al.* (2012). Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries. *Eurosurveillance*. 18(4): 20385. Pasteur-01111429.

LiPuma, J. J. (2012). The new microbiology of cystic fibrosis: It takes a community. *Thorax.* 67: 851–852.

Lundquist, M., Caspersen, M. B., Wikström, P., and Forsman, M. (2005). Discrimination of *Francisella tularensis* subspecies using surface enhanced laser desorption ionization mass spectrometry and multivariate data analysis. *FEMS Microbiol. Lett.* 243: 303–310.

Lynch, S. V., and Bruce, K. D. (2013). The cystic fibrosis airway microbiome. *Cold Spring Harb. Perspect. Med.* 3: a009738.

Matsumura, Y., Yamamoto, Y., and Nagao, M. (2014). Detection of extended-spectrum-beta-lactamase-producing *Escherichia coli* ST131 and ST405 clonal groups by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 52: 1034–1040.

Onteniente, L., Brisse, S., Tassios, P. T., and Vergnaud. G. (2003). Evaluation of the polymorphisms associated with tandem repeats for *Pseudomonas aeruginosa* strain typing. *J. Clin. Microbiol.* 41:4991–4997.

Rogers, G. B., Stressmann, F. A., Walker, A. W., Carroll, M. P., and Bruce, K. D. (2010). Lung infections in cystic fibrosis: Deriving clinical insight from microbial complexity. *Expert. Rev. Mol. Diagn.* 10: 187–196

Rogers, G., and Döring, G. (2015). New methods for detecting and identifying bacteria. In *Hodson and Geddes' Cystic Fibrosis*, Fourth Edition, pp. 186–194. 2015.

Roring, S., Scott, A., and Brittain, D. (2002). Development of variable-number tandem repeat typing of *Mycobacterium bovis*: Comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. *J. Clin. Microbiol.* 40: 2126–2133.

Schmid, O., Ball, G., Lancashire, L., Culak, R. A., and Shah, H. N. (2005). New approaches to identification of bacterial pathogens by surface enhanced laser desorption/ionisation time of flight mass spectrometry in concert with artificial neural networks, with special reference to *Neisseria gonorrhoeae*. *J. Med. Microbiol.* 54: 1205–1211.

Seibold, E., Maier, T., Kostrzewa, M., Zeman, E., and Splettstoesser, W. (2010). Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels. *J. Clin. Microbiol.* 48: 1061–1069.

Sibley, C. D., Grinwis, M. E., Field, T. R., Eshaghurshan, C. S., Faria, M. M., Dowd, S. E., Parkins, M. D., Rabin, H. R., and Surette, M. G. (2011). Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PLoS ONE.* 6: e22702.

Singhal, N., Kumar, M., Kanaujia, P. K. and Virdi, J. S. (2015). MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis. *Front. Microbiol.* 6: 791.

Shah, H. N., Rajakaruna, L., Ball, G., Misra, R., Al-Shahib, A., Fang., M., and Gharbia, S. E. (2011). Tracing the transition of methicillin resistance in subpopulations of *Staphylococcus aureus* using SELDI-TOF mass spectrometry and artificial neutral network analysis. *System. Appl. Microbiol.* 34: 81–86.

Shitikov, E., Ilina, E., Chernousova, L., Borovskaya, A., Rukin, I., Afanas'ev, M., Smirnova, T., Vorobyeva, A., Larionova, E., Andreevskaya, S., Kostrzewa, M., and Govorun, V. (2012). Mass spectrometry based methods for the discrimination and typing of mycobacteria. *Infect. Genet. Evol.* 12: 838–845.

Spiers, A. J., Buckling, A., and Rainey. P. B. (2000). The causes of *Pseudomonas* diversity. *Microbiology.* 146: 2345–2350.

Starostin, K. V., Demidov, E. A., Bryanskaya, A. V., Efimov, V. M., Rozanov A. S., and Peltek, S. E. (2015). Identification of Bacillus strains by MALDI TOF MS using geometric approach. *Sci. Rep.* 5, 16989.

Stephan, R., Cernela, N., Ziegler, D., Pflüger, V., Tonolla, M., Ravasi, D., Fredriksson-Ahomaa M., and Hächler H. (2011). Rapid species specific identification and subtyping of *Yersinia enterocolitica* by MALDI-TOF mass spectrometry. *J. Microbiol. Meth.* 87: 150–153.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yua, Y. N., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E., Lory, S., and Olson. M. V. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature.* 406: 959–964.

Turton, J. F., Turton, S. E., Yearwood, L., Yarde, S., Kaufmann, M. E., and Pitt, T. L. (2010). Evaluation of a nine-locus variable-number tandemrepeat scheme for typing of *Pseudomonas aeruginosa*. *Clin. Microbiol. Infect.*16: 1111–1116.

Vu-Thien, H., Corbineau, G., Hormigos, K., Fauroux, B., Corvol, H., Clément, A., Vergnaud, G., and Pourcel, C. (2007). Multiple-locus variable-number tandem-repeat analysis for longitudinal survey of sources of *Pseudomonas aeruginosa* infection in cystic fibrosis patients. *J. Clin. Microbiol.* 45: 3175–3183.

Wolfgang, M. C., Kulasekara, B. R., Liang, X., Boyd, D., Wu, K., Yang, Q., Miyada, C. G., and Lory, S. (2003). Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA* 100: 8484–8489.

Yahr, T. L., Mende-Mueller, L. M., Friese, M. B., and Frank. D. W. (1997). Identification of type III secreted products of the *Pseudomonas aeruginosa* exoenzyme S regulon. *J. Bacteriol.* 179: 7165–7168.

Zautner, A. E., Masanta, W. O., Michael Weig, M., Uwe Groß, U., Bader, O. (2015). Mass Spectrometry-based PhyloProteomics (MSPP): A novel microbial typing Method. *Sci. Rep.* 5: 13431

Zemanick, E. T., Harris, J. K., Wagner, B. D., Robertson, C. E., Sagel, S. D., Stevens, M. J., Accurso, F. J., and Laguna, T. A. (2013). Inflammation and airway microbiota during cystic fibrosis pulmonary exacerbations. *PLoS ONE.* 8:e62917.

Index

MALDI-TOF and Tandem MS for Clinical Microbiology, First Edition.
Edited by Haroun N. Shah and Saheer E. Gharbia.

b

d

e

g

h

m

n

q