MICROBIOLOGY MONOGRAPHS

ALEXANDER STEINBÜCHEL
Series Editor

Friedhelm Meinhardt
Roland Klassen
Editors

Microbial Linear Plasmids

Springer

7
Microbiology Monographs

Series Editor: Alexander Steinbüchel

Microbiology Monographs

Volumes published in the series

Inclusions in Prokaryotes
Jessup M. Shively (Editor)
Volume 1 (2006)

Complex Intracellular Structures in Prokaryotes
Jessup M. Shively (Editor)
Volume 2 (2006)

Magnetoreception and Magnetosomes in Bacteria
Dirk Schüler (Editor)
Volume 3 (2007)

Predatory Prokaryotes – Biology, Ecology and Evolution
Edouard Jurkevitch (Editor)
Volume 4 (2007)

Amino Acid Biosynthesis – Pathways, Regulation and Metabolic Engineering
Volker F. Wendisch (Editor)
Volume 5 (2007)

Molecular Microbiology of Heavy Metals
Dietrich H. Nies and Simon Silver (Editors)
Volume 6 (2007)

Microbial Linear Plasmids

Volume Editors: Friedhelm Meinhardt, Roland Klassen

With 57 Figures, 6 in Color

Springer

Volume Editors:

Prof. Dr. Friedhelm Meinhardt
Institut für Molekulare Mikrobiologie und Biotechnologie
Westf. Wilhelms-Universität Münster
Corrensstr. 3
48149 Münster
Germany
e-mail: meinhar@uni-muenster.de

Dr. Roland Klassen
Institut für Molekulare Mikrobiologie und Biotechnologie
Westf. Wilhelms-Universität Münster
Corrensstr. 3
48149 Münster
Germany
e-mail: rolando@uni-muenster.de

Series Editor:

Professor Dr. Alexander Steinbüchel
Institut für Molekulare Mikrobiologie und Biotechnologie
Westfälische Wilhelms-Universität
Corrensstraße 3
48149 Münster
Germany
e-mail: steinbu@uni-muenster.de

Library of Congress Control Number: 2007928742

ISSN 1862-5576
ISBN 978-3-540-72024-9 Springer Berlin Heidelberg New York
DOI 10.1007/978-3-540-72025-6

Editor: Dr. Christina Eckey, Heidelberg, Germany
Desk Editor: Dr. Jutta Lindenborn, Heidelberg, Germany
Cover Design: WMXDesign GmbH, Heidelberg, Germany
Typesetting and Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig, Germany

Printed on acid-free paper SPIN 11815754 149/3100 YL – 5 4 3 2 1 0

Preface

Linear plasmids were firstly discovered (in mitochondria of corn) approximately 25 years after the seminal, epoch-making original documentation of circular plasmids in bacteria by Joshua Lederberg and co-workers in the 1950s. This earmarked the advent of the molecular biological age. Though linear plasmids were initially regarded as rare and rather peculiar genetic elements, in the following three decades their frequency in a multitude of eukaryotic and bacterial species became evident. Despite linearity-neglecting definitions, which were even put forward in a number of textbooks, such a configuration was frequently found not only in eukaryotic and certain bacterial chromosomes but especially in extrachromosomal genetic elements of both taxa.

Linear DNA-molecules must meet the challenge of being shortened gradually during successive replication rounds; a problem that is solved for eukaryotic chromosomes by the action of the telomerase. In microbial linear plasmid systems other, probably ancient, mechanisms are instrumental in avoiding impending losses. A great number of linear plasmids from both bacteria and fungi are so-called invertrons, which possess terminal inverted repeat sequences and 5′covalently attached terminal proteins serving as replication primers. Such a replication mode, which is assumed to already have existed before the genesis of eukarya, is conserved not only in linear plasmids but also in adenoviruses and bacteriophages. The major player in replication, the DNA-polymerase encoded by invertron type linear plasmids, displays articulate similarities to enzymes of the above viruses, unambiguously confirming the common ancestry.

Similarly, viral ancestry is obvious for the linear retroplasmids present in several fungal species. Eponymous for such elements, replication involves an RNA intermediate serving as a template for the reverse transcriptase. As for the invertron-type DNA polymerase such a key player of replication substantiates the common ancestry of retroplasmids and retroviruses, which suggests that they are remnants of primitive genetic elements occurring shortly after or in parallel with the transition from the RNA world to the DNA world.

The most convincing argument for the affinity of linear plasmids and viruses originates from elements isolated as linear plasmids in enterobacteria, which subsequently shaped up as prophage stages. Such prophages, as well as linear plasmids from *Borrelia*, have covalently closed hairpin termini allowing DNA

replication to proceed around the turns. Since dimeric replication intermediates require resolution, an enzyme reminiscent of type IB resolvases is encoded by *Borrelia* plasmids as well as linear prophages.

The biological significance of linear plasmids is rather diverse. In yeast they may encode protein toxins, which not only enable autoselection but also have the ability to eliminate competitors. In bacteria they may confer specific catabolic capabilities, antibiotic resistance or enhancement of pathogenicity. Selfish elements without any (obvious) effect on their hosts exist in yeast and filamentous fungi; in the latter they may, however, integrate into the mitochondrial DNA, eventually resulting in a molecular disease causing senescence of the host.

Progenitors of linear plasmids date back to earliest stages of evolution to which they have significantly contributed. This is indicated by the highly transmissible *Streptomyces* elements, which clearly contributed to genome evolution by active DNA transfer and exchange, recruitment and spread of accessory genetic information and pathways involved in secondary metabolism. Some linear plasmids may be regarded as molecular fossils, which provide us with the unique possibility to study molecular details at the earliest stages of life.

The aim of this volume was to assemble a collection of reviews that comprehensively address the diverse structural and functional aspects of the rather heterogeneous group of genetic elements subsumed as "linear plasmids". We wish to express our gratitude to all authors for their substantial efforts and excellent contributions. It was their expertise in the particular areas of linear plasmid biology and their enthusiastic engagement in preparing the individual chapters that made it possible to bring together all these elements in one volume. We are grateful to Springer for giving us the opportunity to edit this book, in particular to Jutta Lindenborn from the editorial office for her support in all aspects of the publishing process.

Münster, May 2007

Roland Klassen
Friedhelm Meinhardt
Alexander Steinbüchel

Contents

Streptomyces Linear Plasmids: Their Discovery, Functions, Interactions with Other Replicons, and Evolutionary Significance
K. F. Chater · H. Kinashi 1

Streptomyces Linear Plasmids: Replication and Telomeres
C. W. Chen 33

Catabolic Linear Plasmids
S. Fetzner · S. Kolkenbrock · K. Parschat 63

Linear Plasmids and Phytopathogenicity
I. Francis · D. Gevers · M. Karimi · M. Holsters · D. Vereecke 99

The Linear Hairpin Replicons of _Borrelia burgdorferi_
K. Kobryn 117

Linear Plasmids and Prophages in Gram-Negative Bacteria
S. Hertwig 141

Retroplasmids: Linear and Circular Plasmids That Replicate via Reverse Transcription
J. T. Galligan · J. C. Kennell 163

Linear Protein-Primed Replicating Plasmids in Eukaryotic Microbes
R. Klassen · F. Meinhardt 187

Hairpin Plasmids from the Plant Pathogenic Fungi _Rhizoctonia solani_ and _Fusarium oxysporum_
T. Hashiba · A. Nagasaka 227

Subject Index 247

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_097/Published online: 27 June 2007

Streptomyces Linear Plasmids: Their Discovery, Functions, Interactions with Other Replicons, and Evolutionary Significance

Keith F. Chater[1] (✉) · Haruyasu Kinashi[2]

[1]John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK
keith.chater@bbsrc.ac.uk

[2]Department of Molecular Biotechnology, Hiroshima University,
739-8530 Higashi-Hiroshima, Japan

1 Introduction: Early Studies of Plasmids 2

2 Exploring the Genetics of Other Bacteria: The Dawn of *Streptomyces* Genetics 3

3 *Streptomyces coelicolor* A3(2) Has a Plasmid Fertility Factor, SCP1, Capable of Integration into the Chromosome to Give High-Fertility Variants 5

4 SCP1 is a Very Large Linear Molecule, and Linear Plasmids are Widespread Among Streptomycetes 7

5 Interactions of Linear Plasmids with the Chromosome: Physical Analysis 12
5.1 Integration of SCP1 Within the Chromosome 12
5.2 Single Crossover Recombination Between a Linear Plasmid and the Chromosome to Generate Molecules with Heterologous Ends . . . 14

6 Linear Plasmids and Antibiotic Production 16
6.1 Genes for Methylenomycin Production 18
6.2 pSLA2-L of *Streptomyces rochei*: A Plasmid Densely Packed with Genes for Secondary Metabolism 19

7 Linear Plasmids and Evolution 20
7.1 Evolutionarily and Adaptively Significant Genes Carried by Linear Plasmids 21
7.2 Possible Role of Linear Plasmids in the Dissemination of TTA-Containing Genes Subject to the Influence of the Developmental Gene *bldA* 23
7.3 Co-evolution of Streptomycetes and Their Linear Replicons 23

References 25

Abstract Unusually among bacteria, streptomycetes possess linear chromosomes, and many of them also carry linear plasmids (circular plasmids are also found). The linear plasmids range in size from tens to hundreds of kilobases. The most studied is SCP1, discovered as a sex factor in the model organism *Streptomyces coelicolor* A3(2). A variety

of co-integrates and hybrids have been found between SCP1 and the host chromosome, which can greatly increase the likelihood of chromosomal transfer. Several examples of the exchange of ends between linear plasmids and linear chromosomes have been documented. These can sometimes bring about the mobilisation of chromosomal genes for antibiotic biosynthesis. Some very large linear plasmids themselves carry genes for the biosynthesis of bioactive small molecules, including antibiotics. For example, such genes occupy about two thirds of the plasmid pSLA2-L. In another case, almost identical gene sets for methylenomycin biosynthesis are present both on the linear plasmid, SCP1, and on a quite different circular plasmid, pSV1, in a related streptomycete. It appears that linear plasmids have played key roles in the architecture, accessory gene content and rapid evolution of *Streptomyces* chromosomes. They may have permitted the diversification and spread of pathways for secondary metabolism, and the evolution of some *Streptomyces*-specific families of paralogous genes. They may also have been the source of most of the genes that, because of their possession of the rare TTA codon, are dependent for expression on the cognate tRNA specified by *bldA*, a gene whose deletion has wide-ranging effects on morphological differentiation and secondary metabolism.

1 Introduction: Early Studies of Plasmids

The brilliant early studies of bacterial genetics (see Hayes, 1964, for a review) were helped by good fortune in Joshua Lederberg's choice of *Escherichia coli* K-12 as the strain for investigation, because this strain happened to have a genetic determinant, termed the F-factor, that permitted genes to be transferred between strains, albeit at a low frequency ($< 10^{-6}$). By the mid-1950s it had been deduced that all the genetic markers of *E. coli* were arranged on a single linkage group, i.e. there was a single chromosome, with the exception of F itself: its transfer was not associated with the transfer of particular chromosomal regions. F was also fairly easily lost, and such F^- variants could readily be converted to F^+ by growth in contact with a differently marked F^+ strain, at frequencies many times higher than the transfer of markers on the chromosome. Thus, F was a plasmid—a genetic element able to replicate separately from the chromosome. However, F^+ cultures contained rare variants that showed a much higher frequency of chromosomal recombination in crosses with F^- strains. In these Hfr (high frequency of recombination) variants, the F-factor was no longer easily lost, and it showed genetic linkage to chromosomal genes, a result which meant that the factor had integrated into the chromosome.

It was subsequently found that the integrated F-factor could sometimes excise from the chromosome of Hfr strains together with adjacent chromosomal DNA, to give rise to F-prime (F′) factors, which provided tools for geneticists to carry out functional tests, such as dominance and complementation. These observations provided early evidence of the kinds of molecular exchange that might be contributing to the evolution of bacterial chromosomes. Moreover, the *E. coli* chromosome was shown both genetically and physically to be cir-

cular, so the fact that the F-factor could integrate into it and excise from it suggested that F was also circular, with a single, reversible recombination event between it and the chromosome accounting for integration. Such single crossover integration events became known as "Campbell integration", after the first clear proposal by Campbell that they could account for the ability of the DNA of bacteriophage lambda to integrate into, and excise from, the *E. coli* chromosome.

Within a few years, Watanabe found that other transmissible plasmids (R-factors) were the agents of the spread of multiple antibiotic resistance among enteric bacteria, and subsequently numerous different phenotypic traits also turned out to be encoded by plasmids. It became obvious that plasmids could be found in bacteria of almost any taxonomic group. Clearly, such elements were contributing in various ways to host evolution and adaptation. This, coupled with the possibility that they might provide model systems for studies of DNA-related physiological questions, made it important to isolate and characterise them physically. The circularity of known plasmids gave rise to the first method for plasmid purification (Clewell and Helinski 1970). The method depends on the intercalating dye ethidium bromide, which is taken up to a reduced extent by covalently closed circular (CCC) plasmid DNA compared to open-circular (nicked) plasmid DNA, or linear DNA fragments such as those inevitably generated by shearing chromosomal DNA during its isolation. The CCC plasmids therefore undergo a smaller decrease in buoyant density than linear DNA during high-speed centrifugation in a CsCl–ethidium bromide solution, and form a separate band below chromosomal DNA when the density gradient reaches equilibrium. Material separated in this way, coupled with other advances such as the development by Kleinschmidt of a method for displaying DNA in the electron microscope, led quickly to a much greater understanding of the genetic and physical organisation of plasmids, and allowed their first use as vectors for gene cloning in the early 1970s. At that time, there was no reason to expect exceptions to the rule that plasmids were circular DNA molecules, and further rapid and convenient methods for their purification, all depending on their CCC nature, became available (e.g. Birnboim and Doly 1979).

2
Exploring the Genetics of Other Bacteria: The Dawn of *Streptomyces* Genetics

After the first discovery that it was possible to carry out genetic analysis in *E. coli*, genetic exchange was soon demonstrated in other phylogenetically distant bacteria, and harnessed for experimental purposes. Against this background, several scientists began in the mid-1950s to look for evidence of genetic exchange in *Streptomyces* (reviewed by Hopwood 2007). There were

three main reasons why this particular group of bacteria attracted attention. Firstly, it had become obvious that streptomycetes were a rich source of antibiotics, so there was a real possibility that success in developing genetic systems would find long-term utility. Secondly, these abundant, widespread, ecologically significant soil organisms were phylogenetically distant from *E. coli* and other genetically studied bacteria, and had an unusual mycelial growth habit like that of fungi (Fig. 1) (in fact, some people still incorrectly thought of them as intermediate between bacteria and fungi, or even as fungi). It was therefore felt that they should, on the one hand, exhibit novel features and, on the other, help to reveal what was universally true of bacteria. Thirdly, they were easy to grow at a reasonable rate on defined media, and went through a stage in their life cycle in which, as spores, they had a single copy of the genome, allowing the progeny of crosses to be classified readily; they were thus potentially experimentally convenient.

The greatest success was had with *Streptomyces coelicolor* A3(2), the organism chosen by D.A. Hopwood (Hopwood 1999). His earliest papers showed that random segments of its chromosome could be transferred between strains by conjugation, with both participants being able to act as donor or recipient (the crosses were "non-polar") and without obvious preference for particular chromosome segments (Hopwood 1967). Subsequent work revealed the involvement of a complex fertility system, involving plasmids, which is described in the next section. It was from this work that the unusual nature of some *Streptomyces* plasmids began to emerge, culminating in the 1980s with the revelation that many of them are linear.

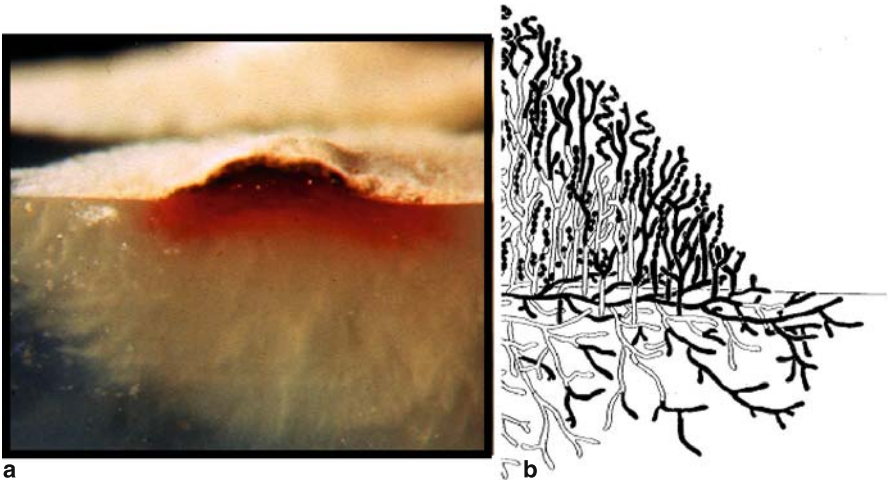


Fig. 1 The *Streptomyces* colony. **a** A mature *S. coelicolor* colony grown on agar for 5 days sliced vertically with a razor (photograph by N.J. Ryding). **b** Diagrammatic cross section showing mycelial branching and sporulation in the aerial mycelium (after Wildermuth 1970)

3
Streptomyces coelicolor A3(2) Has a Plasmid Fertility Factor, SCP1, Capable of Integration into the Chromosome to Give High-Fertility Variants

Sermonti and Casciano (1963) provided the first hints that different fertility types might be represented among strains derived from *S. coelicolor* A3(2). A subsequent large-scale systematic analysis was carried out (see Hopwood et al. 1973, for a review). Colonies from UV-exposed spores of one strain (number 12, and therefore an early derivative of the wild type) were replicated onto plates spread with untreated spores of mating partners with suitably different genetic markers, to allow growth, mating and sporulation to take place. On a further round of replication to a medium selecting for recombinants, it turned out that about 1% of the colonies of strain 12 gave rise to dramatically increased numbers of recombinants, and were similarly highly fertile in crosses with several other multiply marked strains from the *S. coelicolor* lineage. These variants of strain 12 were called "ultrafertile" or "UF" (Hopwood et al. 1969).

Detailed analysis of the ultrafertile crosses showed that the UF strains derived from strain 12 were in fact super-recipients, and that the partner strains (designated "NF" for "normal fertility") behaved as donors. Not only were the crosses polarised in direction, but the pattern of chromosome donation had changed from that observed in the earliest *S. coelicolor* crosses: there was a bidirectional gradient of donor marker inheritance among recombinants, centred on the determinant of the NF fertility type. The crosses were indeed astonishingly fertile: nearly all the spores from a well-balanced mating mixture were recombinants, with the donor and recipient genotypes virtually absent. Strain 12, from which the UF strains had been derived, proved to be about ten times less fertile in crosses with NF strains, and these properties, which were shared with the wild type and other direct derivatives from it, defined the IF type (for initial fertility) (Vivian and Hopwood 1970). A subsequent fuller exploration of pairwise crosses between representatives of all three fertility types showed that UF $\times$ UF and IF $\times$ UF crosses exhibited comparatively low frequencies of chromosomal recombination (Vivian and Hopwood 1970); but, in the IF $\times$ UF crosses, strikingly, close to 100% of all those progeny with the UF chromosomal genotype had acquired IF status (Vivian 1971).

From such results, it was concluded that IF strains, such as the wild type and strain 12, possessed a self-transmissible plasmid, termed SCP1, which replicated independently of the chromosome and could be lost to give UF strains either spontaneously (from about 0.1% of spores) or, with tenfold higher frequency, after UV irradiation. SCP1 transfer from IF strains was associated, albeit infrequently, with co-transfer of chromosomal DNA. In the NF super-donor strains, SCP1 appeared to have integrated into the

chromosome, in a genetically long region otherwise devoid of known genes (this long "silent" region may, at least partially, be attributable to the integrated SCP1 itself, which is about 5% of the length of the chromosome) (Fig. 2). In a successful search for new examples of SCP1 integration, Vivian and Hopwood (1973) and Hopwood and Wright (1976b) isolated further high-frequency donor strains from irradiated spores of the SCP1$^+$ strain 12 (Fig. 2). These differed from NF strains with respect to their stability, apparent point of SCP1 integration (in some cases) and gradient of marker transfer (they appeared to be unidirectional donors), though it was possible to isolate an NF-like bidirectional donor strain (A634) from one of them by a further round of screening (Fig. 2: see also below). Thus, SCP1 appeared to be able to recombine with different regions of the chromosome, perhaps in different configurations to account for unidirectional or bidirectional transfer.

Having found that SCP1 was behaving in a manner at least partially resembling that of the F-factor of *E. coli*, Hopwood and Wright (1973b,1976a)

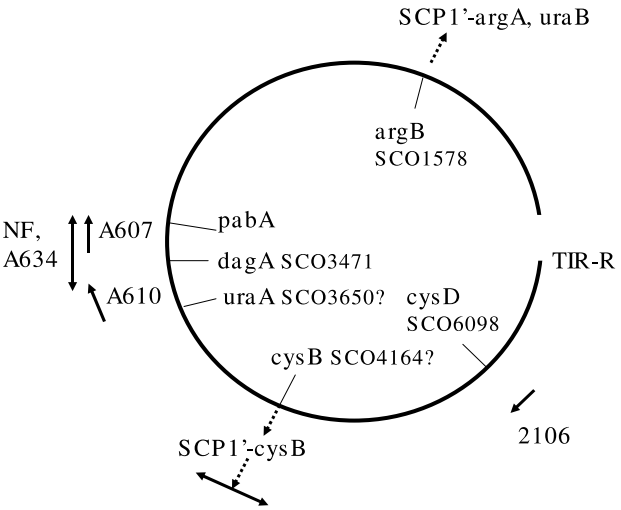


Fig. 2 Fertility associated with recombination of SCP1 with the chromosome. Approximate positions of the regions of the chromosome at which recombination with SCP1 has taken place to give high-frequency donor strains are shown in relation to a chromosome map obtained by linkage analysis. *Arrows* indicate the direction of the gradient of marker inheritance by recombinants. This linkage map is collinear with the sequenced genome, and "classical" gene names are here annotated, where possible, with the numbered open reading frames from the genome sequence, with uncertainty about the congruence indicated by a *question mark*. The distances between markers correspond only approximately to the physical coordinates of the genome. The linkage map is circular, but the sequenced genome is linear, so the linkage map has here been opened at the appropriate position (see Hopwood 2006 for a discussion of this apparent paradox)

went on to find two autonomously replicating SCP1 derivatives that resembled F′ plasmids in carrying genetically identifiable chromosomal markers (*cysB* and *argA-uraB*, hence the respective plasmid designations SCP1′-*cysB* and SCP1′-*argA, uraB*; Fig. 2). These could mobilise the chromosome, the best characterised (SCP1′-*cysB*) giving a bidirectional gradient of chromosomal marker transfer that was highest for genes close to *cysB*. It also proved possible to isolate further derivatives from an SCP1′-*cysB*-carrying strain in which the plasmid appeared to have integrated fairly stably into the chromosome. Although these SCP1′ strains and their derivatives have yet to be characterised physically, some of the other "new" kinds of donors are revisited later in this chapter, when the molecular basis for their properties is discussed.

4
SCP1 is a Very Large Linear Molecule, and Linear Plasmids are Widespread Among Streptomycetes

Since SCP1 behaved in many ways similarly to well-characterised plasmids such as the F-factor of *E. coli*, it was reasonable to suppose that it would be physically similar as well. It was therefore natural to try to isolate it by CsCl–ethidium bromide density gradient centrifugation. Application of this technique to *S. coelicolor* indeed yielded CCC DNA, of about 30 kb (Schrempf et al. 1975), but this proved to be present even in SCP1-free strains, and was designated SCP2 (Bibb et al. 1977). Nevertheless, physical evidence that SCP1 existed as DNA came from two kinds of experiments: DNA renaturation using *S. coelicolor* strains and a strain of another species, *S. parvulus*, into which SCP1 had been transferred by conjugation revealed that the presence of SCP1 was associated with a significant amount of DNA (more than 100 kb of unique sequence) (Hopwood et al. 1979); and fragments of SCP1 were cloned in the early 1980s (Bibb et al. 1980; Chater and Bruton 1983, 1985).

Eventually, the idea that SCP1 might be linear instead of circular was given impetus by the ground-breaking discovery of linear plasmids in *S. rochei* in K. Sakaguchi's laboratory. In the 1970s, the development of restriction enzyme technology, and the resultant recombinant DNA revolution, had been greatly helped by the use of electrophoresis through agarose gels to separate linear DNA fragments. Hayakawa et al. (1979) found that when total DNA preparations from a lankacidin-producing streptomycete (later called *S. rochei*) were subjected to standard agarose gel electrophoresis, the usual loose band corresponding to sheared chromosomal DNA fragments was accompanied by a sharp band of molecules shown by electron microscopy to be about 15 kb (later recalibrated as 17 kb), and revealed by restriction enzyme analysis to consist of one type of molecule. (Note that the use of a pH-neutral DNA extraction procedure was important for the success of this experiment:

the usual alkaline extraction method for plasmid isolation denatures and precipitates most linear DNAs.) The copy number of this DNA was about 60 per chromosome (Hirochika and Sakaguchi 1982). It was deduced that proteins were bound at or close to the ends, since the plasmid end-fragments were not seen on agarose gels when DNA prepared without the usual help of pronase was digested with restriction enzymes, and subsequent pronase digestion resulted in their full representation. Hirochika et al. (1984) went on to show that only the 5′-ends were linked to a protein. They also sequenced about 800 bp from each end, and found that both ends had the same sequence for 614 bp. The presence of 5′-linked proteins, often at the ends of terminal inverted repeats (TIRs) of various lengths, has turned out to be a general feature of linear replicons of actinomycetes, the terminal proteins playing a key role in the replication of the ends (Bao and Cohen 2001; see Chen 2007, in this volume).

Multicopy linear plasmids of tens of kilobases can be detected in some other streptomycetes by conventional gel electrophoresis (e.g. *S. rimosus*, Chardon-Loriaux et al. 1986; *S. clavuligerus*, Keen et al. 1988), but they have not so far been shown rigorously to carry genes for recognisable phenotypic traits, or to interact physically with the host genome. It was not until the advent of various kinds of pulsed field gel electrophoresis (PFGE; Schwartz and Cantor 1984; Carle and Olson 1984) that very large linear DNA molecules were revealed by the work of H. Kinashi. These PFGE techniques all depend on frequent changing of the direction of the current during electrophoresis, which requires linear DNA molecules to realign, a process that takes longer for larger molecules. Shearing of the DNA is minimised by preparing it from mycelium immobilised in an agarose plug. The full preparation in the plugs includes successive treatments with lysozyme, pronase and detergent, before the plugs are embedded in the electrophoresis gel.

The Kinashi laboratory's extraordinary first results were published in two papers (Kinashi and Shimaji 1987; Kinashi et al. 1987). They found one or more linear plasmids ranging from 17 to 590 kb in six out of 12 antibiotic-producing streptomycete (strains) tested (Fig. 3). (Although these have often been called "giant" linear plasmids, it is clear that there is a continuum of sizes of linear plasmids from 17 to hundreds of kb, so we have not retained the term in this chapter.) One of the strains examined was *S. coelicolor* A3(2), in which it was found that a series of plasmids varying in size from 410 to 590 kb in ca. 30-kb increments was made up of forms of SCP1 (using as probe a cloned fragment of SCP1 carrying resistance to methylenomycin—see below and Fig. 4). No bands were found in a strain genetically defined as lacking SCP1, and strains carrying SCP1, but not SCP2, had only a single SCP1 band of 350 kb. Subsequent work showed that the bands forming the ladder in the A3(2) strain arose from the integration of tandem copies of SCP2-derived sequences into SCP1 (see below).

Further examples of linear plasmids have also been described from other species (Table 1), and they have been found in several other actinomycete

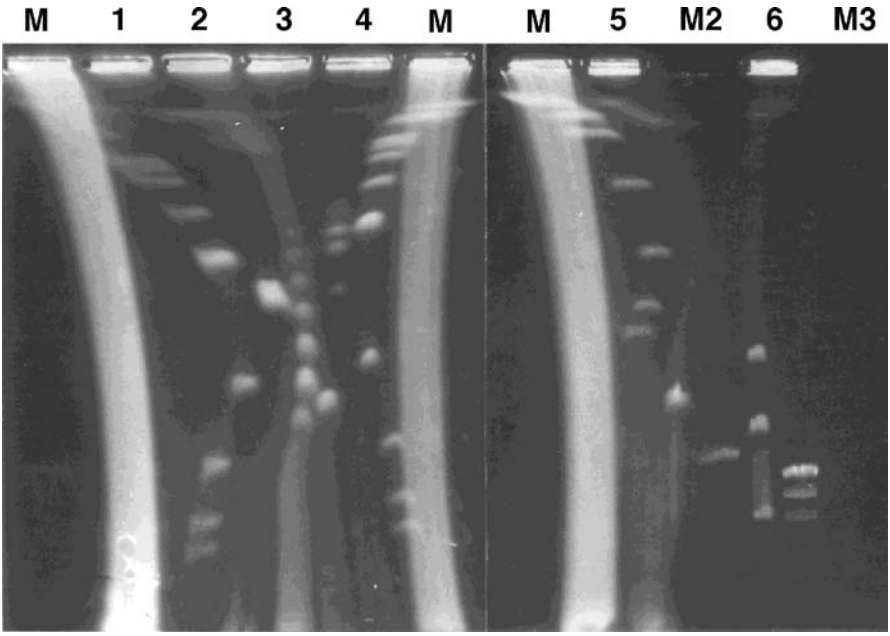


Fig. 3 Separation of linear plasmids from antibiotic-producing *Streptomyces* strains by orthogonal-field-alteration gel electrophoresis. The figure is from Kinashi and Shimaji (1987). Note that in these early days of PFGE technology, the tracks often showed great distortion, which has been largely overcome by later technical improvements. M, *Saccharomyces cerevisiae* chromosomes; M2, λDNA; M3, λDNA digested with HindIII; 1, *Streptomyces lasaliensis* NRRL3382R; 2, *S. violaceoruber* JCM4979; 3, *S. fradiae* C373.1; 4, *S. parvulus* JCM4068; 5, *S. venezuelae* JCM4526; 6, *S. rochei* 7434AN4

genera and other bacteria (see Fetzner et al. 2007, Hertwig 2007, Holsters et al. 2007, all in this volume). Here we briefly expand on a few general aspects of *Streptomyces* linear plasmids. The first such plasmids to be studied all possessed TIRs hundreds of base pairs in length, leading Hirochika et al. (1984) to propose that the terminal repeats were held together by proteins in a "racket frame-like" structure ("invertrons": Sakaguchi 1990); but some other linear plasmids from streptomycetes have TIRs of only a few tens of base pairs, and in the cases of *Rhodococcus opacus* plasmids pHG201 and 205 the TIRs are of only 3 bp (Kalkus et al. 1998), so the racket-frame model in its original incarnation does not really apply. However, the ends ("telomeres") of all actinobacterial linear plasmids do show complex secondary structural features, conserved among most of the plasmids, over several tens of bases (see Chen 2007, in this volume). The fact that the plasmid 5′-ends are always linked to a terminal protein leaves open the possibility that the ends may be held together by terminal-protein-mediated interactions. Significantly, *Streptomyces* chromosomes were also eventually found to be linear, and to have

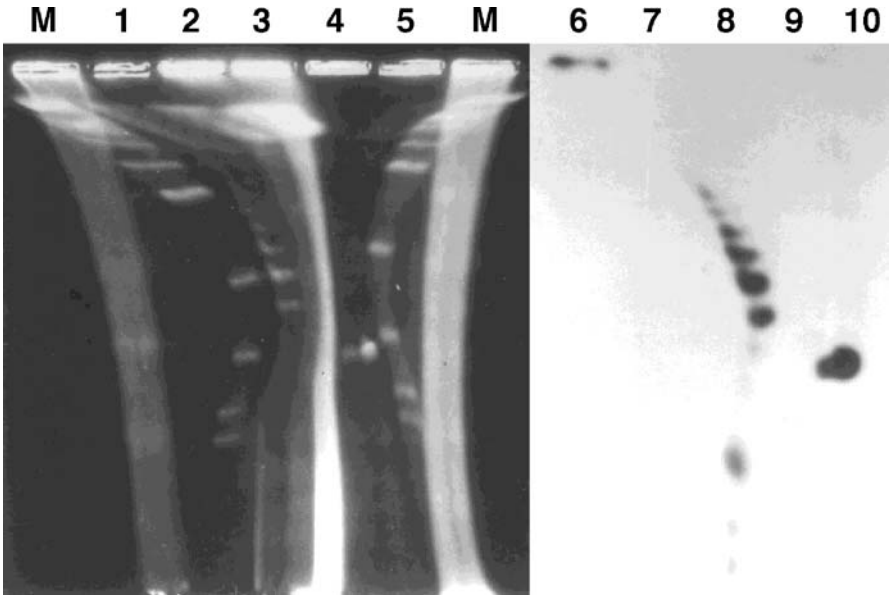


Fig. 4 Separation of SCP1 plasmids by PFGE and their Southern blot analysis probed by the methylenomycin resistance gene (*mmr*). The figure is from Kinashi et al. (1987). In addition to the linear plasmid SCP1, a series of plasmids were detected in *S. violaceoruber* JCM4979, which was originally Sermonti's A3(2) strain. The multiple bands are the result of tandemly integrated sequences derived from the circular plasmid SCP2, which is also present in strain A3(2) (see text for further discussion). M, *Saccharomyces cerevisiae* chromosomes; 1, 6, *S. violaceoruber* JCM4979, originally Sermonti's *S. coelicolor* A3(2); 2, 7, *S. coelicolor* M124; 3, 8, *S. coelicolor* M130; 4, 9, *S. coelicolor* M138; 5, 10, *S. coelicolor* M146

similar protein-linked ends (Lin et al. 1993; Chen 1996; Chen et al. 2002). Cytological evidence in support of spatial co-location of the ends of the *S. coelicolor* chromosome was obtained by Yang and Losick (2001), using fluorescence in situ hybridisation to probes ostensibly specific for each end, but it was subsequently found that the particular strain used in that work (J1508) possessed a duplication compared with the sequenced strain M145 that would have resulted in both probes hybridising to the same end (Weaver et al. 2004; C.W. Chen, personal communication).

At least some of the linear plasmids are evidently transmissible by conjugation (SCP1, Vivian 1971; SLP2, Chen et al. 1993), but the absence of genetic markers on many of them has limited analysis of this aspect. Their copy numbers vary from several tens per chromosome in the case of the small plasmids to just a few-fold for the larger examples; but at least for SCP1, which typically has a copy number of up to seven in *S. coelicolor* (Yamasaki et al. 2003), the copy number was greatly increased in an SCP1$^+$ derivative of *S. parvulus* obtained by interspecific conjugation (Hopwood and Wright 1973a; Chater and Bruton 1983).

Table 1 Linear plasmids of streptomycetes

Plasmid (host)	Size (kb)	Terminal inverted repeats	Copy number	Comments	Refs.
SCP1 (*S. coelicolor*)	356	75 kb	c. 7	Encodes methylenomycin production; many interactions with host chromosome	Bentley et al. 2004
SLP2 (*S. lividans*)	50	44 bp	?	15.7 kb at one end; corresponds to ends of host chromosome	Chen et al. 1993
pPZG101 (*S. rimosus*)	387	95 kb	?	Derived from pPZG102 during strain improvement; many interactions with chromosome including acquisition of oxytetracycline pathway genes	Gravius et al. 1994
pSCL1 (*S. clavuligerus*)	12	c. 900 bp	High	Host produces clavulanic acid	Keen et al. 1988; Wu and Roy 1993
pSLA2-S (*S. rochei*)	17	614 bp	c. 60	Origin of the "racket-frame" model	Hirochika et al. 1984
pSLA2-M (*S. rochei*)	100	?	?		Kinashi et al.1994
pSLA2-L (*S. rochei*)	211	12 bp	?	Mainly occupied by gene sets for production of lankacidin, lankamycin, a type II polyketide, carotenoids, gamma-butyrolactone	Mochizuki et al. 2003
pSRM (*S. rimosus*)	43	?	?	Evidence for pock formation accompanying transfer	Chardon-Loriaux et al. 1986
Unnamed (*S. rimosus*)	255	?	?	Host produces oxytetracycline	Rathos et al. 1989
pKSL (*S. lasaliensis*)	520	?	?	Host produces lasalocid and echinomycin	Kinashi et al. 1987
SAP1 (*S. avermitilis*)	94	?	?	Discovered during genome sequencing	Ikeda et al. 2003
Unnamed (*S. fradiae*)	420	?	?	Host produces tylosin	Kinashi and Shimaji 1987
Unnamed (*S. parvulus*)	520, 560, 580	?	?	Host produces actinomycin D	Kinashi and Shimaji 1987

Table 1 (continued)

Plasmid (host)	Size (kb)	Terminal inverted repeats	Copy number	Comments	Refs.
Unnamed (*S. venezuelae*)	130	?	?	Host produces chloramphenicol	Kinashi and Shimaji 1987
pSV2 (*S. violaceo-*) *ruber*	100	426 bp	?	Plasmid ends similar to those of the *S. coelicolor* chromosome	Spatz et al. 2002

5
Interactions of Linear Plasmids with the Chromosome: Physical Analysis

The physical isolation of SCP1 (Kinashi and Shimaji-Murayama 1991) was followed by the cloning of segments of SCP1 in cosmids and the establishment of a restriction map (Redenbach et al. 1998), and eventually the sequencing of the plasmid (Bentley et al. 2004) and the chromosome of its host (Bentley et al. 2002). It became possible to investigate the physical basis of some of the chromosome-donating strains described above.

5.1
Integration of SCP1 Within the Chromosome

The first SCP1–chromosome chimera to be investigated was that of an NF strain 2612 (Fig. 5a: Hanafusa and Kinashi 1992; Kinashi et al. 1992). Combining that analysis with DNA sequence information (Yamasaki et al. 2000, 2001; Bentley et al. 2002, 2004), it emerged that chromosomal integration of SCP1 had involved (a) interruption of the chromosome by insertion of SCP1 into gene SCO3464, such that the left end of SCP1 at this junction was almost intact, and (b) at the other end of SCP1, recombination between a copy of IS*466* in SCP1 (IS*466*S; SCP1.276) and another copy in the chromosome (IS*466*B; SCO3490) (Fig. 5a). In the course of these events, both the right arm of SCP1 and a ca. 33-kb segment of the chromosome (genes SCO3465–SCO3489) were lost. The chromosomal deletion includes a large series of genes involved in an undefined aspect of carbohydrate metabolism. At one end of the deleted segment is the *dagA* gene encoding a diffusible agarase, which Hodgson and Chater (1982) had earlier found was missing from NF strains.

An independently isolated "NF-like" strain, A634 (Vivian and Hopwood 1973, see above), was obtained in a second step from a donor strain that initially showed a unidirectional gradient of transfer. In the formation of A634,

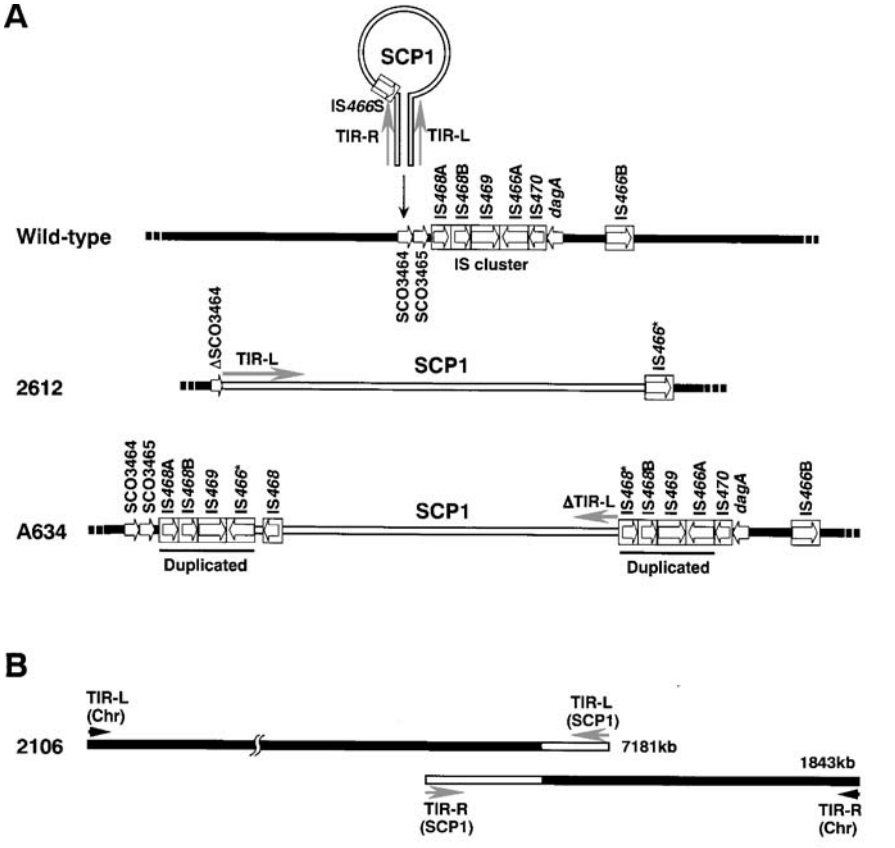


Fig. 5 Structures of some donor strains resulting from interaction of SCP1 with the chromosome (**a**) and two chimeric chromosomes in *S. coelicolor* 2106 (**b**). IS elements with an *asterisk* were involved in recombination between SCP1 and the chromosome. See text for further details

SCP1 recombination with the chromosome again involved interactions of plasmid- and chromosome-borne copies of IS*466*; but this time SCP1 became integrated in the opposite orientation. The multi-step formation of A634 is reflected by the presence of deletions of both ends of the integrated SCP1 (Fig. 5a). Because of these end deletions, it is not possible to deduce the position at which SCP1 first interacted with the chromosome: this may have been in SCO3464, as in the case of NF, but with the opposite orientation of SCP1. Yamasaki et al. (2001) hypothesised a series of subsequent events to give rise to the eventual structure of A634. Their model invokes an exchange between IS*466* copies in two non-identical chromosomes—the A634 progenitor and a wild-type copy—to account for the eventual location of the left-hand junction of the chromosome and SCP1 at a copy of IS*466* (SCO3469), as well as

for the occurrence of a duplicated 5.4-kb chromosome segment on either side of the integrated plasmid. In A634, another transposable element, IS*468*, is present in two places in the integrated SCP1, whereas SCP1 usually contains no IS*468* copies. The two implied IS*468* transposition events must have taken place from the chromosome, which has tandemly arranged copies located in the SCP1 integration target region (IS*468*A, SCO3466; IS*468*B, SCO3467). Recombination of IS*468*A with one of the SCP1-located copies accounts for the deletion of part of the SCP1 TIR-L during the formation of A634. A partially alternative model for the origin of A634 is possible, based on a Campbell integration event (see earlier). Because such events are mediated by a single crossover between homologous sequences present on two different DNA molecules, they result in duplication of the common sequence, such as is seen in A634. Perhaps the lineage of A634 included a circularised SCP1′ plasmid carrying this 5.4-kb segment.

It is not known what gives rise to unidirectional and bidirectional gradients of marker inheritance in crosses of the various donor strains with SCP1$^-$ recipients. The devising of models is limited by a lack of knowledge about the basic enzymology and geometry of SCP1 transfer, which is different from that of the plasmids of enteric bacteria (such as F in *E. coli*), and probably involves double-stranded transfer, though whether starting from one or both ends, or from an internal transfer origin, is unclear (see Hopwood 2006 and the chapter by Chen, in this volume, for more detailed reviews of these aspects).

5.2 Single Crossover Recombination Between a Linear Plasmid and the Chromosome to Generate Molecules with Heterologous Ends

Interestingly, a single exchange between SCP1 and the chromosome (albeit showing complexity at the nucleotide sequence level) can generate a viable strain (Yamasaki and Kinashi 2004). In a unidirectional donor strain 2106 (Hopwood and Wright 1976b) capable of high-frequency donation of the *cysD* gene (SCO6098), and of donation of a gradient of markers clockwise from *cysD* (Fig. 2), the housekeeping genes of *S. coelicolor* have become split between two chromosomes: one of 7181 kb, comprising the left end of the normal chromosome (SCO0001–SCO6389′) and the left end of SCP1 (SCP1.1–SCP1.136′), and the other of 1843 kb, fusing the right end of the normal chromosome (SCO6388′–SCO7846) to the right end of SCP1 (SCP1.136–SCP1.353) (Fig. 5b). The recombination event giving rise to these two complementary fusion chromosomes was an "illegitimate" event, in that there was no sequence homology at the point of crossing over and a small number of bases had been lost at one junction. Strikingly, the two hybrid chromosomes both have non-identical termini (one from SCP1, one from the chromosome), although these termini have some of the general structural features of *Strep-*

tomyces telomeres (though it should be noted that the SCP1 termini do have some atypical features) (see Chen 2007, in this volume).

One might wonder how replication and partitioning of the two chromosomes in 2106 could be stably achieved during various key points in mycelial growth (Fig. 6). Thus, at least one chromosome origin maintains a position closely behind growing hyphal tips, and this positioning must be established at the new tips generated by branch emergence. Also, chromosomes need to be partitioned to either side of newly formed septa, most particularly during sporulation, when each spore usually receives just one copy of the chromosome (the need for a plasmid to partition into spores would also be crucial). In fact, both of the SCP1–chromosome hybrid replicons in strain 2106 appear to be equipped with replication origins and partitioning functions. The larger of the hybrid molecules contains the entire chromosomal replication origin region, including the *parAB* partitioning genes (Kim et al. 2000). The smaller presumably depends on equivalent functions of SCP1: the right end of SCP1 present in this hybrid chromosome contains a segment capable of autonomous replication in an SCP1-free host (extending from within SCP1.194 to a site within SCP1.199, Redenbach et al. 1999), as well as two sets of putative partitioning genes (ORFs SCP1.138–139; SCP1.221–

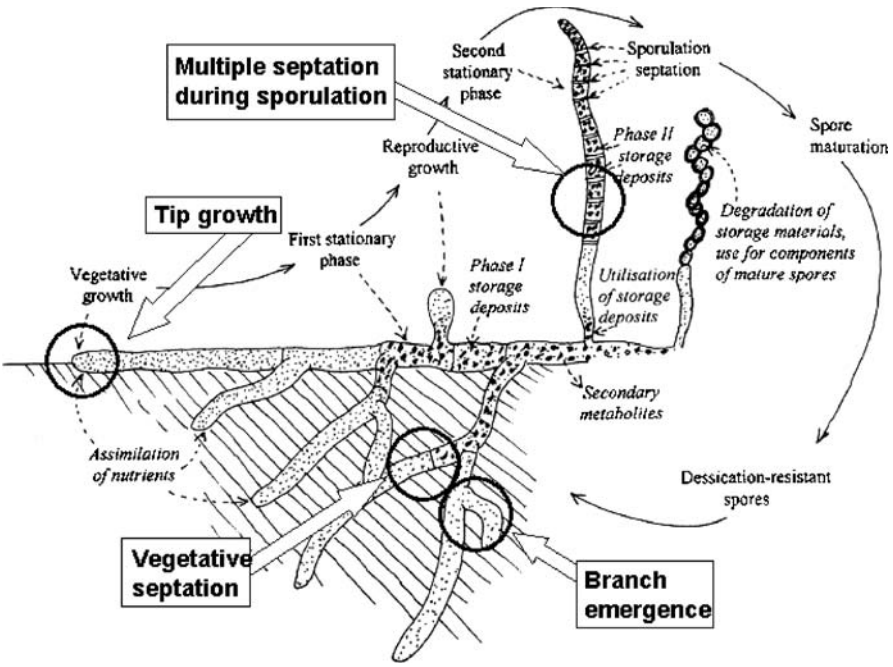


Fig. 6 Critical points for the efficient partitioning of plasmids and the chromosome during growth and development of a *Streptomyces* colony. The diagram is based on that of Chater (1998)

222). (One or both of the latter pairs may be responsible for the observation that the deletion of the chromosomal *parAB* genes, which causes frequent failure of chromosome partitioning during sporulation of SCP1$^-$ strains, had little effect on a strain containing an integrated copy of SCP1 (Bentley et al. 2004).)

Other cases of exchange of a chromosome end with a linear plasmid end have been reported, demonstrating the "viability" of linear replicons with almost no TIRs. In *S. lividans*, this apparently resulted from recombination between chromosomal and plasmid SLP2-borne copies of a 5.3-kb transposable element (Tn*4811*) (resolution of a Tn*4811* transposition event was another possible explanation) (Huang et al. 2003). A further example is provided by the acquisition of an *S. rimosus* chromosome end, including oxytetracycline biosynthetic genes, by the linear plasmid pPZG101: an event involving recombination at a homologous sequence of just four base pairs (Gravius et al. 1994; Pandza et al. 1998; see below). We note that in these cases, physical analysis of terminal structures was confined to the mainly plasmid hybrid molecules, and was not extended to the mainly chromosomal ones.

As already mentioned, SCP1 has also been found to recombine with a circular plasmid, SCP2, in an SCP1$^+$ strain that also carried SCP2; SCP1 was seen as a ladder of bands on PFGE, with increments of ca. 30 kb caused by tandem integrated sequences mostly derived from SCP2 in a complex manner that is not yet fully elucidated, but which may be the result of a single initial crossover between copies of Tn*5714* present in both elements (Kinashi et al. 1993; Bentley et al. 2004; H. Kinashi, unpublished results). This provides a model for the probable origin of linear chromosomes from a single crossover between a circular chromosome and a linear extrachromosomal replicon.

6
Linear Plasmids and Antibiotic Production

In the 1970s it was suggested that antibiotic production might be plasmid-encoded, following the observation that antibiotic production was often easily lost, particularly after treatments with UV or intercalating agents previously used to cure some circular plasmids in other bacteria (Okanishi et al. 1970). The discovery that SCP1 did indeed carry the genes for methylenomycin biosynthesis (Kirby et al. 1975) reinforced this general hypothesis (see Hopwood and Merrick 1977, for a careful review of the evidence at that time). However, further work on the *S. coelicolor* gene sets for production of actinorhodin (Wright and Hopwood 1976b; Rudd and Hopwood 1979), undecylprodigiosin (Rudd and Hopwood 1980) and calcium-dependent antibiotic (CDA) (Hopwood and Wright 1983) showed that they were all chromosomal, and no further clear examples of plasmid-linked antibiotic biosynthesis

emerged for several years. The early papers on PFGE-based identification of linear plasmids from antibiotic-producing streptomycetes reopened this debate (Kinashi 1994), but the initial results were circumstantial and often equivocal. For example, only one of two actinomycin D-producing streptomycetes had such a plasmid (Kinashi and Shimaji 1987), and the same was found for two independently isolated tetracycline producers (Kinashi and Shimaji 1987; Rathos et al. 1989). However, the evidence was stronger in the case of lasalocid biosynthesis by *S. lasaliensis*, in which the absence or presence of such plasmid (after curing treatments or plasmid re-establishment following protoplast fusion) correlated with non-production or production, respectively (Kinashi et al. 1987, 1988). More recently, detailed molecular analysis, culminating in sequencing of the entire 210 614-kb linear molecule, has shown that pSLA2-L of *S. rochei* is mainly composed of gene sets for the biosynthesis of secondary metabolites (Fig. 7; Mochizuki et al. 2003). The next two subsections discuss the specification of antibiotic production by SCP1 and pSLA2-L in more detail.

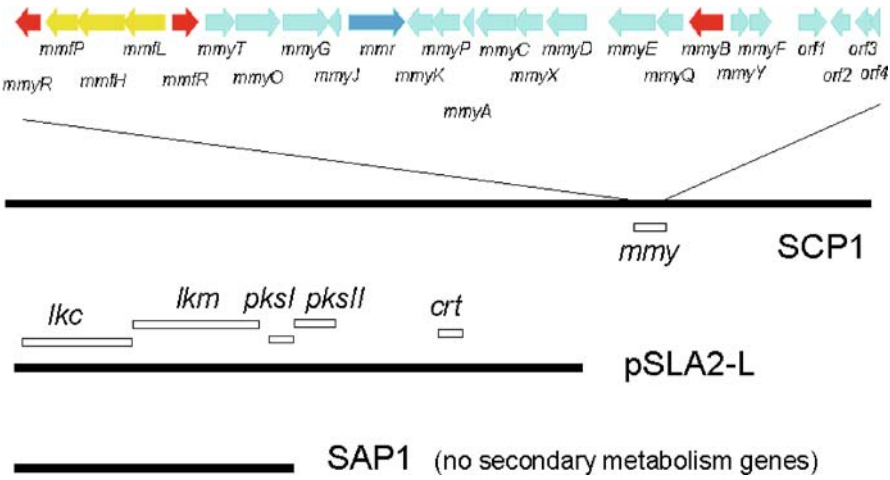


Fig. 7 Large linear plasmids vary greatly in their possession of gene sets for antibiotic biosynthesis. In the examples shown, SCP1 (365 kb) was initially studied for its fertility properties in *S. coelicolor* A3(2); pSLA2-L (211 kb), from *S. rochei*, was identified in a PFGE study of antibiotic-producing streptomycetes, and SAP1 (94 kb) was discovered in the course of genome sequencing of *S. avermitilis* (see Table 1 for references). The three plasmids shown were the first three large linear plasmids to be sequenced. The top of the diagram shows the 20-kb gene set for methylenomycin A biosynthesis. *Light blue arrows*, genes encoding enzymes of methylenomycin biosynthesis; *dark blue*, methylenomycin resistance; *red*, regulatory genes, including the possible M-factor receptor genes *mmfL* and *mmfR*; *yellow*, M-factor biosynthetic genes. pSLA2-L biosynthetic gene abbreviations: lkc, lankacidin; lkm, lankamycin; pksI, type I polyketide (incomplete); pksII, type II polyketide; crt, carotenoid

6.1 Genes for Methylenomycin Production

Following Vivian's initial discovery that *S. coelicolor* A3(2) strains carrying SCP1 inhibit those lacking the plasmid, Wright and Hopwood (1976a) identified the inhibitor as methylenomycin A, an unusual cyclopentanone antibiotic first characterised in *S. violaceoruber* SANK 95570 (to which *S. coelicolor* A3(2) is taxonomically close) by Haneishi et al. (1974). The demonstration that pathway genes for methylenomycin were indeed on SCP1 used a collection of non-producing mutants (*mmy*) (Kirby and Hopwood 1977). Different mutants had different phenotypes in relation to their ability to co-synthesise methylenomycin when grown close to each other, and could be classified as secretors, convertors or non-interacting. Critically, the particular phenotypes were 100% co-transferred with SCP1. Later, segments of *mmy* DNA were cloned from SCP1 in some of the first *Streptomyces* cloning experiments, using plasmid or phage vectors (Bibb and Cohen 1980; Chater and Bruton 1983, 1985), and probes from the cluster were shown to hybridise to SCP1 plasmid DNA (Kinashi et al. 1987). Eventually the entire 20-kb cluster was included in the complete sequence determination of SCP1 (Bentley et al. 2004) (Fig. 7).

The *mmy* gene cluster of *S. violaceoruber* SANK 95770 also lies on a plasmid, pSV1, which is a large *circular* molecule (Okanishi et al. 1980; Aguilar and Hopwood 1882). The pSV1 *mmy* cluster is about 99% identical to that of SCP1 at the DNA sequence level (Chater and Bruton 1985; Yamasaki et al. 2003). (For comparison, most chromosomal housekeeping genes are around 90% identical between *S. coelicolor* and *S. avermitilis*, whose last common ancestor probably existed more than 200 million years ago (Chater and Chandra 2006).) Apart from the *mmy* genes and a set of *parAB* genes, SCP1 and pSV1 share little homology. Examination of the sequences around the ends of the conserved segment revealed no clues to the mechanism of transposition of the cluster between the replicons. This is not very surprising, as it has generally proved difficult to detect such features at the ends of chromosomally located antibiotic biosynthetic clusters, even though their sporadic occurrence suggests that many of them moved into their present host genomes by some kind of lateral transfer. Probably, such lateral transfer events generally took place long enough ago for most insertion-related sequences to have been lost.

Recent studies have shown that the distinction between secretor *mmy* mutants and the single convertor mutant reflects the involvement of an extracellular signalling molecule ("M-factor") that controls the onset of methylenomycin biosynthesis (O'Rouke 2003). Biosynthesis of M-factor depends on a gene of a type, found in many other streptomycetes, that is critical for biosynthesis of gamma-butyrolactone signalling molecules (*mmfL*, Fig. 7). The most well known of these molecules is A-factor, which activates strep-

tomycin production in *S. griseus* (Chater and Horinouchi 2003). Such factors are thought to diffuse easily through membranes, and accumulate around and within the mycelium as it becomes more densely packed, until a concentration is reached at which the factor can interact with a factor-binding protein in the cytoplasm. The best known of these is the A-factor-binding protein specified by the *arpA* gene of *S. griseus*. The factor-binding proteins are usually repressors of other regulatory genes involved in activating antibiotic pathway-specific biosynthetic genes (Takano 2006). The *mmy* gene cluster includes two *arpA*-like genes for likely factor-binding proteins (*mmyR* and *mmfR*, Fig. 7). Thus, SCP1 and pSV1 both confer production of at least two bioactive small molecules on their hosts: M-factor and methylenomycin. Of course, the plasmid linkage of these attributes may have little particular adaptive significance, but it is intriguing that methylenomycin appears selectively to inhibit aerial growth, and hence sporulation, of SCP1$^-$ derivatives of *S. coelicolor* A3(2) (Vivian 1971). We may therefore speculate that methylenomycin production is programmed by the "quorum-sensing" M-factor to take place only when dense mycelium has formed, which is the preferred time for the emergence of sporulating aerial hyphae. At that time the inhibition of growth of SCP1$^-$ aerial hyphae would selectively favour reproduction of SCP1$^+$ hyphae. The absence of such inhibition at earlier growth stages allows the growth of nearby SCP1$^-$ vegetative mycelium, which can potentially be infected by SCP1 to aid plasmid spread.

SCP1 also encodes some spore-associated proteins, SapC, D and E (Guijarro et al. 1988; Bentley et al. 2004; Redenbach et al. 1998). It has tentatively been suggested that these too may be associated with the need for SCP1-containing hyphae to be able to sporulate in the presence of their self-produced antagonist of aerial development (Bentley et al. 2004).

A small number of other antibiotics structurally similar to methylenomycin have been identified (listed in Challis and Chater 2001), but as yet nothing is known about their genetic determinants or biosynthesis. It will be of some interest to know whether any of them are also plasmid-determined, and how their biosynthesis relates to that of methylenomycin.

6.2 pSLA2-L of *Streptomyces rochei*: A Plasmid Densely Packed with Genes for Secondary Metabolism

At 211 kb, pSLA2-L is the largest of the three linear plasmids found in *S. rochei* (Table 1). Using hybridisation probes for polyketide synthase (PKS) genes of two kinds, one often associated with macrolide antibiotics such as erythromycin, the other with polycyclic aromatic antibiotics such as tetracycline, Kinashi et al. (1998) found evidence that both kinds of PKSs were encoded by pSLA2-L. Sequencing and gene disruption experiments showed that the macrolide-related sequence was part of a large gene cluster for biosynthesis

of the 14-membered macrolide lankamycin (Suwa et al. 2000). Completion of the pSLA2-L sequence by Mochizuki et al. (2003) showed that, astonishingly, about three quarters of the entire plasmid was occupied by genes for secondary metabolism (Fig. 7). Thus, in addition to the lankamycin biosynthetic genes and those for an as yet unidentified aromatic polyketide, those for a more unusual 17-membered macrocyclic molecule, lankacidin, were also present, as well as a further gene set for carotenoid biosynthesis. The plasmid also carried genes for the production of a gamma-butyrolactone factor (*srrX*, *S. rochei* regulatory gene X) and its receptor (*srrA*), and these were shown to be needed to activate the synthesis of both lankacidin and lankamycin (Arakawa et al. 2007). The remarkable conjunction of secondary metabolism pathways on pSLA2-L indicates that linear plasmids might provide a means by which different sets of pathway genes can be brought into close proximity, in a context that might facilitate the evolution of new pathways by a natural version of "combinatorial biosynthesis" (a term applied to the artificial mixing and matching of genes from different pathways to generate new and useful metabolites; see Reynolds 1998, for a brief review).

The co-regulation of two adjacent antibiotic production gene sets on pSLA2-L is also reminiscent of the situation in the *S. clavuligerus* chromosome, in which the adjacent gene sets for clavulanic acid and cephamycin C biosynthesis are controlled by a common regulatory gene (Ward and Hodgson 1993; Perez-Llarena et al. 1997). The actions of the latter two secondary metabolites are synergistic, since beta-lactamases that might destroy cephamycin C are inhibited by clavulanic acid. Another example of such a chromosomal supercluster is that for the production of the two-component antibiotic pristinamycin, although in that case the genes for pristinamycin A and B are intermixed rather than being in discrete adjacent clusters (Bamas-Jacques et al. 1999). Perhaps, then, lankacidin and lankamycin may also have synergistic action. Moreover, their apparent regulation by a quorum-sensing molecule might imply that, as hypothesised for methylenomycin, they provide an advantage to pSLA2-L-containing organisms that is best expressed after the main growth period, so avoiding inhibition of the vegetative growth of possible recipient streptomycetes, which might prevent pSLA2-L transfer. As in the case of methylenomycin, no clues to the mechanism by which the different clusters arrived in pSLA2-L could be divined from sequence analysis.

7
Linear Plasmids and Evolution

It is likely that the linearity of *Streptomyces* chromosomes originally resulted from Campbell-type recombination between a circular genome and a linear plasmid (Volff and Altenbuchner 2000; Chen et al. 2002). Although it has been

suggested that this occurred many times independently within the *Streptomyces* lineage (Chen et al. 2002), it is also attractive to consider the case for a unitary linearisation event at the birth of the genus. On the assumption that diverse linear plasmids continued to arise, their ends would have readily been moved into and between the linear chromosomes of streptomycetes, both early and more recently in their evolution, as illustrated by the phenomenology described in this chapter (see also Hopwood 2006). This would have accelerated the rate of diversification of *Streptomyces* genomes at a critical point in their early evolution which, it has been argued, took place when a vast new ecological niche first became available on the earth through colonisation of the land by green plants about half a billion years ago (Chater and Chandra 2006). In this final section, we consider linear plasmids in the context of the evolution of their hosts.

7.1 Evolutionarily and Adaptively Significant Genes Carried by Linear Plasmids

As is often the case for large plasmids, both circular and linear, a relatively high proportion of SCP1 genes (57%) show no significant database matches, and a further 20% resemble database entries of unknown function. The equivalent numbers for the host chromosome are 23 and 30%. This reinforces the idea that plasmids are an important route by which new genes enter their hosts, though it has been argued elsewhere that fundamentally novel genes probably arise more often in bacteriophages (Chater and Chandra 2006). It is interesting to note that one of the linear plasmids, SCP1, is a vehicle for members of several families of regulatory genes that are particularly important in streptomycetes. One such family is the *arpA*-like genes referred to above, representatives of which (*mmfR* and *mmyR*) function in methylenomycin production. Another comprises genes for ECF (extracytoplasmic function) sigma factors, which are particularly abundant in streptomycetes (e.g. nearly 50 each in *S. coelicolor* and *S. avermitilis*, compared to one in *E. coli* K-12; Bentley et al. 2002; Ikeda et al. 2003): three of these are present on SCP1. The plasmid also carries three genes belonging to the *whiB*-like family (*wbl* genes; Soliveri et al. 2000). Such *wbl* genes are confined to actinomycetes, where they are abundant and play important roles in cell division/development and antibiotic resistance (Chater and Chandra 2006). (Remarkably, one of the *wbl* genes (*wblP*; SCP1.161c) encodes a Wbl-ECF sigma factor fusion protein, reinforcing evidence that Wbl proteins directly interact with sigma factors (Steyn et al. 2002).) SCP1 also contains a representative of the family of *abaA*-like clusters found only in streptomycetes, in which there are many such gene sets. In *S. coelicolor*, the prototype of these clusters, *abaA*, is involved in the regulation of antibiotic production (Fernandez-Moreno et al. 1992), while another includes the

developmental gene *whiJ* (Gehring et al. 2000). The SCP1 *abaA*-related paralogues (SCP1.58c–60 and 295–293c: the cluster is present in both TIRs) are co-transcribed with the *sapCDE* genes for spore-associated proteins. Thus, SCP1 may be representative of the mechanism of expansion and spread of these regulatory gene families, so important in the contexts of secondary metabolism, differentiation and cellular homeostasis/stress responses. Frequent lateral transfer of such genes among streptomycetes is also indicated by the fact that, in each species, a set of universally conserved versions of all these families is augmented by species-specific chromosomal versions, often located in the poorly conserved TIR regions.

Other large linear plasmids are much less rich in such families of regulatory genes, but may help to bring about other aspects of genome evolution. As discussed in the preceding section, pSLA2-L illustrates the possibility that linear plasmids help to bring sets of antibiotic production genes together as a prelude to the selection of synergistic cooperations, while the movement of genes for oxytetracycline biosynthesis between the *S. rimosus* chromosome and linear plasmid pPZG101 illustrates how readily such chromosomal gene sets might be mobilised to other streptomycetes. Such recombinational exchanges are likely to be helped by the frequent occurrence of transposable DNA in some (but not all) linear plasmids (Fig. 8) and chromosome ends. The occurrence of transposable elements near chromosome ends may well reflect exchanges with the ends of linear plasmids.

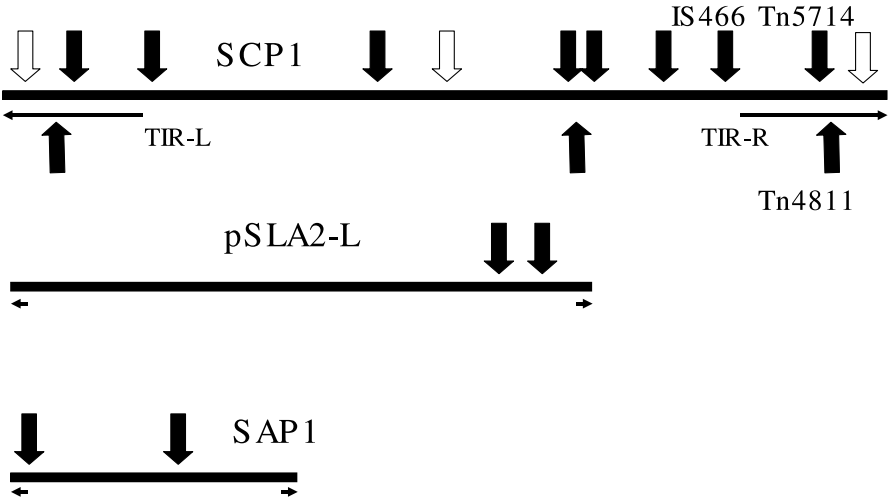


Fig. 8 Transposable elements in three large linear plasmids. See legend to Fig. 7 for details of the three plasmids

7.2 Possible Role of Linear Plasmids in the Dissemination of TTA-Containing Genes Subject to the Influence of the Developmental Gene *bldA*

Bentley et al. (2004) noted that SCP1 contains a higher frequency (5%) of genes carrying the leucine codon TTA than is observed in the chromosome (2%). This codon is rare in GC-rich DNA (*Streptomyces* genomes contain more than 70% GC), and is confined to genes inessential for vegetative growth, since a mutation lacking the gene (*bldA*) encoding the only tRNA capable of efficiently translating UUA codons is fully viable (Lawlor et al. 1987; Leskiw et al. 1991). Such mutants are, however, defective in production of several antibiotics, and fail to develop aerial hyphae on most media. *bldA* mutants of diverse streptomycetes all have a similar phenotype, indicating that the specialised use of TTA codons developed early in the evolution of the genus (reviewed by Chater 2006, and Chater and Chandra 2006). More than 90% of the TTA-containing genes of *S. coelicolor* are non-conserved between species, and therefore likely to have been acquired by lateral transfer, and mutations in most of them have no recognisable phenotype, leading Li et al. (2007) to suggest that they may be important only in particular ecological contexts. The relatively high frequency of TTA-containing genes in SCP1 would be consistent with the notion that some linear plasmids act as vehicles for the movement of ecologically adaptive genes between the chromosomes of different streptomycetes. Two of the TTA-containing genes of SCP1 are in the methylenomycin gene cluster, which causes production of methylenomycin to be defective in a *bldA* mutant (O'Rourke 2003), and another is in the large operon containing the *sapCDE* genes, encoding spore-associated proteins (Bentley et al. 2004).

7.3 Co-evolution of Streptomycetes and Their Linear Replicons

As we have seen, there is abundant evidence of exchanges between linear plasmids and host chromosomes or circular plasmids. "Fossil" evidence that one linear plasmid can integrate into another has also been found in DNA sequences: sequences with the same kind of secondary structure-potentiating features as plasmid ends are embedded within SCP1 (Bentley et al. 2004) (Fig. 9). When this is coupled with the fact that, where adequately studied, most linear plasmids are highly transmissible, it is reasonable to conclude that linear plasmids have contributed hugely to genome evolution, and therefore to host evolution, in streptomycetes (Volff and Altenbuchner 2000). Perhaps the evolutionary emergence of *Streptomyces* owes something to this genome fluidity. Indeed, the apparently continual acquisition of species-specific features, such as antibiotic biosynthetic gene clus-

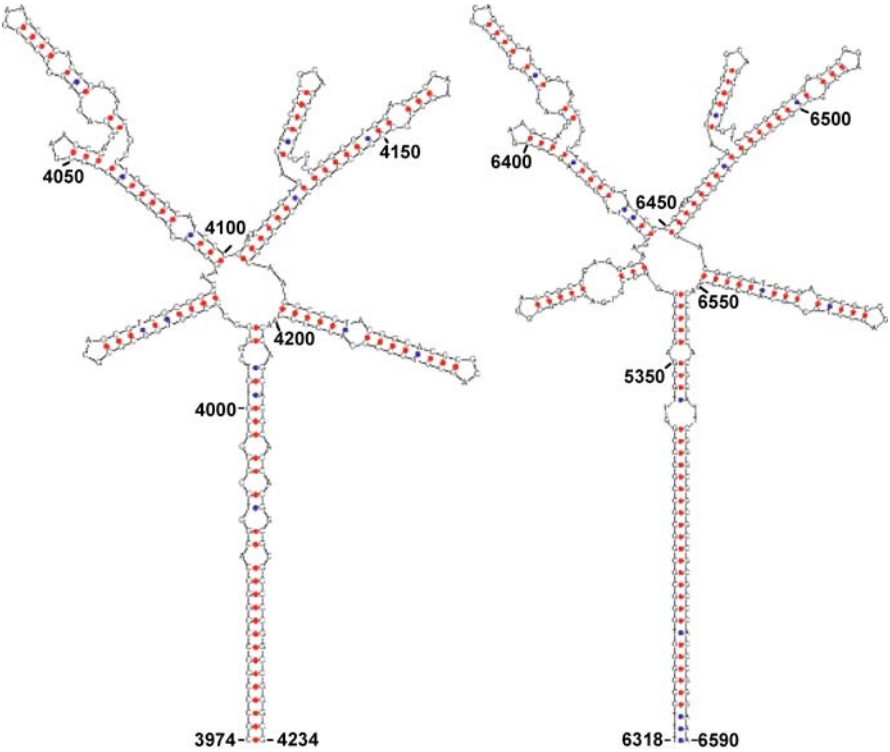


Fig. 9 Pseudotelomeres in SCP1. The predicted secondary structures with the minimum energy of two pseudotelomere sequences are shown

ters and transposable elements, since the origin of the genus suggests that this dynamic interaction continues to the present. Many of the ca. 50% of species-non-conserved genes of streptomycetes are in the arms of the chromosomes, while nearly all the generally conserved genes are in the central ca. 5-Mb core, with a gradient of species-specific genes separating the core from the arms (Choulet et al. 2006). This suggests the idea that laterally acquired arm genes subject to relatively frequent selection would migrate inwards towards the core, as a result of deletions and recombinational exchanges taking place in the arms, and so become increasingly immune from the many kinds of genetic instability events in the chromosome ends (Chater and Chandra 2006).

The possibility of new configurations at the ends of linear chromosomes, particularly assisted by the introduction of linear plasmids, offers great possibilities for genome adaptation that may account for the enormous success of streptomycetes in colonising soils. The ready chromosomal acquisition of genes from linear plasmids would mean that these genes could be retained over many generations even in the absence of any selection, whereas plasmids

themselves, and the genes that they carry, can often be lost when selection is relaxed. This more stable inheritance would be valuable for streptomycetes, which need to be able to respond flexibly, and probably very intermittently, to the great variety of microenvironments that they may encounter in the soil. The multigenomic compartments of hyphae offer havens in which altered, and possibly defective, chromosomes can be supported by non-defective versions for short periods of growth, until they must eventually segregate during sporulation; and to this the TIRs of linear replicons add a kind of partial diploidy capable of transmission through spores, so that recessive mutations in TIRs can potentially be propagated through many generations, as in diploid eukaryotes. This might permit accelerated evolution compared with unicellular organisms possessing circular chromosomes and plasmids. As pointed out by Volff and Altenbuchner (2000), an exchange of plasmid and chromosomal ends could also provide a different kind of partial diploidy, in which genes present both on the chromosome and on an un-rearranged copy of the plasmid could be maintained in the diploid state through sporulation, with the same possibilities for the preservation of recessive mutations.

Why, if linear genomes are beneficial to streptomycetes, are they not a general feature of all bacteria? One possibility is that linear genomes are not very well-suited to the efficient and complete partitioning of chromosomes into daughter cells that is crucial for unicellular organisms. Specialised mechanisms exist for decatenation and resolution of newly replicated circular chromosomes of such organisms (Sherratt et al. 2004). Highly reliable partitioning may be sacrificed in favour of increased adaptability in a vegetative lifestyle involving multigenomic compartments and relatively few cell division events.

Acknowledgements We thank David Hopwood and Carton Chen for helpful comments on the manuscript.

References

Aguilar A, Hopwood DA (1982) Determination of methylenomycin A synthesis by the pSV1 plasmid from *Streptomyces violaceus-ruber* SANK 95570. J Gen Microbiol 128:1893–1901

Arakawa K, Mochizuki S, Yamada K, Noma T, Kinashi H (2007) (gamma)-butyrolactone autoregulator-receptor system involved in lankacidin and lankamycin production and morphological differentiation in *Streptomyces rochei*. Micobiology 153:1817–1827

Bamas-Jacques N, Lorenzon S, Lacroix P, de Swetschin C, Crouzet J (1999) Cluster organization of the genes of *Streptomyces pristinaespiralis* involved in pristinamycin biosynthesis and resistance elucidated by pulsed-field gel electrophoresis. J Appl Microbiol 87:939–948

Bao K, Cohen SN (2001) Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. Genes Dev 15:1518–1527

Bentley SD, Chater KF, Cerdeno-Tarraga A-M, Challis GL, Thompson NR, James KD et al (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147

Bentley SD, Brown S, Murphy LD, Harris DE, Quail MA, Parkhill J, Barrel BG, McCormick JR, Santamaria RI, Losick R, Yamasaki M, Kinashi H, Chen CW, Chandra G, Jakimowicz D, Kieser HM, Kieser T, Chater KF (2004) SCP1, a 356,023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). Mol Microbiol 51:1615–1628

Bibb MJ, Freeman RF, Hopwood DA (1977) Physical and genetical characterisation of a second sex factor, SCP2, for *Streptomyces coelicolor* A3(2). Mol Gen Genet 154:155–166

Bibb M, Schottel JL, Cohen SN (1980) A DNA cloning system for interspecies gene transfer in antibiotic-producing *Streptomyces*. Nature 284:526–531

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Carle GF, Olson MV (1984) Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. Nucleic Acids Res 12:5647–5664

Challis GL, Chater KF (2001) Incorporation of [U-^{13}C]glycerol defines plausible early steps for the biosynthesis of methylenomycin A in *Streptomyces coelicolor* A3(2). Chem Commun, pp 935–936

Chardon-Lauriaux I, Charpentier M, Percheron F (1986) Isolation and characterization of a linear plasmid from *Streptomyces rimosus*. FEMS Microbiol Lett 35:153–155

Chater KF (1998) Taking a genetic scalpel to the *Streptomyces* colony. Microbiology 144:1465–1478

Chater KF (2006) *Streptomyces* inside out: a new perspective on the bacteria that provide us with antibiotics. Phil Trans R Soc Lond B Biol Sci 361:761–768

Chater KF, Bruton CJ (1983) Mutational cloning in *Streptomyces* and the isolation of antibiotic production genes. Gene 26:67–78

Chater KF, Bruton CJ (1985) Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. EMBO J 4:1893–1897

Chater KF, Chandra G (2006) The evolution of development in *Streptomyces* analysed by genome comparisons. FEMS Microbiol Rev 30:651–672

Chater KF, Horinouchi S (2003) Signalling early developmental events in two highly diverged *Streptomyces* species. Mol Microbiol 48:9–15

Chen CW (1996) Complications and implications of linear bacterial chromosomes. Trends Genet 12:192–196

Chen CW (2007) *Streptomyces* Linear Plasmids: Replication and Telomeres (in this volume)

Chen CW, Yu T-W, Lin Y-S, Kieser HM, Hopwood DA (1993) The conjugative plasmid SLP2 of *Streptomyces lividans* is a 50-kb linear molecule. Mol Microbiol 7:925–932

Chen CW, Huang C-H, Lee H-H, Tsai H-H, Kirby R (2002) Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. Trends Genet 18:522–529

Choulet F, Aigle B, Gallois F, Mangenot S, Gerbaut C, Truong C, Francou F-X, Fourrier C, Guerineau M, Decaris B, Barbe E, Pernodet J-L, Leblond P (2006) Evolution of the terminal regions of the *Streptomyces* linear chromosome. Mol Biol Evol 23:2361–2369

Clewell DB, Helinski DR (1969). Supercoiled circular DNA–protein complexes in *Escherichia coli*: purification and conversion to an open-circular DNA form. Proc Natl Acad Sci USA 62:1159–1166

Fernandez-Moreno MA, Martin-Triana AJ, Martinez E, Niemi J, Kieser H, Hopwood DA, Malpartida F (1992) *abaA*, a new pleiotropic regulatory locus for antibiotic production in *Streptomyces coelicolor*. J Bacteriol 174:2958–2967

Fetzner S (2007) Catabolic Linear Plasmids (in this volume)

Gehring AM, Nodwell JR, Beverley SM, Losick R (2000). Genome-wide insertional mutagenesis in *Streptomyces* reveals additional genes involved in morphological differentiation. Proc Natl Acad Sci USA 97:9642–9647

Gravius B, Glocker D, Pigac J, Pandza K, Hranueli D, Cullum J (1994) The 387-kb linear plasmid pPZ101 of *Streptomyces rimosus* and its interactions with the chromosome. Microbiology 140:2271–2277

Guijarro J, Santamaria R, Schauer A, Losick R (1988) Promoter determining the timing and spatial localization of a cloned *Streptomyces coelicolor* gene encoding a spore-associated polypeptide. J Bacteriol 170:1895–1901

Hanafusa T, Kinashi H (1992) The structure of an integrated copy of the giant linear plasmid SCP1 in the chromosome of *Streptomyces coelicolor* 2612. Mol Gen Genet 231:363–368

Haneishi T, Terahara A, Arai M, Hata T, Tamura C (1974) New antibiotics, methylenomycins A and B. 2. Structures of methylenomycins A and B. J Antibiot (Tokyo) 27:393–399

Hayes W (1964) The genetics of bacteria and their viruses. Blackwell, Oxford

Hertwig S (2007) Linear Plasmids and Prophages in Gram Negative Bacteria (in this volume)

Hirochika H, Sakaguchi K (1982) Analysis of linear plasmids isolated from *Streptomyces*: association of protein with the ends of the plasmid DNA. Plasmid 7:59–65

Hirochika H, Nakamura K, Sakaguchi K (1984) A linear plasmid from *Streptomyces rochei* with a terminal inverted repeat of 614 base pairs. EMBO J 3:761–766

Hodgson DA, Chater KF (1981) A chromosomal locus controlling extracellular agarase production by *Streptomyces coelicolor* A3(2), and its inactivation by chromosomal integration of plasmid SCP1. J Gen Microbiol 124:339–348

Holsters M (2007) Linear Plasmids and Phytopathogenicity (in this volume)

Hopwood DA (1967) Genetic analysis and genome structure in *Streptomyces coelicolor* A3(2). Bacteriol Rev 31:373–403

Hopwood DA (1999) Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico. Microbiology 145:2183–2202

Hopwood DA (2006) Soil to genomics: the *Streptomyces* chromosome. Annu Rev Genet 40:1–23

Hopwood DA (2007) *Streptomyces* in nature and medicine: the antibiotic makers. Oxford University Press, New York

Hopwood DA, Merrick M (1977) Genetics of antibiotic production. Bacteriol Rev 41:595–635

Hopwood DA, Wright HM (1973a) Transfer of a plasmid between *Streptomyces* species. J Gen Microbiol 77:187–195

Hopwood DA, Wright HM (1973b) A plasmid of *Streptomyces coelicolor* carrying a chromosomal locus and its inter-specific transfer. J Gen Microbiol 79:331–342

Hopwood DA, Wright HM (1976a) Genetic studies on SCP1-prime strains of *Streptomyces coelicolor* A3(2). J Gen Microbiol 95:107–120

Hopwood DA, Wright HM (1976b) Interactions of the plasmid SCP1 with the chromosome of *Streptomyces coelicolor* A3(2). In: MacDonald KD (ed) Proceedings of the 2nd international symposium on the genetics of industrial micro-organisms. Academic Press, London, pp 607–619

Hopwood DA, Wright HM (1983) CDA is a new chromosomally determined antibiotic from *Streptomyces coelicolor* A3(2). J Gen Microbiol 129:3575–3579

Hopwood DA, Harold RJ, Vivian A, Ferguson HM (1969) A new kind of fertility variant in *Streptomyces coelicolor*. Genetics 62:461–477

Hopwood DA, Chater KF, Dowding JE, Vivian A (1973) Recent advances in *Streptomyces coelicolor* genetics. Bacteriol Rev 37:371–405

Hopwood DA, Bibb MJ, Ward JM, Westpheling J (1979) Plasmids in *Streptomyces coelicolor* and related species. In: Timmis KN, Puhler A (eds) Plasmids of medical, environmental and commercial importance. Elsevier, Amsterdam, pp 245–258

Huang CH, Chen CY, Tsai HH, Chen C, Lin YS, Chen CW (2003) Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. Mol Microbiol 47:1563–1576

Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S (2004) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 21:526–531

Kalkus JR, Menne R, Reh M, Schlegel HG (1998) The terminal structures of linear plasmids from *Rhodococcus opacus*. Microbiology 144:1271–1279

Keen CL, Mendelovitz S, Cohen G, Aharonowitz Y, Roy KL (1988). Isolation and characterization of a linear DNA plasmid from *Streptomyces clavuligerus*. Mol Gen Genet 212:172–176

Kim HJ, Calcutt MJ, Schmidt FJ, Chater KF (2000) Partitioning of the linear chromosome during sporulation of *Streptomyces coelicolor* A3(2) involves an *oriC*-linked *parAB* locus. J Bacteriol 182:1313–1320

Kinashi H (1994) Linear plasmids from actinomycetes. Actinomycetologica 8:87–96

Kinashi H, Shimaji M (1987) Detection of giant linear plasmids in antibiotic producing strains of *Streptomyces* by the OFAGE technique. J Antibiot 40:913–916

Kinashi H, Shimaji-Murayama M (1991) Physical characterization of SCP1, a giant linear plasmid from *Streptomyces coelicolor*. J Bacteriol 173:1523–1529

Kinashi H, Shimaji M, Sakai A (1987) Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. Nature 328:454–456

Kinashi H, Otten SL, Duncan JS, Hutchinson CR (1988) Frequent loss and restoration of antibiotic production by *Streptomyces lasaliensis*. J Antibiot 41:624–637

Kinashi H, Murayama M, Hanafusa T (1992) Integration of SCP1, a giant linear plasmid, into the *Streptomyces* chromosome. Gene 115:35–41

Kinashi H, Murayama M, Matsushita H, Nimi O (1993) Structural analysis of the giant linear plasmid SCP1 in various *Streptomyces coelicolor* strains. J Gen Microbiol 139:1261–1269

Kinashi H, Mori E, Hatani A, Nimi O (1994) Isolation and characterization of linear plasmids from lankacidin-producing *Streptomyces* species. J Antibiot (Tokyo) 47:1447–1455

Kinashi H, Fujii S, Hatani A, Kurokawa T, Shinkawa H (1998) Physical mapping of the linear plasmid pSLA2-L and localization of the *eryAI* and *actI* homologs. Biosci Biotechnol Biochem 62:1892–1897

Kirby R, Hopwood DA (1977) Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). J Gen Microbiol 98:239–252

Kirby R, Wright LF, Hopwood DA (1975) Plasmid-determined antibiotic synthesis and resistance in *Streptomyces coelicolor*. Nature 254:265–267

Lawlor EJ, Baylis H, Chater KF (1987) Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). Genes Dev 1:1305–1310

Leskiw BK, Lawlor EJ, Fernandez-Abalos JM, Chater KF (1991) TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. Proc Natl Acad Sci USA 88:2461–2465

Li W, Wu J, Tao W, Zhao C, Wang Y, He X, Chandra G, Zhou X, Deng Z, Chater KF, Tao M (2007) A genetic and bioinformatic analysis of *Streptomyces coelicolor* genes containing TTA codons, possible targets for regulation by a developmentally significant tRNA. FEMS Microbiol Lett 266:20–28

Lin YS, Kieser HM, Hopwood DA, Chen CW (1993) The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol Microbiol 10:923–933

Mochizuki S, Hiratsu K, Suwa M, Ishii T, Sugino F, Yamada K, Kinashi H (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. Mol Microbiol 48:1501–1510

Molle V, Palframan WJ, Findlay KC, Buttner MJ (2000) WhiD and WhiB, homologous proteins required for different stages of sporulation in *Streptomyces coelicolor* A3(2). J Bacteriol 182:1286–1295

Okanishi M, Ita T, Umezawa H (1970) Possible control of formation of aerial mycelium and antibiotic production in *Streptomyces* by episomic factors. J Antibiot (Tokyo) 23:45–47

Okanishi M, Manome T, Umezawa H (1980) Isolation and characterization of plasmid DNAs in actinomycetes. J Antibiot 33:88–91

O'Rourke S (2003) PhD Thesis, University of East Anglia, Norwich, UK

Pandza S, Biukovic G, Paravic A, Dadbin A, Cullum J, Hranueli D (1998) Recombination between the linear plasmid pPZG101 and the linear chromosome of *Streptomyces rimosus* can lead to exchange of ends. Mol Microbiol 28:1165–1176

Perez-Llarena FJ, Liras P, Rodriguez-Garcia A, Martin JF (1997) A regulatory gene (*ccaR*) required for cephamycin and clavulanic acid production in *Streptomyces clavuligerus*: amplification results in overproduction of both beta-lactam compounds. J Bacteriol 179:2053–2059

Rathos MJ, Verma NC, Notani NK (1989) Separation by pulsed-field gradient gel electrophoresis of giant linear plasmids from antibiotic-producing strains of *Streptomyces* and *Nocardia*. Curr Sci 58:1235–1239

Redenbach M, Ikeda K, Yamasaki M, Kinashi H (1998) Cloning and physical mapping of the *Eco*RI fragments of the giant linear plasmid SCP1. J Bacteriol 180:2796–2799

Redenbach M, Bibb M, Gust B, Seitz B, Spychaj A (1999) The linear plasmid SCP1 of *Streptomyces coelicolor* A3(2) possesses a centrally located replication origin and shows significant homology to the transposon Tn4811. Plasmid 42:174–185

Reynolds KA (1998) Combinatorial biosynthesis: lesson learned from nature. Proc Natl Acad Sci USA 95:12744–12746

Rudd BAM, Hopwood DA (1979) Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). J Gen Microbiol 114:35–43

Rudd BAM, Hopwood DA (1980) A pigmented mycelial antibiotic in *Streptomyces coelicolor*: control by a chromosomal gene cluster. J Gen Microbiol 119:333–340

Sakaguchi K (1990) Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. Microbiol Rev 54:66–74

Schrempf H, Bujard H, Hopwood DA, Goebel W (1975) Isolation of covalently closed circular deoxyribonucleic acid from *Streptomyces coelicolor* A3(2). J Bacteriol 121:416–421

Schwarz DC, Cantor CR (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37:67–75

Sermonti G, Casciano S (1963) Sexual polarity in *Streptomyces coelicolor*. J Gen Microbiol 33:293–301

Snyder L, Champness W (2002) Molecular genetics of bacteria, 2nd edn. ASM, Washington DC

Sherratt DJ, Soballe B, Barre FX, Filipe S, Lau I, Massey T, Yates J (2004) Recombination and chromosome segregation. Philos Trans R Soc Lond B Biol Sci 359:61–69

Soliveri JA, Gomez J, Bishai WR, Chater KF (2000) Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene *whiB* are present in *Streptomyces* and other actinomycetes. Microbiology 146:333–343

Spatz K, Kohn H, Redenbach M (2002) Characterization of the *Streptomyces violaceoruber* SANK 95570 plasmids pSV1 and pSV2. FEMS Microbiol Lett 213:87–92

Steyn AJ, Collins DM, Hondalus MK, Jacobs WR Jr, Kawakami RP, Bloom BR (2002) *Mycobacterium tuberculosis* WhiB3 interacts with RpoV to affect host survival but is dispensable for in vivo growth. Proc Natl Acad Sci USA 99:3147–3152

Suwa M, Sugino H, Sasaoka A, Mori E, Fujii S, Shinkawa H, Nimi O, Kinashi H (2000) Identification of two polyketide synthase gene clusters on the linear plasmid pSLA2-L in *Streptomyces rochei*. Gene 246:123–131

Takano E (2006) Gamma-butyrolactones: small signalling molecules that regulate antibiotic production and differentiation. Curr Opin Microbiol 9:287–294

Vivian A (1971) Genetic control of fertility in *Streptomyces coelicolor* A3(2): plasmid involvement in the interconversion of UF and IF strains. J Gen Microbiol 69:353–364

Vivian A, Hopwood DA (1970) Genetic control of fertility in *Streptomyces coelicolor* A3(2): the IF fertility type. J Gen Microbiol 64:101–117

Vivian A, Hopwood DA (1973) Genetic control of fertility in *Streptomyces coelicolor* A3(2): new kinds of donor strains. J Gen Microbiol 76:147–162

Volff JN, Altenbuchner J (2000) A new beginning with new ends: linearisation of circular chromosomes during bacterial evolution. FEMS Microbiol Lett 186:143–150

Ward JM, Hodgson JE (1993) The biosynthetic genes for clavulanic acid and cephamycin production occur as a "super-cluster" in three *Streptomyces*. FEMS Microbiol Lett 110:239–242

Weaver D, Karoonuthaisiri N, Tsai HH, Huang CH, Ho ML, Gai S, Patel KG, Huang J, Cohen SN, Hopwood DA, Chen CW, Kao CM (2004) Genome plasticity in *Streptomyces*: identification of 1 Mb TIRs in the *S. coelicolor* A3(2) chromosome. Mol Microbiol 51:1535–1550

Wildermuth H (1970) Development and organization of the aerial mycelium in *Streptomyces coelicolor*. J Gen Microbiol 60:43–50

Wright LF, Hopwood DA (1976a) Identification of the antibiotic determined by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). J Gen Microbiol 95:96–106

Wright LF, Hopwood DA (1976b) Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolor* A3(2). J Gen Microbiol 96:289–297

Wu X, Roy K (1993) Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus* and characterisation of its RNA transcripts. J Bacteriol 175:37–52

Yamasaki M, Kinashi H (2004) Two chimeric chromosomes of *Streptomyces coelicolor* A3(2) generated by single crossover of the wild-type chromosome and linear plasmid SCP1. J Bacteriol 186:6553–6559

Yamasaki M, Miyashita K, Cullum J, Kinashi H (2000) A complex insertion sequence cluster at a point of interaction between the linear plasmid SCP1 and the linear chromosome of *Streptomyces coelicolor* A3(2). J Bacteriol 182:3104–3110

Yamasaki M, Redenbach M, Kinashi H (2001) Integrated structures of the linear plasmid SCP1 in two bidirectional donor strains of *Streptomyces coelicolor* A3(2). Mol Gen Genet 264:634–642

Yamasaki M, Ikuto Y, Ohira A, Chater K, Kinashi H (2003) Limited regions of homology between linear and circular plasmids encoding methylenomycin biosynthesis in two independently isolated streptomycetes. Microbiology 149:1351–1356

Yang MC, Losick R (2001) Cytological evidence for association of the ends of the linear chromosome in *Streptomyces coelicolor*. J Bacteriol 183:5180–5186

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_090/Published online: 23 June 2007

Streptomyces Linear Plasmids: Replication and Telomeres

Carton W. Chen

Department of Life Sciences and Institute of Genome Sciences,
National Yang-Ming University, Shih-Pai, 112 Taipei, Taiwan
cwchen@ym.edu.tw

1 Introduction . . . 34
1.1 Unique System of Replication . . . 34

2 Terminal DNA . . . 35
2.1 Archetypal Telomeres . . . 35
2.2 Nonarchetypal Telomeres . . . 38
2.3 Terminal Inverted Repeats . . . 38
2.4 Pseudotelomeres . . . 39

3 Terminal Proteins . . . 40
3.1 Archetypal TPs . . . 40
3.2 Nonarchetypal TPs: The TP of SCP1 . . . 42
3.3 TP–TP Interactions . . . 43

4 Replication . . . 44
4.1 Replication Origin and Replication Initiation . . . 44
4.2 Elements Required for Replication in Linear Form . . . 45

5 End Patching . . . 46
5.1 The Mechanisms . . . 46
5.2 Protein Players . . . 48
5.3 DNA Templates . . . 49
5.4 DNA Polymerase . . . 49

6 Postreplicational Segregation . . . 50
6.1 The Plasmid-Encoded *parAB–parS* System . . . 50
6.2 Resolution of Sister Plasmids . . . 52

7 Conjugal Transfer . . . 54
7.1 Mode of Transfer . . . 54
7.2 Nuclear Localization . . . 55

8 Concluding Remarks . . . 56

References . . . 57

Abstract Vegetative replication of terminal protein (TP)-capped linear plasmids (and linear chromosomes) of *Streptomyces* proceeds in two steps: a classical bidirectional replication from an internal origin followed by a novel TP-primed DNA synthesis that patches the resulting single-strand gaps at the 3′ ends ("end patching"). Replication initiation systems found on different linear *Streptomyces* plasmids consist of helicase-like genes and

iterons of relatively diverse origins. In contrast, the end patching system (including the telomeres and the TPs) is highly conserved in most linear replicons in *Streptomyces* with only a single exception so far. Both the TPs and the telomeric DNAs have evolved structural features to serve replication as well as protection of the telomeres. Interaction of TPs shapes the linear replicon into a circular form, which would allow generation of biologically important superhelicity in terminal DNA, but would also create complications during postreplicational segregation of the daughter DNA. TPs may also be involved in priming initiation of replication during conjugal transfer. Like the T-DNA transfer system in *Agrobacterium tumefaciens*, the TPs can target themselves and the attached DNA into eukaryotic nuclei, thus suggesting a possibility of interkingdom conjugal transfer.

1 Introduction

1.1 Unique System of Replication

Linear plasmids are very common among species of *Streptomyces* and related actinomycetes (such as mycobacteria and rhodococci). They are often accompanied by circular plasmids in the same host. The history of the discovery of linear plasmids in *Streptomyces* is described in this volume by Chater and Kinashi.

Like their circular counterparts in *Streptomyces*, linear plasmids are generally conjugative, but rarely carry genes conferring resistance to toxic compounds or other obvious selective advantage (Hopwood and Kieser 1993; Chater and Kinashi, in this volume). Their evolutionary advantage perhaps lies mainly in their ability to promote efficient genetic exchange between their hosts and other species in the soil.

Linear plasmids of *Streptomyces* and related actinomycetes represent a novel class of prokaryotic linear replicons capped by "terminal proteins" (TPs) covalently bound at the 5′ ends (Hirochika et al. 1985; Sakaguchi 1990; Yang et al. 2002). The same structure is shared by the linear chromosomes of *Streptomyces* (reviewed by Chaconas and Chen 2005) and some closely related actinomycetes, such as *Actinoplanes philippinensis*, *Micromonospora chalcea*, *Nocardia asteroides*, *Streptoverticillium abikoense*, *Streptoverticillium cinnamoneus* (Chen 2000; Redenbach et al. 2000), and *Rhodococcus* sp. RHA1 (McLeod et al. 2006).

Such TP-capped linear DNA molecules are also found as plasmids and viruses of both eukaryotes and other prokaryotes (Fetzner et al., in this volume; Klassen and Meinhardt, in this volume). The best-studied are bacteriophage ϕ29 of *Bacillus subtilis* and the adenoviruses. However, beyond their TP-capped structures, the linear plasmids of actinomycetes are very different from these two model systems in structure and function. The fundamental difference concerns their replication.

ϕ29 and adenovirus genomes replicate by an end-to-end mechanism, in which the TP serves as the primer for replication initiation. In such replication, no discontinuous synthesis is involved, and therefore there are no Okazaki fragments. The telomeres are the origin of replication. The replication of the linear plasmids and chromosomes of *Streptomyces* (and presumably other actinomycete linear replicons also), on the other hand, is initiated from an internal origin, and proceeds bidirectionally toward the ends. Such inside-out replication would leave single-stranded gaps at the 3′ ends of the DNA molecules, creating a classical end problem that must be solved so that the DNA would not be shortened with each round of replication. In eukaryotic chromosomes, the solution is extension of the 3′ ends by the action of telomerase. In *Streptomyces*, instead, the single-stranded gaps of the linear plasmids (and chromosomes) are filled in ("end patching") to produce termini of defined length and sequence.

End patching of the 3′ overhangs during replication of *Streptomyces* linear replicons appears to use a previously unknown mechanism. Unlike TP-primed DNA synthesis during replication of ϕ29 and adenoviruses, which uses double-stranded DNA templates, the substrate for end patching in *Streptomyces* is a single-stranded template of 250 to 320 nucleotides (nt) (Chang and Cohen 1994; C.-H. Huang, unpublished results). The structure of this stretch of telomere DNA is much more complicated and information-rich than those of ϕ29 and adenoviruses. It must not only provide a proper substrate for the end patching reaction but also help to escape attack by cellular nucleases.

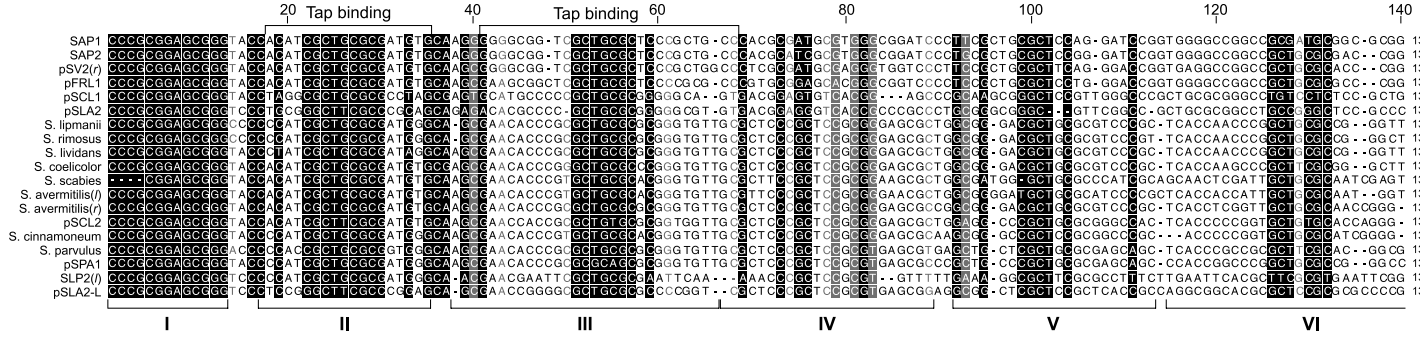
In this chapter, I summarize progress in studying replication of the linear plasmids of *Streptomyces* during growth and conjugal transfer with particular emphasis on the events at the telomeres and the roles of two of the major players: the telomere DNA and the TP. Postreplicational processes, such as segregation and partitioning, and some remarkable functions of the TPs will also be discussed.

2
Terminal DNA

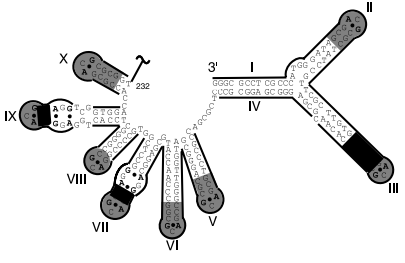
2.1
Archetypal Telomeres

The telomeric sequences of many linear chromosomes and plasmids of *Streptomyces* are available. The telomeric sequences of most *Streptomyces* linear plasmids characterized so far resemble those of *Streptomyces* chromosomes. The term "archetypal" is used here to describe this group of conserved telomeric sequences, which need to be distinguished from a number of atypical telomeres (such as SCP1, pRL1, pRL2, and the *S. griseus* chromosome; see

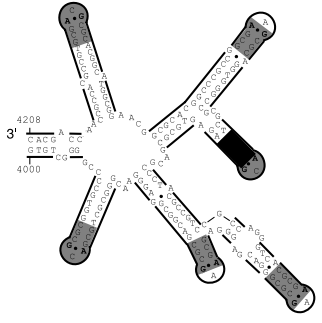
A. Archetypal telomeres



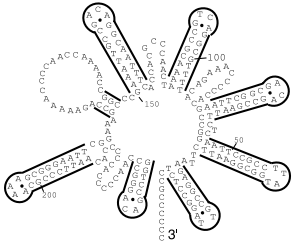
B. *S. lividans*



C. SCP1 pseudotelomere



D. SCP1



below). The archetypal telomeric sequences are most conserved over approximately the first 40 nt, and diverge increasingly away from this point (Fig. 1a). In most archetypal telomeres, the first 13 nucleotides form a palindromic sequence (Palindrome I). Palindrome I is followed by several tightly packed palindromes, some lying next to each other without interruption. In many (but not all) cases, another (usually the fourth) palindrome is complementary to Palindrome I in sequence.

When the 3′ strand of the telomere sequences is folded into a secondary structure using an energy optimization algorithm, the most striking feature is the "rabbit ear" (or "Y-shaped") structure formed by the first four palindromes (with paired Palindromes I and IV; Fig. 1b). The general shape and loop sequences of this part resemble those at the 3′ end of autonomous parvoviruses, which are single-stranded DNA molecules flanked by stable duplex hairpins at both ends (Astell et al. 1979). Autonomous parvovirus-like loops are also found in other hairpins of the folded archetypal *Streptomyces* telomere.

Most of the hairpins are closed by the YGNAR sequence (loop underlined; N being most often C). In such hairpins, the G and A residues have been shown by ^{13}C NMR studies to form "sheared purine–purine pairing", resulting in a single C-residue loop (Zhu et al. 1995; Chou et al. 1997). Interestingly, GCA hairpins are the most stable among all GNA hairpins with a sheared G : A pair. These single-nt hairpin loops closed by sheared purine pairs resist single-strand specific nuclease attacks (Chou et al. 1997). The overwhelming presence of GCA loops at the 3′ overhangs strongly implies their importance in shaping the 3′ overhangs for structural integrity during replication, and possibly for the proper conformation for end patching.

◀ **Fig. 1** *Streptomyces* telomere DNA. **a** Aligned archetypal telomere sequences. For the telomeres of plasmids, the plasmid names are given, and for those of the chromosomes, the species names are given. When the left ("*l*") and right ("*r*") terminal sequences of a replicon are different, they are distinguished in *parentheses*. The degree of relative conservation is indicated by the *shading of the letters and the background*. The *lightest letters with the darkest background* indicate the most conserved sites, and the *darkest letters with the lightest background* indicate the least conserved sites. Six palindromes (I–VI) are indicated based on the *S. lividans* chromosome (and SLP2 right end) sequence. These palindromes are preserved in most but not all the archetypal telomeres. The binding regions based on the pSLA2 sequence of Tap (Bao and Cohen 2003) are marked. **b** Predicted secondary structures formed by the 3′ strands of the *S. lividans* chromosome. The *Roman numerals* indicate the palindrome numbers. The proposed Pu : Pu sheared pairings are indicated by the *solid dots*. The hairpin structures identical to those in the autonomous parvoviruses are *shaded*. **c** Secondary structures formed by the 3′ strands of a pseudotelomere (nt 3974–4234) found on SCP1. The long stem structure is omitted. **d** Secondary structures formed by the 3′ strands of the SCP1 telomere. Note the unpaired 3′ end and the 4-nt loops

2.2 Nonarchetypal Telomeres

Four nonarchetypal *Streptomyces* telomere sets have been characterized so far: three are on the plasmids, and one on a chromosome. The most striking set is that of SCP1. It differs remarkably from the archetypal telomeric sequences in primary structures and potential secondary structures formed by the 3′ overhangs (Kinashi et al. 1991; Bentley et al. 2004). It begins with a string of four to six Gs instead of three Cs in the archetypal telomeres. In the predicted secondary structure of the SCP1 telomere sequence, the 3′ end is unpaired, in contrast to the fully duplexed 3′ ends formed by many archetypal telomeres (Fig. 1d). Moreover, all the loops consist of four rather than three nucleotides at the archetypal telomeres. Most of the 4-nt loops may still form purine–purine sheared pairing between the first and fourth nucleotides. These distinct features hint at a very different capping TP for this plasmid (see below).

The telomere sequences on two linear plasmids, pRL1 and pRL2 (Zhang et al. 2006) and on the *S. griseus* chromosome (Goshi et al. 2002) also share no homology in primary and secondary structures with either the archetypal telomeres or the SCP1 telomere. Although multiple palindromes are present in these telomere sets, they are relatively sparse and simple compared to those in the archetypal telomeres. Unlike the SCP1 telomere, the predicted secondary structures of these telomeres contained 3-nt instead of 4-nt loops.

2.3 Terminal Inverted Repeats

The telomeres on a linear plasmid (or a chromosome) of *Streptomyces* are often part of longer terminal inverted repeats (TIRs), i.e., identical terminal sequences. The sizes of most TIRs range widely from tens of base pairs (bp) to hundreds of kilobases (kb). The sequences of the TIRs are also very heterogeneous with few constraints, except for the telomere sequences and perhaps a terminally located helicase gene *ttrA* that is highly conserved (Bey et al. 2000).

The size and sequence variations turn out to be the results of intermolecular exchanges that are frequent enough to be observed in the laboratory (reviewed by Chen et al. 2002). In general, TIR shortening may result from internal deletions or from recombination between heterologous linear replicons (with different terminal sequences). This is illustrated by the chromosomes of *S. coelicolor–S. lividans* hybrid strains produced by conjugation, which contain a telomere from each parent and exhibit imperfect TIRs of less than 200 bp (Wang et al. 1999). The 44-bp imperfect TIR of the SLP2 plasmid (Chen et al. 1993) and the 174-bp TIR of the *S. avermitilis* chromosome (Ikeda et al. 2003) were presumably generated in a similar fashion. More recently, Ya-

masaki and Kinashi (2004) discovered two chimeric chromosomes produced by recombination between SCP1 and the host chromosomes in *S. coelicolor* 2106, each of which contains a telomere of SCP1 and a telomere of the chromosome. An artificially constructed linear plasmid containing this mixed pair of telomeres can also replicate in vivo (Huang et al. 2007). These two telomeres share no homology down to the terminal nucleotide (C for *S. lividans* chromosome and G for SCP1), and therefore the recombinant plasmid and the 2106 chimeric chromosomes have no TIR. Thus, TIRs appear to have no significant biochemical function, although the duplication may serve as a template for repair of damage in the TIR (Qin and Cohen 2002), and so provide an evolutionary advantage.

On the other hand, TIR may also be increased by unequal crossovers between sister replicons, as was observed by Fischer et al. (1998) in *S. ambofaciens* and Uchida et al. (2003) in *S. griseus*. The unusually large (1-Mb) TIR found in some strains of *S. coelicolor* (Weaver et al. 2004) may also be the result of such an event.

2.4 Pseudotelomeres

"Pseudotelomeres" (Fig. 1c) were originally discovered as internal sequences of SLP2, which exhibit remarkable similarity to archetypal telomeres in both primary and secondary structures, such as the Y-shaped "rabbit ear" structures and the stems with 3-nt (often GCA) loops (Huang et al. 2003). Subsequently, two pairs of pseudotelomeres were also identified in the TIRs of SCP1 (Bentley et al. 2004). These pseudotelomeres lie near the end of both plasmid DNAs (at 1.6 kb on SLP2 and at 4.0 kb on SCP1). It is interesting that these pseudotelomeres all occur in pairs (and in the same sequence orientation) and are separated from each other by only 1.8–2.1 kb. They are presumably the remnants of integrated linear replicons. It is not known whether these "pseudotelomeres" are still functional in *Streptomyces* when placed at the end of linear replicons.

Recently, on a subclone of SLP2, Xu et al. (2006) discovered a 287-bp internal deletion corresponding to 91% of one of the predicted pseudotelomeres. Perhaps these pseudotelomeres readily form secondary structures under negative superhelicity in vivo, which are prone to either nucleolytic attacks or replication slippage. A more speculative alternative would be that these elements are mobile.

3
Terminal Proteins

3.1
Archetypal TPs

The first *Streptomyces* TPs characterized were found to be highly conserved in size (184–185 amino acids (aa)) and sequence (Fig. 2a) (Bao and Cohen 2001; Yang et al. 2002). No similarity to other known protein sequences in the databases (including the TPs of adenoviruses and ϕ29) was found. They are herein designated archetypal TPs to distinguish them from the heterologous TP that caps SCP1 (see below). Bao and Cohen designated the gene encoding archetypal TPs as *tpg* followed by a letter (and sometimes a number) to identify its origin. For example, the three TP genes for the linear plasmids pSLA2-L, pSLA2-M, and chromosome in *S. rochei* were named *tpgR1*, *tpgR2*, and *tpgR3,* respectively (Bao and Cohen 2001; Mochizuki et al. 2001; original assignments of Bao and Cohen corrected based on unpublished results communicated by H. Kinashi), and the TP genes for the *S. coelicolor* and *S. lividans* chromosomes were designated *tpgC* and *tpgL*. This genetic nomenclature is unconventional and is proving confusing as the catalog grows. A more conventional use of only "*tpg*" (gene) and "Tpg" (gene product) for generic designation is therefore used here. TP (terminal protein) is a general term applicable (as it has been elsewhere) to all those proteins that cap the DNA of linear plasmids and chromosomes, including those of adenoviruses, ϕ29 phage, and any other such replicons. In *Streptomyces*, TPs include Tpgs as well as others of heterologous origins (e.g., TP of SCP1—see below). To specify the origin of a *tpg* gene or Tpg protein, the plasmid name or a three-letter abbreviation of the species (as for restriction enzyme nomenclature) will be added in superscript. For example, the Tpgs of SLP2 and the *S. coelicolor* chromosome are designated Tpg^{SLP2} and Tpg^{Sco}, respectively.

The amino acid sequences of the Tpg homologs contain a putative helix-turn-helix domain that shows limited similarity to the DNA-binding "thumb" domain of HIV reverse transcriptase (Bao and Cohen 2001; Yang et al. 2002) and putative amphiphilic β-sheets (Fig. 2a). The latter may participate in protein–protein or protein–membrane interactions, as observed in adenoviruses and ϕ29 phage (Ortin et al. 1971; Schaack et al. 1990; Bravo and Salas 1997).

Most functional *tpg* form a polycistronic operon with an upstream gene *tap*, which is also essential for replication of the linear replicons in *Streptomyces* (Bao and Cohen 2003). In *S. coelicolor* and *S. lividans*, the *tap–tpg* operon is located about 100 and 120 kb, respectively, from the "right" end of the chromosome (based on the conventional orientation of the *S. coelicolor* chromosome). On the *S. avermitilis* chromosome, the tap^{Sav}–tpg^{Sav} operon lies about 8 kb from the right end—corresponding to the left end of the

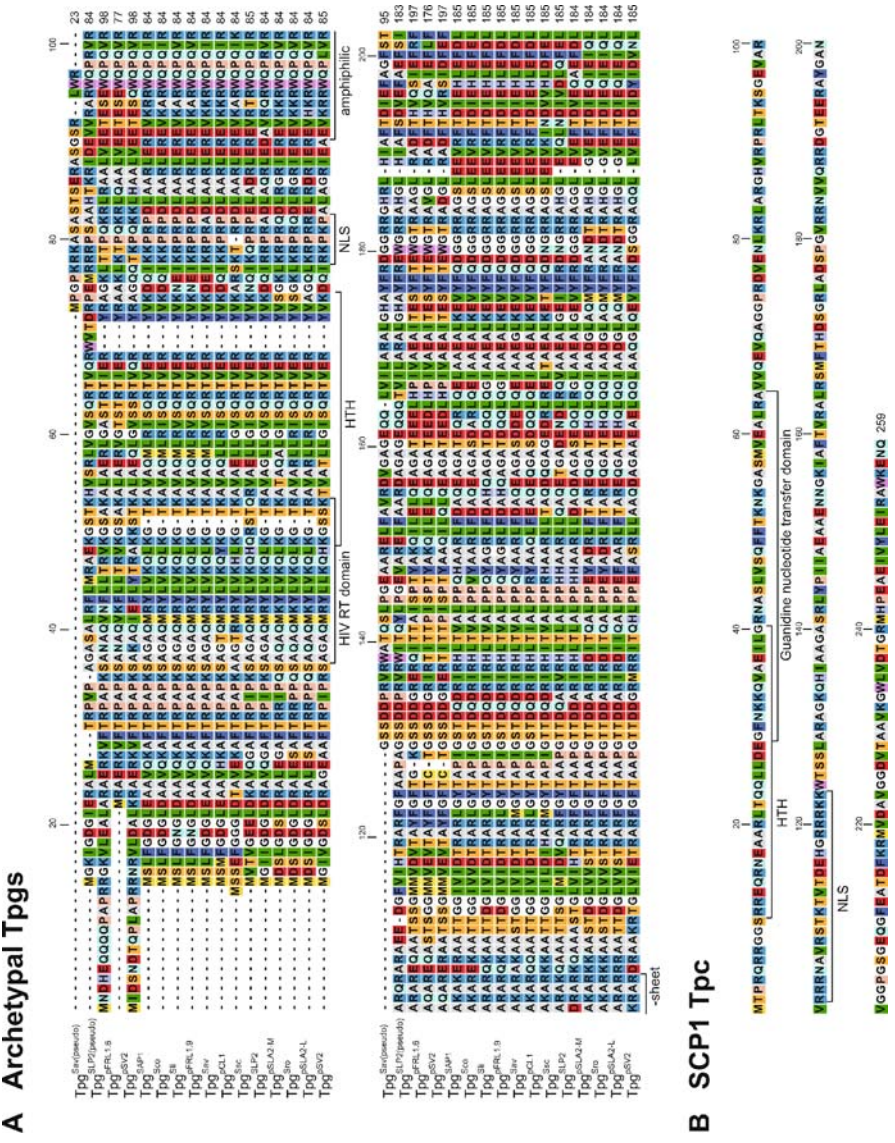


Fig. 2 Structure of TPs. **a** Aligned archetypal Tpg sequences. For a plasmid-encoded Tpg, the plasmid name is given, and for a chromosome-encoded Tpg, the three-letter abbreviation of the species is given (Sav, *S. avermitilis*; Sco, *S. coelicolor*; Sli, *S. lividans*; Sro, *S. rochei*; Ssc, *S. scabies*). pFRL1 encodes two Tpgs. These two Tpgs are distinguished by their gene numbers (6 and 9). The amino acid residues are *shaded* using the Rasmol color scheme. Four identified features (based on Tpg^{Sco} and Tpg^{Sli}) are indicated. HIV RT, human immunodeficiency virus reverse transcriptase; HTH, helix-turn-helix DNA-binding domain; NLS, nuclear localization signal. **b** Tpc of SCP1

S. coelicolor chromosome. Some linear *Streptomyces* plasmids also contain *tpg* homologs, which are sometimes pseudogenes. For example, SLP2 contains a unique tpg^{SLP2} gene on the left arm plus a *tpg* pseudogene on the 15.3-kb right arm, which is identical to the termini of the *S. lividans* chromosome. Smaller linear plasmids tend to lack a *tpg* homolog. They appear to be capped by the Tpgs encoded by larger plasmids or the host chromosome (Bao and Cohen 2001). Many *tpg* pseudogenes and at least one functional *tpg* gene (tpg^{SLP2}; C.-C. Yang, personal communication) are not accompanied by a *tap* gene.

Pseudogenes of *tpg* appear at unusually high frequencies in *Streptomyces* genomes. Their aberrant (usually shortened) size and the lack of a typical *Streptomyces* codon preference hint at their degeneracy. The first *tpg* pseudogene discovered was in the 15.3-kb terminal sequence shared by both arms (in TIR) of the *S. lividans* chromosome and the right arm of SLP2. This *tpg* pseudogene (encoding 183 aa) is about 8 kb from the telomere, and has been proven to be defective based on its inability to support replication of a recombinant linear plasmid (C.-C. Yang, unpublished results).

Putative *tpg* pseudogenes were also found in the chromosome of *S. avermitilis* (Ikeda et al. 2003). Their amino acid sequence is more divergent than that of the pseudogene in *S. lividans*, although their lack of function has not been confirmed. The terminal location of the *tpg* pseudogenes suggests that they are remnants of previous exchanges with other linear plasmids or chromosomes. Pseudogenes in the genomes of free-living bacteria are relatively rare, presumably because of their rapid degeneration under evolutionary selection. The finding of the *tpg* pseudogenes supports the occurrence of relatively recent end exchange events. The wide distributions of *tpg* genes and pseudogenes probably reflect the high mobility and frequent intermolecular exchanges of linear plasmids.

3.2 Nonarchetypal TPs: The TP of SCP1

The nonarchetypal telomere sequence of SCP1 and the absence of a *tpg* homolog on it suggested that it may be capped by a different type of TP. Recently the TP that caps SCP1 was isolated and identified. It is encoded by a gene (SCP1.127, designated *tpc*) on SCP1 (Huang et al. 2007). The gene product (Tpc; 259 aa) is significantly longer than the archetypal TPs (Fig. 2b), and does not resemble any protein in the databases. It contains an HTH DNA-binding domain in the N-terminal region, followed by a so-called guanine nucleotide transfer domain. These motifs are presumably involved in anchoring to the telomeric DNA with the guanine nucleotide transfer domain providing specificity for the terminal guanidine residues at the 5′ end. A gene *tac* (SCP1.125) upstream of *tpc*, encoding a hypothetical protein, is also essential for replication of SCP1. Therefore, the SCP1 telomeric DNA and *tac–tpc*

set represent a heterologous system that has evolved convergently to form a novel *Streptomyces* telomere system.

Plasmids pRL1 and pRL2 also contain nonarchetypal telomeres, and therefore may be capped by novel TPs. pRL2 nevertheless encodes a Tpg homolog (Tpg^{pRL2}), which is significantly larger than archetypal Tpgs (216 vs 184–185 aa). Tpg^{pRL2} could support replication of an artificially constructed linear plasmid with pRL2 telomeres in the absence of *tpg*, suggesting that it capped the pRL2 telomeres (Z. Qin, personal communication). However, it is not clear whether in vivo pRL2 is also capped by Tpg^{pRL2} or by another TP. In contrast, pRL1 contains no homolog of *tpg* or *tap*.

3.3 TP–TP Interactions

Interaction between the TPs at the two telomeres was initially suggested based on the immobilization of the TP–DNA complex during electrophoresis without a denaturing or proteolytic treatment (for example, Lin et al. 1993). TP–TP interaction (between two Tpg^{Sco}) at the telomeres in vivo was recently demonstrated by chemical cross-linking experiments and atomic force microscopy by Tsai et al. (unpublished results). If such a TP–TP interaction persists during recombination, an even number of crossovers would be required to resolve two recombining chromosomes, and this would automatically lead to circular genetic maps (Stahl and Steinberg 1964; Stahl 1967). Indeed, the genetic maps obtained from various *Streptomyces* species have always been circular (Hopwood 1967) despite the linearity of their chromosomes (Lin et al. 1993). This model was further supported by the observed strong bias toward an even number of crossovers during recombination of *Streptomyces* chromosomes (Wang et al. 1999). For a more comprehensive explanation, see the recent review by Hopwood (2006).

TP–TP interactions have previously been demonstrated in other TP-capped genomes, including ϕ29 phage (Ortin et al. 1971), GA-1 phage (Arnberg and Arwert 1976), and adenoviruses (Robinson et al. 1973). In the absence of proteolytic treatment, these linear genomes assume a circular form (Arnberg and Arwert 1976; Keegstra et al. 1977; Wong and Hsu 1989). The identical functions of these heterologous TPs in diverse biological systems (bacterial chromosomes, plasmids and phages, and animal viruses) indicate the importance of this function.

Association of telomeres in vivo makes sense in that it provides topological constraints that make superhelicity in the linear DNA possible, which would be important for replication, transcription, and/or recombination. Indeed, superhelical structures of adenovirus DNA were detected in the virus particles (Wong and Hsu 1989). Persistent telomere interaction, on the other hand, creates a postreplicational segregation complication that is also generally faced by circular replicons (see below).

4
Replication

4.1
Replication Origin and Replication Initiation

Unlike ϕ29 phage and adenoviruses, replication of the linear plasmids (and linear chromosomes) of *Streptomyces* is initiated from an internal origin. Such an internal replication origin was first identified by Shiffman and Cohen (1992) in pSCL1 of *S. clavuligerus* by its ability to support replication of a circular DNA construct in *Streptomyces*. This autonomously replicating sequence (ARS) contains two *rep* genes essential for replication. One is a DnaB (replicative helicase) homolog and the other resembles another *rep* gene of pSLA2 of *S. rochei* (Chang et al. 1996).

Subsequently, ARSs were identified in several other linear plasmids of *Streptomyces*. Without exception, they contain at least one gene encoding a protein necessary for replication (presumably its initiation) and probably the replication origin in the form of a series of direct repeats (iterons) or AT-rich sequences or both. The ARS of pSLA2 contains an operon consisting of $rep1^{pSLA2}$ encoding a DNA-binding protein and $rep2^{pSLA2}$ encoding a helicase (Chang et al. 1996). The ARS of SCP1 is central and contains an open reading frame (ORF) (SCP1.196) encoding a 506-aa putative primase/helicase, an overlapping ORF (SCP1.197) encoding a 133-aa hypothetical protein, and a putative iteron nearby (Redenbach et al. 1999). The same gene pair encoding a primase/helicase and a ca. 130-aa protein is also found in two ARSs (*ori2* and *ori3*) on the 210-kb linear plasmid pSLA2-L in *S. rochei* (Mochizuki et al. 2001). The ca. 330-aa proteins are of limited occurrence, but the primase/helicase proteins are widespread, particularly among plasmids and bacteriophages.

pSLA2-L contains a third ARS (*ori1*) adjacent to *ori2*. *ori1* includes an essential gene, *repL1*, a subsidiary gene, *repL2*, and a putative replication origin lying about 800 bp from *repL1* (Hiratsu et al. 2000). Search of the database found only a homologous pair of *repL1* and *repL2* on pSCL1 in a syntenous arrangement (including a 4-bp overlap). It is not known which of the three ARSs (all of which can support replication of plasmids in a circular form) is functional in vivo for pSLA2-L replication. In view of the great difference between pSLA2-L and pSCL1 in size and genetic content, it is likely that *ori1* was acquired during evolution. However, the joining of *ori1* and *ori2* next to each other is puzzling.

The SLP2 ARS represents yet a different type. It contains an iteron and a *dnaB*-type helicase gene (rep^{SLP2}) (Xu et al. 2006), which is found in a wide variety of bacterial plasmids, but this is the only example among linear *Streptomyces* plasmids characterized so far.

The diversity of the genetic constituents (*oriP* and *rep* enzymes) involved in initiation of replication of *Streptomyces* linear plasmids is in contrast to the greater homogeneity of those (telomeres and Tpg/Tap proteins) involved in end patching (at least of the archetypal telomeres). This indicates that there have been diverse origins of the *rep*–iteron system of the linear *Streptomyces* plasmids. A recent study shows that new *rep*–iteron combinations produced by artificial exchanges of *rep* genes or iterons among three plasmids, SLP2, SCP1, and pSLA2, are competent to support plasmid replication, although the plasmid transformation efficiencies were decreased by two to three orders of magnitude (Xu et al. 2006). This indicates a relatively loose Rep–iteron specificity for these replication initiation systems.

4.2 Elements Required for Replication in Linear Form

Qin et al. (2003) discovered a novel gene $rlrA^{pSLA2}$ on pSLA2, which is required for replication of the plasmid in linear but not circular form. $RlrA^{pSLA2}$ is a 308-aa protein with two LuxR family regulatory domains, which on overexpression interferes with the replication of pSLA2 in circular form. On pSLA2, *rlrA* appears to be regulated by a divergently transcribed gene *rorR* encoding a KorA-like regulator.

No *rlrA* homolog has been found in the database except for a weak homology (27% identity) with a hypothetical protein encoded by pSCL1 (ORF pSCL_p5), suggesting that other linear plasmids may encode their own functional counterparts of $rlrA^{pSLA2}$. Indeed, replication of SLP2 in linear form does require an additional sequence containing an ORF (SLP2.13, $ilrA^{SLP2}$, with no homology to $rlrA^{pSLA2}$) adjacent to a KorA-like regulatory gene (SLP2.14) on SLP2 (Xu et al. 2006). Interestingly, this requirement for linear replication may be substituted by either $rlrA^{pSLA2}$ (Qin et al. 2003) or by a truncated *tap* homolog on SLP2 ($mtap^{SLP2}$) that encodes an 88-aa polypeptide corresponding to the N terminus of Tap proteins (Xu et al. 2006). In a yeast two-hybrid assay, $IlrA^{SLP2}$ interacts with Tap^{Sli}. Therefore, it has been suggested that $IlrA^{SLP2}$ and $RlrA^{pSLA2}$ are involved in coordinating initiation from the origin with end patching during replication of linear plasmids. However, in these studies, no *trans*-acting function of these genetic elements has been tested, and therefore the possibility that $rlrA^{pSLA2}$, $ilrA^{SLP2}$, and $mtap^{SLP2}$ contain an essential *cis*-acting element, such as an origin of replication, which is essential specifically for replication in a linear form, cannot be ruled out.

Like their circular counterparts, the linear plasmids of *Streptomyces* generally do not encode a DNA polymerase, and presumably rely on one of the chromosomally encoded polymerases (such as Pol III; see below) for replication. An exception so far has been SCP1, which encodes α (catalytic) and β (clamp) subunits of Pol III (Bentley et al. 2004), the significance of which is not clear.

5
End Patching

5.1
The Mechanisms

The internally initiated bidirectional replication of linear replicons of *Streptomyces* encounters a classical problem of dealing with the gaps at the 3′ ends of the replicons, which cannot be replicated by the replisomes. Chang et al. (1994) determined the size of the 3′ overhangs during replication of pSLA2 plasmid to be about 280 nt. The sizes of the 3′ overhangs of other linear *Streptomyces* plasmids and chromosomes have been found to be similar and range from 250 to 320 nt (C.-H. Huang, unpublished results).

Unlike the telomeres of eukaryotic chromosomes, which are heterogeneous in size because of variable extension by telomerase, the telomeres of *Streptomyces* linear replicons have fixed sequences. Therefore, the 3′ overhangs of *Streptomyces* linear replicons must be patched during replication. Several biochemical mechanisms for end patching were originally proposed (Chen 1996). These models center on the role of the TP, which must end up attached covalently to the first nucleotide at the 5′ end. There are two possible mechanisms that may account for this result. Firstly, TP may nick a double-stranded intermediate and remain covalently linked to the 5′ end at the cut site, similarly to the action of many nickases, topoisomerases, and recombinases. Secondly, TP may act as a primer for DNA synthesis that fills the 3′ gaps. Such a primer function is seen with the TPs of ϕ29, adenoviruses, and other related linear replicons. In this regard, the amino acid sequences of TP do not resemble any known nickase or primer TP in other systems.

TP as a nickase has no experimental support. No nuclease domain is found in the *Streptomyces* TP sequences, and no DNA nicking activity has been detected with any *Streptomyces* TP. The simplest model involving TP as a nickase is analogous to the "rolling hairpin" model for replication of helper-dependent parvoviruses (Kornberg and Baker 1992), in which the terminal palindrome folds back and provides a primer for the end patching, followed by nicking by TP and fill-in synthesis that restores the telomere. This model predicts that each round of replication will flip the terminal palindrome once, and therefore the terminal palindrome would exist in two opposite orientations among the population. Testing this flip-flop model with an artificially constructed pSLA2 derivative, Qin and Cohen (1998) ruled out this model.

The complicated "modified rolling hairpin model" (Huang et al. 1998) involving TP as a nickase is based on the asymmetric resolution model for autonomous parvovirus replication, which avoids flip-flops of the terminal palindromes (Astell et al. 1985). This model is also disfavored by the failure to detect the proposed intermediates (Huang 2002).

Currently, TP-primed DNA synthesis models are favored and there are two versions (Fig. 3). In the straightforward version, TP-primed synthesis takes place at the 3′ end in the secondary structural complex formed by the single-strand overhang (Fig. 3a). A second model involves two steps: DNA synthesis followed by recombination. Firstly, a TP-primed synthesis takes place at the fully duplex telomere resulting from leading strand synthesis, which displaces the 5′ strand in the process. The displaced 5′ parental strand with the attached TP then anneals with the exposed 3′ overhang, forming a half Holliday junction, which is subsequently resolved by a resolution system (e.g., the RuvABC resolvasome; Fig. 3b).

Two interesting features distinguish the recombinational model from the straightforward model. Firstly, the end result of the recombinational model is a conservatively replicated telomere DNA (Fig. 3b). Secondly, the process depends on a homologous recombination system. If such recombination were general and *recA*-dependent, then *recA* would be essential for growth of wild-type strains of *Streptomyces* with linear chromosomes. Several laboratories found *recA* to be refractory to complete deletion in a number of different species, suggesting an essential role for *recA* in the viability of *Streptomyces* (Muth et al. 1997). Recently, however, Huang and Chen (2006) were able to isolate null *recA* mutants in different strains of *S. coelicolor*, and the chromosomes of these mutants remained linear. This suggests that *recA* is not

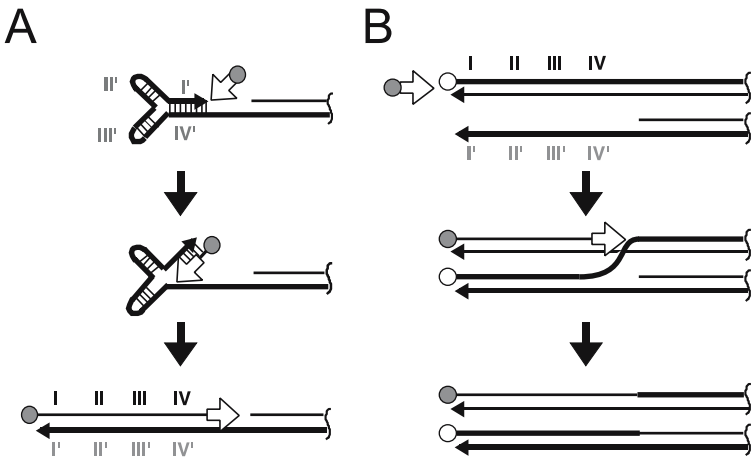


Fig. 3 Two models of end patching involving TP-primed DNA synthesis. **a** TP-primed synthesis on the 3′ single-strand overhangs. **b** TP-primed synthesis on a double-strand telomere followed by homologous recombination. The parental DNA strands are represented by *thick lines*, and the daughter strands by *thin lines*. The TP on the parental DNA is depicted by *open circles*, and the new TP by *filled circles*. The *open arrows* represent the DNA polymerase that catalyzes the TP-primed synthesis. The *Roman numerals* indicate the first four palindromes at the telomeres. See text for detailed descriptions

involved in the end patching of *Streptomyces* chromosome (and thus also of linear plasmids). While all the *recA* mutants exhibited an expected increase in sensitivity to UV and mitomycin C and blocks in chromosomal recombination during conjugation, one pair of mutants displayed *recA*-independent recombination during conjugation, even when both parents were *recA* mutants. It still remains possible that the *recA*-independent homologous recombination is involved in end patching.

The recombination model was rejected by Qin and Cohen (1998) based on a lack of homogenotization between the two nonidentical terminal sequences of a linear plasmid, which is expected if recombination occurs between the two telomeres in each replication cycle. However, this study concerns only recombination between the telomeres on opposite ends, and therefore has not ruled out recombination between telomeres at the same end of the sister molecules, as proposed in the original model. Therefore, the recombination model is still valid as a working hypothesis.

5.2 Protein Players

Regardless of whether homologous recombination is involved or not, TP-primed synthesis is the key step in both of the end patching working models. The key players in this reaction are the TP (Tpg or Tpc), Tap (or a functional homolog for the SCP1 system), a DNA polymerase, and the DNA template.

In vitro deoxynucleotidylation of Tpg^{Sco} has been established recently using crude extracts of *Streptomyces* harboring a high-copy-number linear plasmid (providing necessary enzymes and template) and *Escherichia coli* (expressing Tpg^{Sco}). In this system, dCMP is specifically incorporated onto a Thr residue present in Tpg^{Sco} (Yang et al. 2006). This differs from the situation of ϕ29 and adenoviruses, in which deoxynucleotidylation occurs at a Tyr residue. A phospho-Thr linkage is alkali labile. The TP that caps *Streptomyces* linear replicons is readily removed at alkaline pH (Lin et al. 1993).

Both Tpg^{Sli} and Tap^{Sli} bind to DNA in vitro, as expected from their sequences (Bao and Cohen 2003). Tpg^{Sli} exhibits no sequence specificity, whereas Tap^{Sli} interacts strongly with single-strand telomeric overhangs at regions corresponding to Palindromes II and III. Immunoprecipitation and yeast two-hybrid studies indicate that Tap^{Sli} interacts with Tpg^{Sli} in vivo. This leads to a model that Tap positions at a specific region of the 3′ overhang and recruits Tpg to the telomere for end patching (Bao and Cohen 2003).

Using Tap^{Sco} as a scaffold, Bao and Cohen (2004) isolated several telomere-binding proteins from *S. lividans*, including DNA polymerase I (Pol I) and topoisomerase I. Interestingly, they discovered reverse transcriptase activities in both proteins. In view of the fact that the DNA-binding domain in the Tpg proteins also resembles the thumb domain of the reverse transcriptase of HIV,

this discovery is intriguing. However, there is no simple model for the involvement of reverse transcriptase activity in end patching, which would require a RNA version of the telomere sequence for the template. The biological significance of the reverse transcriptase activities remains obscure, and a simple model of end patching involving a reverse transcription step remains to be envisioned.

5.3 DNA Templates

The template requirement for end patching of pSLA2 was examined by Qin and Cohen (1998). The minimum length of the telomere necessary for end patching in vivo was determined to be 144 bp, which spans the first six palindromes. However, not all the 144 bp are required. For example, Palindrome IV, but not Palindromes II, III, and V, may be removed without affecting viability.

Using a recombinant plasmid to study the *S. lividans* telomere sequence that is shared by the right end of SLP2 and the *S. lividans* chromosome, H.-H. Tsai (unpublished results) found that Palindromes IV to VII may be deleted without affecting replication of the linear plasmid. The lack of Palindrome IV makes the pairing of Palindromes I and IV impossible (Fig. 1b) and, in the predicted secondary structure of this shortened sequence, Palindrome I instead forms a stem-loop by itself. This result indicates that the pairing of Palindromes I and IV in the predicted optimal secondary structure is not essential. In this regard, the 3′ end of the SCP1 telomere is single-stranded in the predicted secondary structure (Fig. 1c). The minimum length requirement for the SCP1 telomere determined so far is only 84 bp, spanning the first three palindromes (C.-H. Huang, unpublished results).

Shiffman and Cohen (1992) showed that functional linear plasmids can be produced when a replication-proficient linear plasmid sequence containing telomeres bracketed by adventitious DNA is introduced into *Streptomyces*. The technique has since been widely used to construct recombinant linear plasmids. The success of this procedure indicates that in vivo end patching may be achieved on an internal telomere DNA sequence that is not capped by a TP. This suggests that the presence of TP caps on the parental DNA during replication is not a critical signal, but the telomere sequence alone (even if it is internal) is sufficient to be recognized as the template for end patching.

5.4 DNA Polymerase

It is not clear which cellular DNA polymerase is involved in end patching. The sequenced chromosomes of *Streptomyces* contain a *polA* gene for DNA poly-

merase I (Pol I). Bao et al. (2004) found that Pol I was associated with the *Streptomyces* telomere complex, and *polA* could not be deleted from either a linear or circular chromosome without affecting viability. However, more recently, complete deletion of *polA* was achieved in certain strains of *S. coelicolor* (Huang and Chen, manuscript in preparation), and cell extract prepared from the *polA* null mutant remained competent in supporting in vitro deoxynucleotidylation of Tpg^{Sco} (Yang et al. 2006). These results suggest that either Pol I is not involved in end patching or it may be substituted by another polymerase. A close homolog of *polA* (SCO3434) is present in the *S. coelicolor* chromosome. However, a chromosome with double knockout mutations in *polA* and SCO3434 remains linear. This rules out the possibility that SCO3434 may substitute for *polA* in end patching. In *E. coli*, DNA polymerase II (Pol II, encoded by *polB*) can substitute for Pol I when performing fill-in synthesis during replication and repair. However, no *polB* homolog is present in any sequenced *Streptomyces* genome.

Flett et al. (1992; 1999) showed that replication of the *Streptomyces* chromosomes requires a DnaE-type Pol III enzyme (*dnaE1*). Interestingly, duplicate copies of genes encoding some subunits of Pol III are present in all sequenced *Streptomyces* genomes. There are two genes for the catalytic subunit α (*dnaE1-2*) and clamp subunit β (*dnaN1-2*), and three genes for the $3' \rightarrow 5'$ exonuclease subunit ε (*dnaQ1-3*). Similar duplication of Pol III subunit genes is also seen in most other actinomycetes with circular or linear chromosomes. It is not clear whether one of the Pol III holoenzymes is involved in end patching. Genes encoding a *polC*-type α subunit, present in low G+C gram-positive bacteria, have not been found in *Streptomyces* genomes. Of the error-prone translesion repair DNA polymerases, Pol IV-encoding *dinB* is present in two copies, but Pol V-encoding *umuC* is absent.

6
Postreplicational Segregation

6.1
The Plasmid-Encoded *parAB–parS* System

Low-copy-number linear plasmids of *Streptomyces* contain a pair of partitioning genes, *parA* and *parB*, which are involved in active segregation and thus stable inheritance of these plasmids. During segregation, ParB binds to a specific palindromic DNA sequence (*parS*) on the replicons and forms a partitioning complex, while ParA, an ATPase, forms cytoskeleton-like structures and is responsible for segregating the two daughter replicons (for review, see Gerdes et al. 2000; Bignell and Thomas 2001; Shih and Rothfield 2006). The vegetative mycelium of *Streptomyces* contains few septa with each compart-

ment containing many copies of chromosomes, and therefore precise partitioning of the genomes appears not to be critical during vegetative growth. Proper segregation and partitioning, on the other hand, is important during sporulation, when multiple septa are formed in the aerial hyphae to produce unigenome spores. Limited studies on partitioning of linear plasmids have been done on SCP1 and SLP2.

Interestingly, SCP1 contains two partitioning loci, each consisting of a *parA–parB* operon (Bentley et al. 2004). The protein pairs encoded by the two operons are more than 50% identical to each other (and more than 30% identical to the chromosomally encoded ParAB^{Sco}), and both pairs are imperfectly related to the type Ia partitioning system (Gerdes et al. 2000). The presence of two partitioning systems is unusual, and may reflect the mosaic nature of this plasmid as a result of past recombination between heterologous plasmids.

The partitioning function encoded by SCP1 has been demonstrated by the ability of the integrated SCP1^{NF} (Chater and Kinashi, in this volume, for details on SCP1^{NF}) to suppress partially the chromosome partitioning defect during sporulation that is caused by a chromosomal Δ*parAB*^{Sco} mutation (Bentley et al. 2004). No *parS*^{Sco} is present in SCP1, and presumably ParAB^{SCP1} acts on different sequences. Therefore, one may assume that the partitioning activity restored to the Δ*parAB* host is exerted on an as yet unidentified *parS*^{SCP1} element present in the integrated SCP1^{NF}.

SLP2 of *S. lividans* also contains a type Ia-like locus in the central region. In a typical type Ia locus, *parS* is adjacent to and downstream from the *parAB* operon. *parS*^{SLP2} (two short inverted repeats), however, lies within the coding sequence of *parB* and encodes the same amino acid sequences (C.-C. Hsu, unpublished results). The binding of ParB to its own coding sequence raises the possibility that it may regulate its own synthesis by blocking transcription elongation.

Deletion of *parAB*^{SLP2} resulted in loss of SLP2 from spores of *S. coelicolor*, and the loss was more severe in an *S. coelicolor* mutant with a chromosomal Δ*parAB*^{Sco} mutation, indicating that the chromosomal partitioning system may affect partitioning of SLP2, although *parS*^{Sco} is likewise absent from SLP2. The interesting cooperative interactions between the chromosomal and plasmid partitioning systems await further elucidation.

On linear plasmid SAP1 (94 kb) of *S. avermitilis* (Ikeda et al. 2003), *parA*^{SAP1} exists without an accompanying *parB* gene. Such unpaired *parA* genes are also found in some linear plasmids in other actinomycete hosts, e.g., pCLP in *Mycobacterium celatum* (Accession number: AF312688), pRHL3 in *Rhodococcus* sp. RHA1 (http://www.rhodococcus.ca), pBD2 (NC_005073) in *Rhodococcus erythropolis* BD2, and pREL1 in *R. erythropolis* PR4 (Sekine et al. 2006). The biological significance of these lone *parA* genes on linear plasmids is not clear.

6.2 Resolution of Sister Plasmids

The apparent interaction between the TPs at the telomeres of a linear *Streptomyces* replicon poses three potential complications during segregation. During and after replication, if the parental TPs persist in associating with each other, the 5′ ends of the two parental DNA would also remain linked with each other, thus creating a novel Möbius strip-like situation, i.e., the two daughter DNA molecules cannot be separated from each other (Fig. 4A–D). To resolve this Möbius strip-like structure, the two parental TPs perhaps dissociate from each other and each finds a new TP partner (Fig. 4E). In this case, a mechanism must exist to allow the recognition and sorting of the old and new TPs. Alternatively, an exchange may occur between the daughter DNA molecules to resolve the structure (Fig. 4F). Such a solution is similar to the resolution of dimeric circular replicons by the XerCD recombinase (reviewed by Barre et al. 2001). Is there such an analogous site-specific recombination system in *Streptomyces*, or is the resolution carried out by less efficient homologous recombination? It is noteworthy that in the latter scenario, replication of the *Streptomyces* linear replicons would be semi-conservative with respect to DNA but conservative with respect to TPs.

A second complication may arise from homologous recombination. Between two *circular* sister chromosomes or plasmids, homologous recombination generates dimers that must be resolved to achieve segregation. The resolution is carried out by the site-specific recombinase XerCD on a specific sequence—*dif* on chromosomes and *cer* on plasmids. Linear DNA molecules are not expected to dimerize through homologous recombination. However, a persistent TP–TP interaction during and after replication may also result in pseudodimer formation through homologous recombination. Interestingly, the recombination products in fact would be identical to the replication products from a persistent TP–TP association (Fig. 4D), which thus may be resolved by the two alternatives outlined above: TP exchange or DNA exchange. No homologs of *xerCD* genes are found in the *S. coelicolor* genomes. Perhaps dimer resolution is solved by a different recombination system, or by TP exchange.

Another topological problem encountered during the replication of circular DNA is the formation of concatemers between two intertwined sister DNA molecules. In *E. coli*, an essential topoisomerase Topo IV (encoded by *parC* and *parD*) is responsible for decatenation of such concatemers (Adams et al. 1992; Zechiedrich and Cozzarelli 1995). *Streptomyces* genomes contain a pair of *parC* and *parD* homologs (SCO5822 and SCO5836 in *S. coelicolor*; SAV2422 and SAV2423 in *S. avermitilis*), presumably encoding Topo IV. Attempts to knock out these genes in *S. coelicolor* have not been successful (T.-W. Huang, unpublished results), suggesting that Topo IV is also essential for decatenation of the linear chromosomes (and presumably linear plasmids)

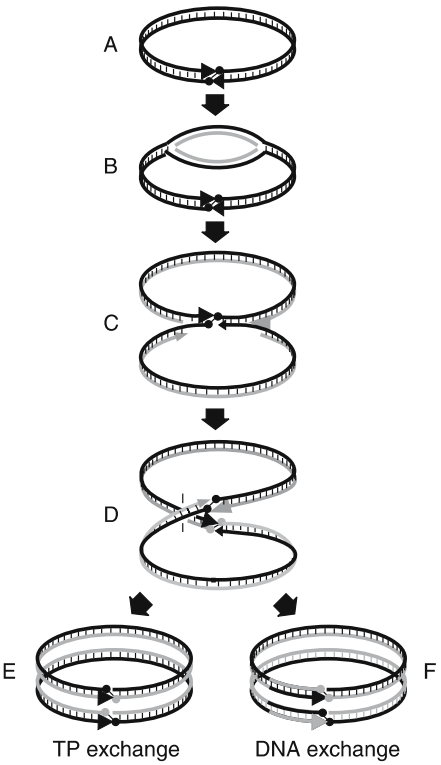


Fig. 4 Postreplicational Möbius strip-like problem. *A* A linear replicon assumes a circular configuration caused by interactions between the terminal proteins. The DNA duplex is colored in *solid black*, and the TPs are represented by the *filled black circles*. The association between them is emphasized by the connecting line. *B* Replication begins at an internal origin. The newly synthesized strands are colored *gray*. *C* Bidirectional replication reaches the telomeres. *D* When end patching is complete, if the parental TPs remain associated to each other, the two sister replicons form a Möbius strip-like structure with 5′–5′ linkages. The Möbius strip-like structure may be resolved by means of two possible alternative pathways. *E* In one pathway, the parental TPs dissociate from each other and find proper "daughter" TPs (*filled grey circles*) to pair with. *F* In the other pathway, a recombination (at sites indicated by the vertical lines in *D*) resolves the pseudodimer

of *Streptomyces* after replication. This may reflect the circular configuration formed by the linear *Streptomyces* replicons through persistent TP–TP interactions. Alternatively, a decatenase such as Topo IV is required to untangle the large chromosomes after replication irrespective of their topology, as is the case with eukaryotic chromosomes.

7
Conjugal Transfer

7.1
Mode of Transfer

In addition to cell cycle-associated replication, a linear plasmid must also be replicated during conjugal transfer to deliver a new copy to the recipient mycelium. Most linear plasmids in the *Streptomyces* investigated are conjugative, being able to promote conjugation, self-transfer, and chromosome mobilization.

In typical circular plasmids of other bacteria, replication during conjugal transfer differs from the vegetative replication in initiation and mode of synthesis. Conjugal transfer is initiated by nicking of one strand of the plasmid DNA at a specific site, *oriT*, by a nickase (TraI in *E. coli*), which remains covalently attached to the 5′ end of the nicked strand. This starts a rolling circle replication around the whole plasmid, and the displaced strand is led into the recipient cell by the nickase.

Such unidirectional replication scheme, if initiated at an internal *oriT* of a linear plasmid, would proceed unidirectionally and replicate only part of the DNA molecule. To replicate the complete plasmid for transfer, internally initiated replication must proceed bidirectionally (as in vegetative replication). Thus, linear plasmid DNA transferred by such a mechanism would need to be in a double-stranded form to maintain continuity. There is no evidence for the double-strand transfer for *linear Streptomyces* plasmids, but there is experimental support for such double-strand transfer of *circular* plasmids in *Streptomyces* (Possoz et al. 2001; Hopwood 2006; Reuther et al. 2006).

Alternatively, linear *Streptomyces* plasmids may be transferred from the telomeres as proposed by Chen (1996). In this "end first" model (Fig. 5), transfer of linear DNA is initiated at a telomere by a TP-primed DNA synthesis, which continues through the whole replicon and displaces the nontemplate strand from the TP-capped 5′ end. As in the rolling circle model, the displaced strand is led by a protein covalently linked to the 5′ end (TP) into the recipient. This conjugal replication differs from vegetative replication not only in the site of initiation, but possibly also in the substrate used by the TP-primed synthesis involved. TP-primed synthesis in the straightforward model of end patching acts on single-stranded DNA, whereas that in the proposed conjugal transfer mechanism acts on double-stranded DNA (with a TP cap), as in the case of ϕ29 and the adenoviruses. The latter would require a helicase for unwinding. In most linear *Streptomyces* plasmids, a conserved helicase gene (*ttrA*) is present near one or both ends. Deletion of *ttrA* reduces the efficiency of transfer of the plasmid, but does not affect replication of linear plasmids or chromosomes (Huang et al. 2003). *ttrA* homologs are also found at the ends of all *Streptomyces* chromosomes that have been sequenced. This,

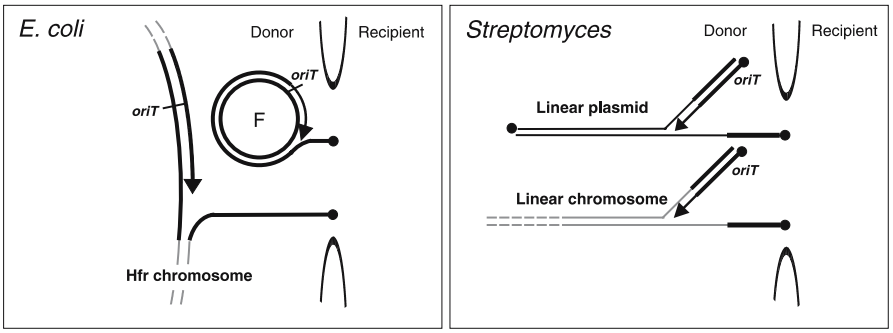


Fig. 5 "End first" model of conjugal transfer of linear plasmids and chromosomes of *Streptomyces* compared to the rolling circle model of conjugal transfer of circular replicons. *Left*: In *E. coli*, F plasmid-specific proteins carry out a nicking–priming event at *oriT* on F DNA, where the 5′ end is covalently attached to the nickase (TraI), and a rolling circle replication is initiated at the 3′ end. The TraI-capped 5′ strand is continuously displaced and transferred to the recipient. In Hfr cells, the chromosomal DNA is led by the F sequence into the recipient cell as an exogenote. *Right*: For linear plasmids and linear chromosomes of *Streptomyces*, the TP-capped telomere may act as the origin of transfer (*oriT*), where TP-primed replication is initiated, and the TP-capped 5′ strand is displaced and transferred, as for F/Hfr transfer

so far, is the only experimental support for the end-first model of transfer. As in the case of vegetative replication, it is not clear which DNA polymerase(s) is involved in replication during conjugal transfer.

7.2 Nuclear Localization

Both Tpgs and Tpc contain a nuclear localization signal (NLS) or NLS-like sequence. A monopartite NLS was initially predicted in the TP of *S. coelicolor* (Tpg^{Sco}) and *S. lividans* (Tpg^{Sli}) (Yang et al. 2002). Similar motifs have been found at the same locations in many other Tpgs (Fig. 2a). A bipartite NLS is also found on Tpc (Fig. 2b). The finding of two different types of NLSs in two convergently evolved TP families raised the likelihood that their occurrences are not fortuitous, although *Streptomyces*, being prokaryotic, lacks nuclei. A recent study (H.-H. Tsai et al., unpublished results) demonstrated that these NLSs can function in targeting the TPs (Tpg^{Sco}, Tpg^{Sav}, and Tpc) and the attached DNA into the nuclei of human cells (Fig. 6). The NLS alone is sufficient to perform nuclear targeting. On the other hand, the NLSs are dispensable for replication of the linear plasmids, suggesting that the nuclear localization function in these TPs has evolved independently of their replication function.

The ability of the TPs to carry the attached DNA into eukaryotic nuclei is analogous to the interkingdom transfer of T-DNA from *Agrobacterium tumefaciens* to plant nuclei by its capping protein (VirD), which also possess NLSs

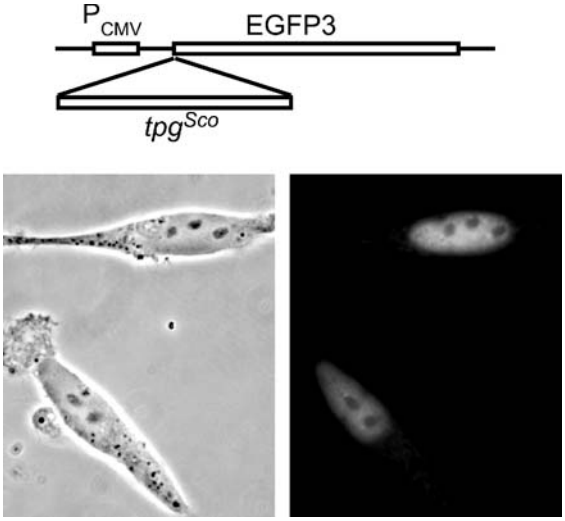


Fig. 6 Nuclear targeting function of TPs. *tpg*^{Sco} was translationally fused to the N-terminus of a reporter gene, EGFP3 (encoding a triple green fluorescence protein concatamer), under the control of the CMV immediate early promoter (P_{CMV}) on pEGFP3 vector (*top*), and used to transfect human HeLa cells. In the transfected cells, green fluorescence accumulated in the nuclei (*bottom*). On the *left* is a phase-contrast microscopic image, and on the *right* is a fluorescence microscopic image of the same transformed cells. In the cells transfected by pEGFP3 vector, the fluorescence was present mainly in the cytosol

(for review, see Zambryski et al. 1989). This raises the possibility that TP-mediated DNA transfer may also take place in nature, between *Streptomyces* and, for example, plants in soil. This possibility, which has great evolutionary and ecological significance, remains to be investigated.

8 Concluding Remarks

The *Streptomyces* TPs share no homology with those of ϕ29 and adenoviruses, indicating a convergent evolution of TP-primed synthesis across various systems. Their small size (184–185 aa for Tpgs and 259 aa for Tpc) are amazing, considering the diversity of possible functions they may possess including: (1) priming DNA synthesis during end patching (and perhaps replication during conjugal transfer); (2) protecting against exonucleolytic attacks; (3) providing topological constraints through TP–TP interactions to achieve superhelicity; and (4) nuclear targeting in eukaryotic cells.

As expected from their frequent exchanges, the linear plasmids of *Streptomyces* have the same telomere structures as their host chromosomes. The plasmid telomere sequences exhibit wider divergence (Huang et al. 1998),

probably reflecting their high mobility and a larger latitude for experimentation. Initiation of replication of the linear chromosomes of *Streptomyces* is similar to that of circular bacterial chromosomes in terms of participating enzymes and *oriC* organization (Jakimowicz et al. 1998). Similarly, studies on the initiation of replication of the linear plasmids, although limited, have so far revealed no surprises. The major interest in replication of these linear replicons has thus been focused on the novel end patching process, and some progress has been made in this area. Although the studies are applicable to both linear plasmids and chromosomes, the plasmids are much easier to manipulate in the laboratory.

So far, most end patching studies have focused on the archetypal telomeres and TPs. The appearance of novel telomeres and capping TP systems in linear plasmids (and probably chromosomes) promises to provide interesting variations on the same theme with their own clues. Beyond end patching, there lie a number of more complicated issues. Due to the presence of the telomere–telomere interaction, these linear replicons may face the same postreplicational topological issues as circular DNA molecules, such as resolution of the concatemers (by Topo IV) and pseudodimers (by recombination or TP swapping; Fig. 4).

In addition to vegetative replication, another type of replication takes place during conjugal transfer. Such replication is typically initiated at a specialized origin (*oriT*) on the plasmid (free or integrated), different from that (*oriP*) for vegetative replication. Although the end first model of transfer is attractive, it has not been confirmed. Finally, if the nuclear targeting of TPs in eukaryotic cells is part of a natural exchange system with certain eukaryotes, such an interkingdom transfer system begs for rigorous investigation.

Acknowledgements I thank Ralph Kirby, Hung-Wei Hsu, and Ting-Wen Chen for bioinformatic assistance, Zhongjun Qin and Haruyasu Kinashi for communications of unpublished results, and David Hopwood, Keith Chater, Haruyasu Kinashi, Ralph Kirby, and Hsiu-Hui Tsai for critical reading of the manuscript and suggestions for improvements. This study is supported by grant from the National Science Council, R. O. C. (NSC95-2321-B010-005,) and a grant from the Ministry of Education, R. O. C. (Aim for the Top University Plan).

References

Adams DE, Shekhtman EM, Zechiedrich EL, Schmid MB, Cozzarelli NR (1992) The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. Cell 71:277–288

Arnberg A, Arwert F (1976) DNA–protein complex in circular DNA from Bacillus bacteriophage GA-1. J Virol 18:783–784

Astell CR, Smith M, Chow MB, Ward DC (1979) Structure of the 3′ hairpin termini of four rodent parvovirus genomes: nucleotide sequence homology at origins of DNA replication. Cell 17:691–703

Astell CR, Chow MB, Ward DC (1985) Sequence analysis of the termini of virion and replicative forms of minute virus of mice suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. J Virol 54:171–177

Bao K, Cohen SN (2001) Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. Genes Dev 15:1518–1527

Bao K, Cohen SN (2003) Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. Genes Dev 17:774–785

Bao K, Cohen SN (2004) Reverse transcriptase activity innate to DNA polymerase I and DNA topoisomerase I proteins of *Streptomyces* telomere complex. Proc Natl Acad Sci USA 101:14361–14366

Barre FX, Soballe B, Michel B, Aroyo M, Robertson M, Sherratt D (2001) Circles: the replication–recombination–chromosome segregation connection. Proc Natl Acad Sci USA 98:8189–8195

Bentley SD et al (2004) SCP1, a 356,023 base pair linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). Mol Microbiol 51:1615–1628

Bey S-J, Tsou M-F, Huang C-H, Yang C-C, Chen CW (2000) The homologous terminal sequence of the *Streptomyces lividans* chromosome and SLP2 plasmid. Microbiology 146:911–922

Bignell C, Thomas CM (2001) The bacterial ParA–ParB partitioning proteins. J Biotechnol 91:1–34

Bravo A, Salas M (1997) Initiation of bacteriophage 29 DNA replication in vivo: assembly of a membrane-associated multiprotein complex. J Mol Biol 269:102–112

Chaconas G, Chen CW (2005) Linear chromosomes in bacteria: no longer going around in circles. In: Higgins NP (ed) The bacterial chromosome. American Society for Microbiology, Washington, DC, pp 525–539

Chang P-C, Cohen SN (1994) Bidirectional replication from an internal origin in a linear *Streptomyces* plasmid. Science 265:952–954

Chang P-C, Kim ES, Cohen SN (1996) *Streptomyces* linear plasmids that contain a phage-like, centrally located replication origin. Mol Microbiol 22:789–800

Chater K, Kinashi H (2007) Streptomyces Linear Plasmids: Their Discovery, Functions, Interactions with other Replicons, and Evolutionary Significance. Microbiol Monogr 7. Springer, Heidelberg

Chen CW (1996) Complications and implications of linear bacterial chromosomes. Trends Genet 12:192–196

Chen CW, Yu T-W, Lin Y-S, Kieser HM, Hopwood DA (1993) The conjugative plasmid SLP2 of *Streptomyces* lividans is a 50-kb linear molecule. Mol Microbiol 7:925–932

Chen CW, Huang C-H, Lee H-H, Tsai H-H, Kirby R (2002) Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes. Trends Genet 18:522–529

Chen Y-T (2000) Linear chromosomes and linear plasmids in actinomycetes. In: Institute of Genetics. National Yang-Ming University, Taipei

Chou S-H, Zhu L, Reid BR (1997) Sheared purine–purine pairing in biology. J Mol Biol 267:1055–1067

Fetzner S, Kolkenbrock S, Parschat K (2007) Catabolic Linear Plasmids. Microbiol Monogr 7. Springer, Heidelberg

Fischer G, Wenner T, Decaris B, Leblond P (1998) Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. Proc Natl Acad Sci USA 95:14296–14301

Flett F, Platt J, Oliver SG (1992) Isolation and characterization of temperature-sensitive mutants of *Streptomyces coelicolor* A3(2) blocked in macromolecular synthesis. J Gen Microbiol 138:579–585

Flett F, de Mello Jungmann-Campello D, Mersinias V, Koh SL, Godden R, Smith CP (1999) A "gram-negative-type" DNA polymerase III is essential for replication of the linear chromosome of *Streptomyces coelicolor* A3(2). Mol Microbiol 31:949–958

Gerdes K, Moller-Jensen J, Jensen RB (2000) Plasmid and chromosome partitioning: surprises from phylogeny. Mol Microbiol 37:455–466

Goshi K et al (2002) Cloning and analysis of the telomere and terminal inverted repeat of the linear chromosome of *Streptomyces griseus*. J Bacteriol 184:3411–3415

Hiratsu K, Mochizuki S, Kinashi H (2000) Cloning and analysis of the replication origin and the telomeres of the large linear plasmid pSLA2-L in *Streptomyces rochei*. Mol Gen Genet 263:1015–1021

Hirochika H, Nakamura K, Sakaguchi K (1985) A linear DNA plasmid from *Streptomyces rochei* with an inverted repetition of 614 base pairs. EMBO J 3:761–766

Hopwood DA (1967) Genetic analysis and genome structure in *Streptomyces coelicolor*. Bacteriol Rev 31:373–403

Hopwood DA (2006) Soil to genomics: the *Streptomyces* chromosome. Annu Rev Genet 40:123

Hopwood DA, Kieser T (1993) Conjugative plasmids of *Streptomyces*. In: Clewell DB (ed) Bacterial conjugation. Plenum, New York, pp 293–311

Huang C-H (2002) Structures and functions of *Streptomyces* telomeres. PhD-Thesis. In: Institute of Genetics. National Yang-Ming University, Taipei, p 111

Huang C-H, Lin Y-S, Yang Y-L, Huang S-W, Chen CW (1998) The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. Mol Microbiol 28:905–926

Huang C-H, Chen C-Y, Tsai H-H, Chen C, Lin Y-S, Chen CW (2003) Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. Mol Microbiol 47:1563–1576

Huang T-W, Chen CW (2006) A recA null mutation may be generated in *Streptomyces coelicolor*. J Bacteriol 188:6771–6779

Huang C-H, Tsai H-H, Tsay Y-G, Chien Y-N, Wang S-L, Cheng M-Y, Ke C-H, Chen CW (2007) The Telomere System of the *Streptomyces* Linear Plasmid SCP1 Represents a Novel Class. Mol Microbiol 63:1710–1718

Ikeda H et al (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces* avermitilis. Nat Biotechnol 21:526–531

Jakimowicz D et al (1998) Structural elements of the *Streptomyces oriC* region and their interactions with the DnaA protein. Microbiology 144:1281–1290

Keegstra W, Van Wielink P, Sussenbach J (1977) The visualization of a circular DNA–protein complex from adenovirions. Virology 76:444–447

Klassen R, Meinhardt F (2007) Linear Protein-primed Replicating Plasmids in Eukaryotic Microbes. Microbiol Monogr 7. Springer, Heidelberg

Kinashi H, Shimaji-Murayama M, Hanafusa T (1991) Nucleotide sequence analysis of the unusually long terminal inverted repeats of a giant linear plasmid, SCP1. Plasmid 26:123–130

Kornberg A, Baker TA (1992) DNA replication. Freeman, New York

Lin Y-S, Kieser HM, Hopwood DA, Chen CW (1993) The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol Microbiol 10:923–933

McLeod MP et al (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. Proc Natl Acad Sci USA 103:15582–15587

Mochizuki S, Hiratsu K, Akita A, Kinashi H (2001) Complete nucleotide sequence of the linear plasmid pSLA2-L revealed unusual gene organizations for secondary metabolism. In: Twelfth international symposium on the biology of actinomycetes, Vancouver, Canada, p 60

Muth G, Frese D, Kleber A, Wohlleben W (1997) Mutational analysis of the *Streptomyces* lividans recA gene suggests that only mutants with residual activity remain viable. Mol Gen Genet 255:420–428

Ortin J, Vinuela E, Salas M, Vasquez C (1971) DNA–protein complex in circular DNA from phage 29. Nat New Biol 234:275–277

Possoz C, Ribard C, Gagnat J, Pernodet JL, Guerineau M (2001) The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. Mol Microbiol 42:159–166

Qin Z, Cohen SN (1998) Replication at the telomeres of the *Streptomyces* linear plasmid pSLA2. Mol Microbiol 28:893–904

Qin Z, Cohen SN (2002) Survival mechanisms for *Streptomyces* linear replicons after telomere damage. Mol Microbiol 45:785–794

Qin Z, Shen M, Cohen SN (2003) Identification and characterization of a pSLA2 plasmid locus required for linear DNA replication and circular plasmid stable inheritance in *Streptomyces lividans*. J Bacteriol 185:6575–6582

Redenbach M, Bibb M, Gust B, Seitz B, Spychaj A (1999) The linear plasmid SCP1 of *Streptomyces coelicolor* A3(2) possesses a centrally located replication origin and shows significant homology to the transposon Tn4811. Plasmid 42:174–185

Redenbach M, Scheel J, Schmidt U (2000) Chromosome topology and genome size of selected actinomycetes species. Antonie Van Leeuwenhoek 78:227–235

Reuther J, Gekeler C, Tiffert Y, Wohlleben W, Muth G (2006) Unique conjugation mechanism in mycelial streptomycetes: a DNA-binding ATPase translocates unprocessed plasmid DNA at the hyphal tip. Mol Microbiol 61:436–446

Robinson AJ, Younghusband HB, Bellett AJ (1973) A circular DNA–protein complex from adenoviruses. Virology 56:54–69

Sakaguchi K (1990) Invertrons, a class of structurally and functionally related genetic elements that include linear plasmids, transposable elements, and genomes of adenotype viruses. Microbiol Rev 54:66–74

Schaack J, Ho WY, Freimuth P, Shenk T (1990) Adenovirus terminal protein mediates both nuclear matrix association and efficient transcription of adenovirus DNA. Genes Dev 4:1197–1208

Sekine M et al (2006) Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. Environ Microbiol 8:334–346

Shiffman D, Cohen SN (1992) Reconstruction of a *Streptomyces* linear plasmid replicon from separately cloned DNA fragment: existence of a cryptic origin of circular replication within the linear plasmid. Proc Natl Acad Sci USA 89:6129–6133

Shih YL, Rothfield L (2006) The bacterial cytoskeleton. Microbiol Mol Biol Rev 70:729–754

Stahl FW (1967) Circular genetic maps. J Cell Physiol 70:1–12

Stahl FW, Steinberg CM (1964) The theory of formal phage genetics for circular maps. Genetics 50:531–538

Uchida T, Miyawaki M, Kinashi H (2003) Chromosomal arm replacement in *Streptomyces griseus*. J Bacteriol 185:1120–1124

Wang S-J, Chang H-M, Lin Y-S, Huang C-H, Chen CW (1999) *Streptomyces* genomes: circular genetic maps from the linear chromosomes. Microbiology 145:2209–2220

Weaver D et al (2004) Genome plasticity in *Streptomyces*: identification of 1-Mb TIRs in the *S. coelicolor* A3(2) chromosome. Mol Microbiol 51:1530–1550

Wong M, Hsu M (1989) Linear adenovirus DNA is organized into supercoiled domains in virus particles. Nucleic Acids Res 17:3535–3550

Xu M et al (2006) Characterization of the genetic components of *Streptomyces lividans* linear plasmid SLP2 for replication in circular and linear modes. J Bacteriol 188:6851–6857

Yamasaki M, Kinashi H (2004) Two chimeric chromosomes of *Streptomyces coelicolor* A3(2) generated by single crossover of the wild-type chromosome and linear plasmid scp1. J Bacteriol 186:6553–6559

Yang C-C, Huang C-H, Li C-Y, Tsay Y-G, Lee S-C, Chen CW (2002) The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins. Mol Microbiol 43:297–305

Yang C-C, Chen Y-H, Tsai H-H, Huang C-H, Huang T-W, Chen CW (2006) In vitro deoxynucleotidylation of terminal protein of *Streptomyces* linear chromosomes. J Bacteriol 72:7959–7961

Zambryski P, Tempe J, Schell J (1989) Transfer and function of T-DNA genes from agrobacterium Ti and Ri plasmids in plants. Cell 56:193–201

Zechiedrich EL, Cozzarelli NR (1995) Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in *Escherichia coli*. Genes Dev 9:2859–2869

Zhang R et al (2006) Diversity of telomere palindromic sequences and replication genes among *Streptomyces* linear plasmids. Appl Environ Microbiol 72:5728–5733

Zhu L, Chou S-H, Xu J, Reid BR (1995) Structure of a single-cytidine hairpin loop formed by the DNA triplet GCA. Nat Struct Biol 2:1012–1017

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_091/Published online: 1 June 2007

Catabolic Linear Plasmids

Susanne Fetzner (✉) · Stephan Kolkenbrock · Katja Parschat

Institut für Molekulare Mikrobiologie und Biotechnologie,
Westfälische Wilhelms-Universität Münster,
Corrensstraße 3, 48149 Münster, Germany
fetzner@uni-muenster.de

1 Introduction 64

2 Linear Plasmids Involved in the Degradation of Halogenated Alkenes and Alkanes 65
2.1 Degradation of Short-Chain Alkenes and Chloroalkenes via the Coenzyme M Pathway of Epoxide Metabolism is Encoded on Linear Plasmids in Phylogenetically Different Bacteria 65
2.2 Genes Encoding 1,2-Dichloroethane Degradation by *Xanthobacter autotrophicus* GJ10 Are Segregated Between the Chromosome and pXAU1 73

3 *Rhodococcus* spp.: Linear Plasmids and Other Genetic Determinants of Their Catabolic Versatility 74
3.1 Alkane Oxidation Genes Located on pREL1 of *Rhodococcus erythropolis* PR4 75
3.2 Linear Plasmids Involved in the Degradation of Biphenyl and Monocyclic Aromatic Compounds by *Rhodococcus* sp. Strains 76
3.2.1 Catabolic Gene Clusters of *Rhodococcus* sp. RHA1 Are Distributed on Several Replicons 76
3.2.2 Megaplasmids pDK2 and pRHL2, as well as pDK3 and pRHL1, Share Almost Identical Catabolic Gene Clusters 79
3.2.3 Multiple Extradiol Dioxygenase Genes Are Distributed on the Genomes of the Biphenyl Degrading Strains *Rhodococcus erythropolis* TA421 and *Rhodococcus globerulus* P6 81
3.2.4 Alkylbenzene Degradation and Co-oxidation of Trichloroethene by *Rhodococcus erythropolis* BD2 is Encoded on pBD2 81
3.2.5 Plasmid p1CP of *Rhodococcus opacus* 1CP Codes for Two Distinct Chlorocatechol Pathways 82
3.3 Linear Plasmids Involved in the Degradation of Polycyclic Aromatic Compounds by *Rhodococcus* sp. Strains 84
3.3.1 Modular Organization and Different Genomic Localization of Naphthalene Degradation Genes in Strains I24, P200, P400, NCIMB12038, CIR2, TKN14, R7, and SAO101 84

4 *Terrabacter* sp. 87
4.1 Dibenzofuran and Fluorene Oxidation via Angular Dioxygenation as well as the Protocatechuate Pathway are Encoded on pDBF1 of *Terrabacter* sp. DBF63 87

5 *Arthrobacter* sp. 90
5.1 Linear Plasmid pAL1 of *Arthrobacter nitroguajacolicus* Rü61a Codes for Degradation of 2-Methylquinoline 90

6 Concluding Remarks 92

References 92

Abstract Catabolic gene clusters localized on invertron-type linear plasmids significantly contribute to the ability of actinobacteria to degrade a wide range of organic compounds. Especially in *Rhodococcus* spp., several large linear plasmids have been identified which are involved in alkane oxidation, or in the degradation of aromatic compounds like fluorene, dibenzofuran, naphthalene, biphenyl, or alkylbenzenes and other monocyclic aromatic compounds. Rhodococci often contain multiple copies of key catabolic genes, which is thought to be an important factor for their catabolic efficiency and versatility. Other actinobacteria with catabolic linear plasmids include *Terrabacter* sp. DBF63 containing the pDBF1 plasmid that codes for the degradation of fluorene, and *Arthrobacter nitroguajacolicus* Rü61a harboring pAL1 encoding 2-methylquinoline conversion. Remarkably, linear replicons carrying genes for the degradation of short-chain alkenes and chloroalkenes via the coenzyme M pathway were identified not only in Gram-positive bacteria, such as *Mycobacterium* sp. strains, *Gordonia rubripertincta* B-276, and *Nocardioides* sp. JS614, but also in the Gram-negative strains *Xanthobacter* sp. Py2, *Pseudomonas putida* AJ, and *Ochrobactrum* sp. TD. In *Xanthobacter autotrophicus* GJ10, genes encoding 1,2-dichloroethane degradation are segregated between the chromosome and the linear plasmid pXAU1. The presence of highly homologous gene clusters on catabolic plasmids of phylogenetically different bacteria and the genetic organization of some linear plasmids sequenced as yet suggest that horizontal transfer of mobile genetic elements and genomic rearrangements significantly contribute to the evolution of catabolic diversity.

1 Introduction

Catabolic linear plasmids have been found in various soil bacteria, particularly in rhodococci. The linear plasmids of actinobacteria with biodegradative capacities are typical invertrons (Sakaguchi 1990), containing terminal inverted repeats and terminal proteins covalently bound to each 5′ end. Some Gram-negative bacteria, e.g., the haloalkane degrader *Xanthobacter autotrophicus* GJ10, were also reported to harbor linear catabolic plasmids, but it is not yet known whether proteins are attached to their ends, and the sequences of their telomeres also have not been investigated yet. Some other linear replicons of Gram-negative bacteria, such as N15 of *Escherichia coli*, are prophages which are not integrated into the bacterial chromosome but exist as linear plasmid molecules with covalently closed hairpin ends (Hertwig 2007, in this volume); however, note that these DNA elements do not code for catabolic traits.

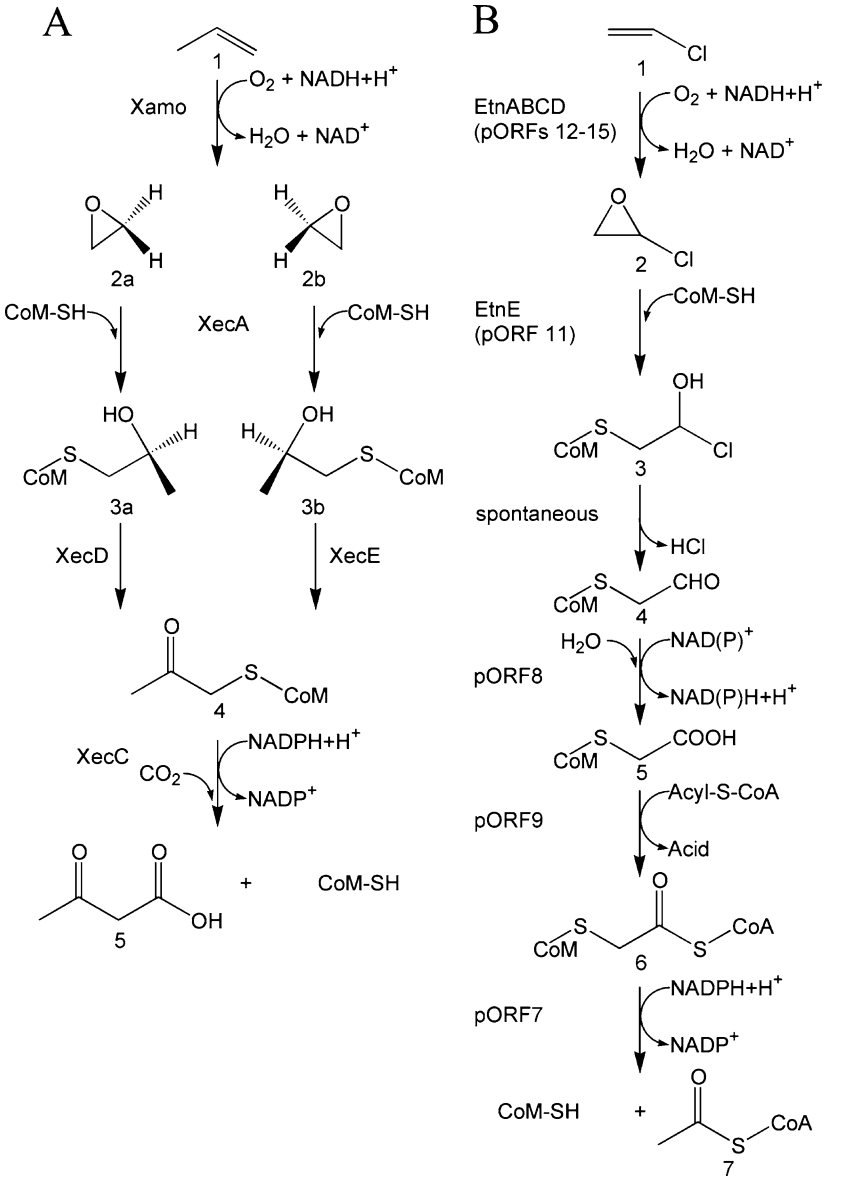
There is much more information on circular plasmids involved in the biodegradation of carbon compounds (especially xenobiotics) than on linear catabolic plasmids (for a review, see, e.g., Nojiri et al. 2004), but some of the catabolic plasmids mentioned in the literature have not been fully characterized, and it is not always clear whether they are circular or linear. Reports on catabolic plasmids have often focused on the characterization of degradative genes or gene clusters and the phenotypic properties conferred by these genes; however, recent analyses of the complete nucleotide sequences of linear catabolic plasmids have significantly enhanced our understanding of these replicons. Ongoing and future genomic approaches will further improve insight into the structure and evolution and—together with biochemical studies—into the function of linear plasmids. This review gives an overview of catabolic linear plasmids, with an emphasis on the organization and physiological role of catabolic gene clusters.

2
Linear Plasmids Involved in the Degradation of Halogenated Alkenes and Alkanes

2.1
Degradation of Short-Chain Alkenes and Chloroalkenes via the Coenzyme M Pathway of Epoxide Metabolism is Encoded on Linear Plasmids in Phylogenetically Different Bacteria

Since chloroethenes, which include known or suspected carcinogens, are widely distributed at many contaminated sites, their microbial degradation has been investigated intensely, and aerobic as well as anaerobic transformation pathways have been described (for a review, see Fetzner 1998a). Linear plasmids have been implicated in alkene and chloroethene metabolism by phylogenetically very different bacteria, such as *Mycobacterium* strains, *Gordonia rubripertincta* B-276 (formerly *Rhodococcus corallinus, Nocardia corallina*), *Nocardioides* sp. JS614, *Xanthobacter* sp. Py2, *Pseudomonas putida* AJ, and *Ochrobactrum* sp. TD (Saeki et al. 1999; Krum and Ensign 2001; Danko et al. 2004, 2006; Mattes et al. 2005).

The Gram-negative bacterium *Xanthobacter* sp. Py2 oxidizes propene via epoxidation and carboxylation to form acetoacetate (Fig. 1A), which can be easily converted to central metabolites (for a review, see Ensign and Allen 2003). Loss of the 320-kb megaplasmid pEK1, proposed to be linear, was accompanied by loss of alkene monooxygenase and carboxylase activity, and loss of coenzyme M biosynthesis. The alkene monooxygenase genes *xamoABCDEF*, as well as the *xecABCDEFG* cluster that codes for the four enzymes of epoxide metabolism (XecA, -D, -E, -C), two proteins of unknown function (XecB, -F), and a putative (2*R*)-phospho-3-sulfolactate synthase



(XecG = ComA) involved in biosynthesis of coenzyme M, are located on pEK1 (Krum and Ensign 2001) (Table 1).

Gordonia rubripertincta B-276, which utilizes propene and oxidizes trichloroethene (TCE), contains four linear plasmids (Table 1). A three-component alkene monooxygenase (Saeki and Furuhashi 1994) catalyzes the oxidation of propene, trichloroethene, and probably dichloroethenes and

◀ **Fig. 1** Alkene degradation via the coenzyme M pathway. (**A**): Propene degradation by *Xanthobacter* sp. strain Py2 (Ensign and Allen 2003). CoM, coenzyme M (2-mercaptoethanesulfonate); **1**, propene; **2a**, (*R*)-epoxypropane; **2b**, (*S*)-epoxypropane; **3a**, 2-(*R*)-hydroxypropyl-CoM; **3b**, 2-(*S*)-hydroxypropyl-CoM; **4**, 2-ketopropyl-CoM; **5**, acetoacetate. Xamo, alkene monooxygenase; XecA, epoxyalkane:CoM transferase; XecD, 2-(*R*)-hydroxypropyl-CoM dehydrogenase; XecE, 2-(*S*)-hydroxypropyl-CoM dehydrogenase; XecC, NADPH:2-ketopropyl-CoM carboxylase/oxidoreductase. (**B**): Hypothetical pathway of vinyl chloride degradation by *Nocardioides* sp. JS614 (Mattes et al. 2005). CoM, coenzyme M (2-mercaptoethanesulfonate); **1**, vinyl chloride; **2**, chloroepoxyethane; **3**, 2-chloro-2-hydroxyethyl-CoM; **4**, 2-ketoethyl-CoM; **5**, carboxymethyl-CoM; **6**, CoM-acetyl-coenzyme A; **7**, acetyl-coenzyme A. EtnABCD, alkene monooxygenase; EtnE, epoxyalkane:CoM transferase. For proteins of open reading frames (ORFs) 7–9, see text

dichloropropenes to the corresponding epoxides. The alkene monooxygenase genes *amoABCD*, as well as genes involved in the downstream catabolism of epoxides, are encoded by the 185-kb linear plasmid pNC30 (Saeki et al. 1999).

The linear plasmid pNoc614 of *Nocardioides* sp. JS614 was discussed to encode the complete pathway of vinyl chloride and ethene oxidation to central metabolites (Mattes et al. 2005; see Fig. 1B for vinyl chloride oxidation). In the first degradation step, alkene monooxygenase (EtnABCD) catalyzes the formation of epoxyethane and chlorooxirane from ethene and vinyl chloride, respectively. Subsequent coenzyme M transfer to these epoxyalkanes, catalyzed by EtnE, yields 2-hydroxyethyl-CoM from epoxyethane, and 2-chloro-2-hydroxyethyl-CoM from chlorooxirane. Whereas the latter may eliminate HCl to form 2-ketoethyl-CoM (Fig. 1B), formation of 2-ketoethyl-CoM from 2-hydroxyethyl-CoM would require the action of a dehydrogenase. Further steps of the hypothetical pathway may involve oxidation to carboxymethyl-CoM, activation to CoM-acetyl-CoA, and reductive cleavage of the CoM-acetyl-CoA thioether to acetyl-CoA and CoM-SH (Mattes et al. 2005). Hybridization studies and sequencing revealed that pNoc614 carries the *etnEABCD* locus, putative acyl-CoA transferase and acyl-CoA synthetase genes (ORFs 9 and 10), and two putative oxidoreductase genes (ORFs 7 and 8) (Table 1; Fig. 2). The deduced gene products of ORF9, ORF10, and *etnEABCD* of strain JS614 show significant similarity to the corresponding proteins of *Mycobacterium rhodesiae* JS60 (36–83% amino acid (aa) identity), which are also encoded on a linear plasmid (Coleman and Spain 2003b). The EtnABCD proteins of strain JS60 also share 41–59% sequence identity with the corresponding gene products of *amoABCD* of *G. rubripertincta* B-276, suggesting a common ancestry (Fig. 2). Remarkably, the ORF7 and ORF8 proteins of pNoc614 of *Nocardioides* sp. JS614 also show some similarity to NADPH:2-ketopropyl-CoM carboxylase/oxidoreductase (XecC; 24% aa identity) and 2-(*S*)-hydroxypropyl-CoM dehydrogenase (XecE; 47% aa identity), respectively, of the Gram-negative strain *Xanthobacter* sp. Py2. A gene presumed to code for (2*R*)-phospho-3-sulfolactate synthase, which catalyzes the first step in CoM biosynthesis, likewise is localized on the plasmids of both

Table 1 Overview of linear catabolic plasmids

Organism	Plasmid [a]	Plasmid size (kb)	Significance in catabolism; catabolic genes identified [accession number]	Refs.
Xanthobacter sp. Py2	pEK1	320	(Chloro)alkene metabolism via coenzyme M pathway: *xamoABCDEF* encoding alkene monooxygenase [AJ012090]; *xecABCDEFG* involved in epoxide metabolism and coenzyme M synthesis [X79863, AY024334].	Swaving et al. 1995 Zhou et al. 1999 Krum and Ensign 2001
Gordonia rubripertincta B-276	pNC10 pNC20 pNC30	70 85 185	Alkene metabolism and co-oxidation of trichloroethene: *amoABCD* encoding alkene monooxygenase [D37875].	Saeki et al. 1999
	pNC40	235		
Nocardioides sp. JS614	pNoc614	290	(Chloro)alkene metabolism via coenzyme M pathway: *etnEABCD* (ORFs 11–15) encoding alkene monooxygenase (EtnABCD) and epoxyalkane: coenzyme M transferase (EtnE), genes for putative CoA transferase (ORF9), putative acyl-CoA synthetase (ORF10), two putative oxidoreductases (ORF7, *xecC*-like, and ORF8, *xecE*-like) [AY772007].	Mattes et al. 2005
Xanthobacter autotrophicus GJ10	pXAU1	225	1,2-Dichloroethane metabolism: haloalkane dehalogenase gene *dhlA* [M26950]; *aldA* encoding 2-chloroacetaldehyde dehydrogenase [AF029733].	Tardif et al. 1991 van der Ploeg et al. 1994 Bergeron et al. 1998
Rhodococcus erythropolis PR4	pREL1	271.577	Alkane oxidation to acyl-CoA (putative): ORFs pREL1_0256–pREL1_0261 and pREL1_0282–pREL1_0286 [AP008931].	Sekine et al. 2006
	pREC1 [a]	104.014	Complete set of genes for β-oxidation of fatty acids (ORFs pREC1_0064–pREC1_0074) [AP008932].	
	pREC2 [a]	3.637	[AP008933].	

Table 1 (continued)

Organism	Plasmid[a]	Plasmid size (kb)	Significance in catabolism; catabolic genes identified [accession number]	Refs.
Rhodococcus sp. RHA1	pRHL1	1123.075	(Polychlorinated) biphenyl and ethylbenzene metabolism: *etbD1, bphAaAbAcAdC1B1, bphS1, bphT1, bphF3G3E3, bphRG1F1E1*. Putative extradiol dioxygenases (ro08079, ro09005), catechol 1,2-dioxygenases (ro08069, ro08511), C–C hydrolases (ro08081, ro09014), maleylacetate reductase (ro08510), phenol monooxygenase (ro08076–08078). Terephthalate and phthalate metabolism: *tpaKBCAbAa, tpaR, patDABCE, padR, padAaAbBAcAdC*. Possible carveol dehydrogenases *limC1* and *limC2* (ro08210, ro08632) and limonene monooxygenase (ro09085). Putative ring hydroxylating oxygenase (ro08637–08639); genes for cytochrome P450 systems (ro08067–08068, ro08984–08986); genes involved in steroid catabolism (ro09001–09005) [CP000432].	Masai et al. 1997 Shimizu et al. 2001 Warren et al. 2004 Patrauchan et al. 2005 McLeod et al. 2006 www.rhodococcus.ca
	pRHL2	442.536	(Polychlorinated) biphenyl and ethylbenzene metabolism: *bphF4, bphG4E4, bphT2, bphS2, etbAd, bphB2, etbAa1Ab1C, bphD1E2F2, etbAa2Ab2AcD2* (formerly *ebdA1A2A3D2*); extradiol dioxygenase gene (ro10315). Terephthalate and phthalate metabolism: *tpaKBCAbAa, tpaR, patDABCE, padR, padAaAbBAcAdC*. Putative 2,3-dichlorophenol 6-monooxygenase (ro10313) [CP000433].	
	pRHL3	332.361	Putative carveol dehydrogenase gene (RHL3.41, *limC3*) and limonene monooxygenase gene (RHL3.42, *limB*), genes for cytochrome P450 proteins (RHL3.62, 3.246, 3.287). Metabolism of aromatic compounds: putative maleylacetate reductase, intradiol dioxygenase, (indole) dioxygenase genes (RHL3.276–3.280) [CP000434].	
Rhodococcus sp. DK17	pDK1 (?)	380		Kim et al. 2002, 2004 Choi et al. 2005
	pDK2 (?)	330	Alkylbenzene metabolism: *akbTS*, (ORF14)*akbA4B*, (ORFs15–20), $akbA1_bA2_bCDEF$, (ORF21), $akbA1_aA2_aA3$; putative genes for 4-hydroxy-2-oxovalerate aldolase, acetaldehyde dehydrogenase, 2-hydroxypenta-2,4-dienoate hydratase, enoyl-CoA hydratase, acyl-CoA ligase. Phthalate and terephthalate metabolism: *ophA1A2*, (ORF0), *ophBA3A4C, ophR, pehA, ptrDABC, tphR, tphA1A2BA4* [AY502075].	

Table 1 (continued)

Organism	Plasmid[a]	Plasmid size (kb)	Significance in catabolism; catabolic genes identified [accession number]	Refs.
	pDK3 (?)	750	Phthalate and terephthalate metabolism: *ophA1A2*, (ORF0), *ophBA3A4C*, *ophR*, *pehA*, *ptrDABC*, *tphR, tphA1A2BA4* [AY502076].	
R. erythropolis TA421	pTA421	560	Biphenyl metabolism: *bphC2* [D88014]; *bphA1A2A3A4BC3STD* [D88020, D88015, AB014348]; *bphA1A2A3BA4C4D* [D88021, D88016].	Kosono et al. 1997 Arai et al. 1998
R. globerulus P6	pLP6 pSP6	650 360	Biphenyl metabolism: *bphC2* (extradiol dioxygenase). Biphenyl metabolism: *bphC2, bphC3, bphC4* (extradiol dioxygenases).	Kosono et al. 1997
R. rhodochrous K37		270 200	Biphenyl degradation: *bphC6* [AB117724]; *bphC8* [AB117726] (extradiol dioxygenases).	Taguchi et al. 2004
R. erythropolis BD2	pBD2	210.250	Isopropylbenzene metabolism and co-oxidation of trichloroethene: *ipbA1A2A3A4CB* (PBD2.153-158), *ipbST* (PBD2.159-169), *ipbD* (PBD2.174) [AY223810].	Dabrock et al. 1994 Kesseler et al. 1996 Stecker et al. 2003
Rhodococcus sp. I24	(?) (?)	340 80	Toluene-inducible dioxygenase, indene oxidation: *tidABCD* [AF452376]. Naphthalene metabolism and indene oxidation: *nidABCD* [AF121905]; *nimR* (transcriptional regulator), *nimA* (naphthalene inducible monooxygenase), *nimB* (extradiol dioxygenase) [AF452375.2].	Priefert et al. 2004
R. opacus 1CP	p1CP	740	4-Chloro-/3,5-dichlorocatechol metabolism (*clcBRAD*) [AF003948]; 3-chloro-catechol metabolism (*clcA2D2B2F*) [AJ439407]; *macA* encoding maleylacetate reductase [AF030176].	Moiseeva et al. 2002 König et al. 2004
Rhodococcus sp. NCIMB12038	p2SL1	380	Naphthalene metabolism: *rub1, narR1, narR2, rub2,* (ORF7), *narAaAbB, ΔnarC*; [AF082663.3]	Kulakov et al. 2005

Table 1 (continued)

Organism	Plasmid[a]	Plasmid size (kb)	Significance in catabolism; catabolic genes identified [accession number]	Refs.
Rhodococcus sp. P400	P40L1 P40C1[a] P40L2 P40L3	420 180 45 20	Naphthalene metabolism: *rub1, narR1, narR2,* (ORF7), *narAaAbBC* [AY392423.2].	Kulakov et al. 2005
R. opacus SAO101	pWK301 pWK302 pWK303	1100 1000 700	Naphthalene, dibenzofuran, and dibenzo-*p*-dioxin metabolism: *rub1*, *narR1*, *narR2,* ORF7, *narAaAbBC* (ORF6, *dodR1, dodR2*, ORF7, *dodA, dodB, dodC*, ORF8 [AB110633]).	Kitagawa et al. 2004 Kimura et al. 2006
R. opacus M213	pNUO1 pNUO2 (?)	750 200–300	Metabolism of aromatic compounds: *edoD* encoding catechol 2,3-dioxygenase.	Uz et al. 2000
Terrabacter sp. DBF63	pDBF1 pDBF2	160 190	Angular dioxygenation of dibenzofuran to 2,2′,3-trihydroxybiphenyl (*dbfA1A2*). Fluorene oxidation to phthalate: *flnR*, *flnB*, *dbfA1A2*, *flnED1*, ORF16, *flnC*; phthalate conversion to protocatechuate: *phtA1A2*, (ORF11), *BphtA3A4CR*; protocatechuate branch of β-ketoadipate pathway: *pcaR*, *pcaHGBDCFIJ*; putative extradiol dioxygenase gene (ORF62) [AP008980]. Derivative of pDF1 (*dbf-fln* and *pht* genes)	Nojiri et al. 2002 Habe et al. 2004 Habe et al. 2005
Arthrobacter nitro-guajacolicus Rü61a	pAL1	112.992	Quinaldine conversion to anthranilate: *xdhC*-like gene; *qoxS*, *qoxM*, *qoxL*, *moq*, *hod*, *amq* (ORF1, *qoxS*, *qoxM*, *qoxL*, ORF2, *hod*, ORF4 [AJ537472]); putative anthranilate metabolism via anthranoyl-CoA and 2-amino-5-oxo-cyclohex-1-ene-carbonyl-CoA.	Overhage et al. 2005 Parschat et al. 2007

[a]: Indicates circularity; (?): Denotes unclear topology

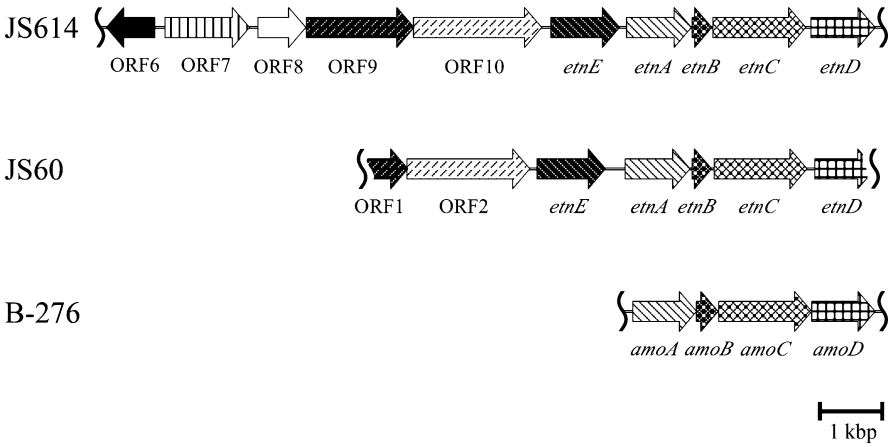


Fig. 2 Organization of genes involved in alkene degradation. JS614, *etn* locus on the linear plasmid pNoc614 of *Nocardioides* sp. strain JS614 (Mattes et al. 2005; GenBank accession number AY772007); JS60, homologues on a linear plasmid of *Mycobacterium rhodesiae* JS60 (Coleman et al. 2003a; AY243034); B-276, *amo* genes located on pNC30 of *Gordonia rubripertincta* B-276 (Saeki and Furuhashi 1994; D37875). The *amoABCD* and *etnABCD* genes code for multicomponent alkene monooxygenases; *etnE* encodes epoxyalkane:CoM transferase. For the other ORFs, see text. The *arrows* indicate the size, location, and direction of transcription of the genes and ORFs; homologous ORFs feature the same pattern. Truncated regions are marked with the symbol ⟆

strains JS614 and Py2 (*comA* = *xecG*, ORF6 on the pNoc614 segment; Mattes et al. 2005).

When six vinyl chloride degrading mycobacteria, which all possess homologous epoxyalkane:CoM transferase genes (*etnE*), were screened for plasmids, large linear DNA elements (110–330 kb) were found in all strains, namely, *Mycobacterium* sp. strains JS61, JS616, JS617, JS619, JS621, and *M. rhodesiae* JS60. In fact, almost identical *etnE* sequences were found in a total of ten ethene/vinyl chloride degrading isolates that were obtained from geographically distant locations. The phylogeny of the 16SrDNA of the strains significantly differed from the phylogeny of the *etnE* gene sequences (Coleman and Spain 2003a,b), indicating horizontal gene transfer among these mycobacteria. Notably, a recent report suggests that gene transfer may have occurred even between Gram-positive and Gram-negative bacteria. Plasmid DNA of *Pseudomonas putida* AJ and *Ochrobactrum* sp. TD was found to contain *etnE* genes almost identical to each other and to those of *Mycobacterium* strains JS60 and JS621 (Danko et al. 2006). Based on their behavior in pulsed-field gel electrophoresis, the plasmids of *P. putida* AJ and *Ochrobactrum* sp. TD, which are essential for the ability to degrade vinyl chloride and ethene, were reported to have a linear topology. These plasmids, however, appear to readily undergo deletions and rearrangements, depending on the carbon

source used for growth; in complex medium, they were cured rapidly (Danko et al. 2004, 2006).

2.2 Genes Encoding 1,2-Dichloroethane Degradation by *Xanthobacter autotrophicus* GJ10 Are Segregated Between the Chromosome and pXAU1

Xanthobacter autotrophicus GJ10 is able to metabolize 1,2-dichloroethane (Janssen et al. 1985, 1989), a compound produced by the chemical industry primarily for use in the synthesis of vinyl chloride. Degradation (Fig. 3) is initiated by haloalkane dehalogenase DhlA, which catalyzes a hydrolytic dehalogenation reaction to form 2-chloroethanol (Janssen et al. 1989; Verschueren et al. 1993). Subsequent dehydrogenation by an alcohol dehydrogenase (Mox)

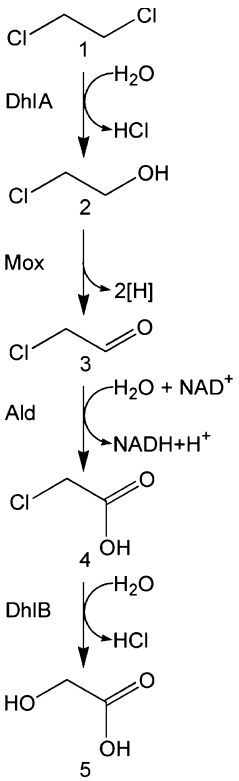


Fig. 3 Degradation of 1,2-dichloroethane by *Xanthobacter autotrophicus* GJ10 (Janssen et al. 1985). **1**, 1,2-Dichloroethane; **2**, 2-chloroethanol; **3**, chloroacetaldehyde; **4**, chloroacetate; **5**, glycolate. DhlA, haloalkane dehalogenase; Mox, (pyrroloquinoline quinone-dependent) alcohol dehydrogenase; Ald, chloroacetaldehyde dehydrogenase; DhlB, 2-haloacid dehalogenase

results in formation of the toxic compound 2-chloroacetaldehyde, which subsequently is oxidized to 2-chloroacetate (van der Ploeg et al. 1994). Strain GJ10 harbors a second hydrolytic dehalogenase, namely, haloacid dehalogenase DhlB (Ridder et al. 1997), which converts 2-chloroacetate to glycolate (van der Ploeg et al. 1991). Two chloroacetaldehyde dehydrogenases have been identified: one is encoded by the chromosomal *aldB* gene, the other by *aldA* located on the 225-kb plasmid pXAU1 which was shown to be linear by pulsed-field gel electrophoresis (Bergeron et al. 1998). Haloalkane dehalogenase DhlA likewise is encoded by pXAU1, whereas the *mox* and *dhlb* genes are located on the chromosome (Tardif et al. 1991; van der Ploeg et al. 1994).

3 *Rhodococcus* spp.: Linear Plasmids and Other Genetic Determinants of Their Catabolic Versatility

Rhodococci are aerobic, nocardioform actinomycetes with the ability to assimilate a wide variety of organic compounds, including hydrophobic xenobiotics like polychlorinated biphenyls (PCBs) or dibenzo-*p*-dioxins. Their general tolerance to toxic substrates, solvents, desiccation, and other (environmental) stress factors, as well as their ability to produce mycolic acid surfactants which may facilitate uptake of hydrophobic compounds, are thought to contribute to the success of members of this genus in biodegradation and bioremediation (Patrauchan et al. 2005; Larkin et al. 2005).

A large number of plasmids have been identified in different *Rhodococcus* spp., which range from small circular to large linear DNA molecules. Many linear plasmids of rhodococci code for catabolic functions. In contrast, the linear conjugative plasmid pFiD188 of the phytopathogenic *R. fascians* strain D188 carries different pathogenicity loci involved in induction of leafy gall formation (reviewed by Goethals et al. 2001; Francis et al. 2007, in this volume).

The termini of rhodococcal linear plasmids, in contrast to those of many *Streptomyces* replicons (Chen 2007, in this volume), generally do not contain long inverted repeats; however, alignments of terminal sequences of linear plasmids from rhodococci and other actinobacteria revealed the presence of palindromes with the common central motif GCTXCGC (Stecker et al. 2003; Warren et al. 2004; König et al. 2004). This highly conserved motif was originally identified by Kalkus et al. (1998) in pHG201 of *R. opacus* MR11. In palindromic sequences of single-stranded terminal 3′ overhangs, it has the potential to form a stable single-residue hairpin loop closed by sheared purine:purine pairing, which may have a role in replication at the telomeres (Huang et al. 1998). Remarkably, the two binding sites of single-stranded DNA on telomeric 3′ overhangs of the *Streptomyces* plasmid pSLA2, which

are recognized by the telomere-associated protein (Tap) involved in telomere patching, comprise this conserved GCTXCGC sequence as a core motif (Bao and Cohen 2003).

The ability of rhodococci to attack a wide range of aromatic compounds is in part due to the abundant presence of several types of oxygenases (Larkin et al. 2005); for example, 122 oxygenase genes were annotated to the genome of *Rhodoccoccus* sp. strain RHA1 (www.rhodococcus.ca). Moreover, their genomes often contain multiple copies of homologous genes and degradative gene clusters (contributing to their large genome sizes), resulting in genetic redundancy which is believed to be an important factor for their catabolic efficiency and versatility (van der Geize and Dijkhuizen 2004; Larkin et al. 2005). Well-studied examples are the presence of multiple alkane monooxygenase genes (*alkB*) (Whyte et al. 2002; van Beilen et al. 2002) and extradiol dioxygenase genes (Irvine et al. 2000; Vaillancourt et al. 2003; McKay et al. 2003; Iida et al. 2002b; Taguchi et al. 2004). Multiple copies of biodegradative genes appear to be "stored" on large linear plasmids (Table 1), but circular plasmids encoding catabolic traits likewise exist in rhodococci, such as pRTL1 (chloroalkane degradation), pTE1 (atrazine degradation), or pTC1 (2-methylaniline metabolism) (for reviews, see Larkin et al. 1998; Nojiri et al. 2004).

As exemplified in Sects. 3.2 and 3.3, some catabolic gene clusters found on linear plasmids of different *Rhodococcus* strains share an extremely high degree of similarity in gene organization and sequence, suggesting that they may have been distributed via horizontal gene transfer. Moreover, rhodococcal plasmids may have undergone rearrangements, since catabolic genes were often found to be flanked by transposon-related ORFs and putative insertion sequences (ISs). Genome analyses, and the observation of frequent nonhomologous illegitimate recombination, led to the hypothesis that rhodococci have adopted an evolutionary strategy of gene storage and hyperrecombination, i.e., acquisition and accumulation of many—even multiple—catabolic genes in large genomes, and genomic rearrangements to promote adaptation to organic substrates (Larkin et al. 1998, 2005).

3.1
Alkane Oxidation Genes Located on pREL1 of *Rhodococcus erythropolis* PR4

Rhodococcus erythropolis PR4, an alkane degrading strain, contains the linear plasmid pREL1 and two circular plasmids pREC1 and pREC2 (Table 1), which have been sequenced completely (Sekine et al. 2006). The plasmids contain modular segments that are homologous to corresponding regions of other rhodococcal plasmids. While pREL1 shares several regions of homology with pBD2 of *R. erythropolis* strain BD2, putative alkane degradation genes, organized in two clusters and flanked by transposon and insertion sequences, are located outside these common regions. The first degradative

cluster includes genes presumed to code for fatty acid-CoA ligase, alcohol dehydrogenase, aldehyde dehydrogenase, and a three-component *n*-alkane monooxygenase system. The second putative catabolic gene cluster comprises genes of another alkane monooxygenase and a truncated fatty acid-CoA ligase. Differences in the G+C contents of the two gene clusters and differences in the deduced amino acid sequences of the gene products suggested that independent gene transfer events have led to the recruitment of these two clusters. A complete set of genes for the β-oxidation of fatty acids, which has been discussed to represent a genomic island of extraneous origin, is located on the circular plasmid pREC1, suggesting that the combined action of the gene products of pREL1 and pREC1 contributes to alkane mineralization. On the other hand, the genome contains additional (and in some cases multiple) copies of genes encoding AlkB-like alkane monooxygenase, AlkG- and AlkT-like rubredoxin and rubredoxin reductase, respectively, putative alcohol dehydrogenase AlkJ, and aldehyde dehydrogenase AlkH, as well as AlkK-like fatty acid-CoA ligase, indicating that *R. erythropolis* PR4 may be able to degrade alkanes even without the proteins encoded on pREL1 (Sekine et al. 2006).

3.2 Linear Plasmids Involved in the Degradation of Biphenyl and Monocyclic Aromatic Compounds by *Rhodococcus* sp. Strains

3.2.1 Catabolic Gene Clusters of *Rhodococcus* sp. RHA1 Are Distributed on Several Replicons

The genome of *Rhodococcus* sp. strain RHA1, a very potent biphenyl and PCB degrader, comprises a total of 9.7 Mb and consists of the 7.8-Mb chromosome and three linear plasmids pRHL1, pRHL2, and pRHL3 (McLeod et al. 2006; www.rhodococcus.ca). Genome analysis indicated that the three linear plasmids carry 11 of the 26 peripheral aromatic degradation pathways of strain RHA1, suggesting that they have a significant catabolic role and may serve as reservoirs for catabolic functions (McLeod et al. 2006). Plasmid pRHL3 was predicted to contain three P450 monooxygenase genes, as well as a gene cluster coding for putative aromatic ring dioxygenase, dehydrogenase, and intradiol dioxygenase enzymes that may be involved in oxidation and degradation of aromatic compounds (RHL3.277–3.281). Products of further putative catabolic genes may contribute to glucose metabolism. Two ORFs (RHL3.41 and RHL3.42) presumed to encode a carveol dehydrogenase and a limonene monooxygenase are clustered together with a series of dehydrogenase genes. Therefore, it has been suggested that the physiological role of pRHL3 is to increase the catabolic capabilities of strain RHA1 (Warren et al. 2004).

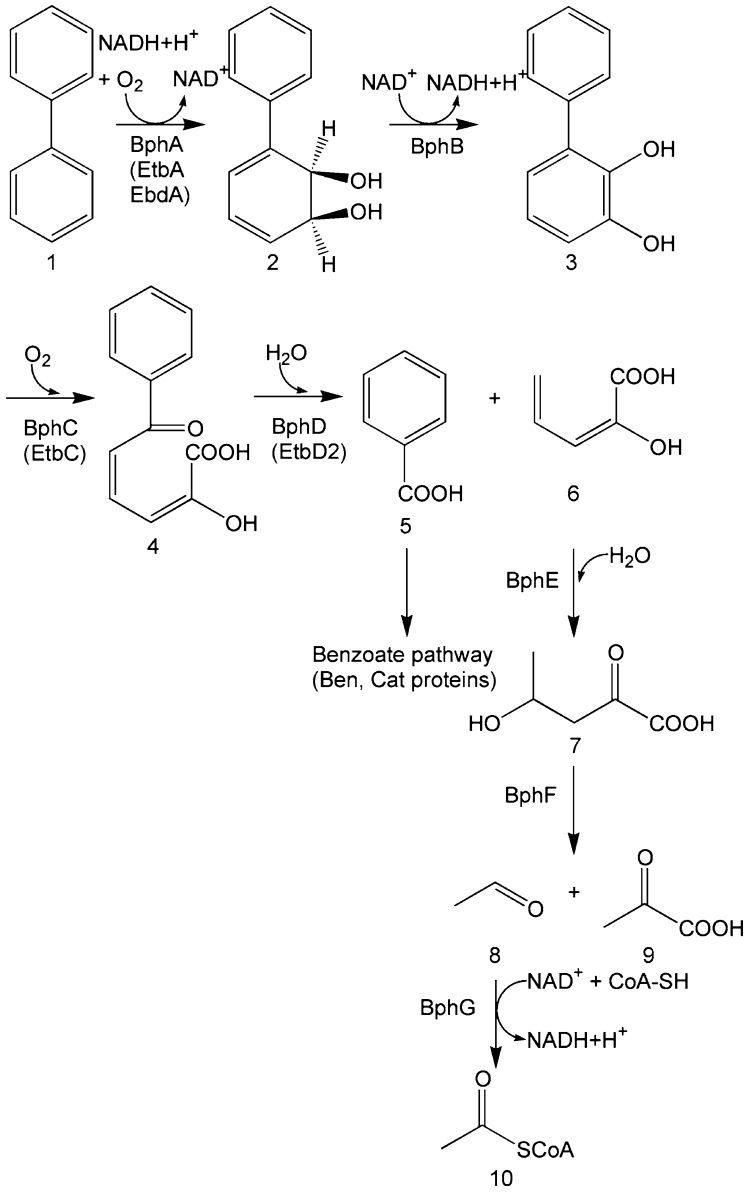


Fig. 4 Aerobic bacterial degradation of biphenyl (http://umbbd.msi.umn.edu/). **1**, Biphenyl; **2**, *cis*-2,3-dihydro-2,3-dihydroxybiphenyl; **3**, 2,3-dihydroxybiphenyl; **4**, 2-hydroxy-6-oxophenylhexa-2,4-dienoate; **5**, benzoate; **6**, *cis*-2-hydroxypenta-2,4-dienoate; **7**, 4-hydroxy-2-oxovalerate; **8**, acetaldehyde; **9**, pyruvate; **10**, acetyl-CoA. BphA (EtbA, EbdA), biphenyl 2,3-dioxygenase; BphB, 2,3-dihydro-2,3-dihydroxybiphenyl dehydrogenase; BphC (EtbC), 2,3-dihydroxybiphenyl 1,2-dioxygenase; BphD (EtbD2), 2-hydroxy-6-oxophenylhexa-2,4-dienoate hydrolase; BphE, 2-hydroxypenta-2,4-dienoate hydratase; BphF, 4-hydroxy-2-oxovalerate aldolase; BphG, acetaldehyde dehydrogenase (acylating)

For the two other linear plasmids of strain RHA1, their role in catabolism, especially in the degradation of biphenyl/PCBs, has been established fairly well. Strain RHA1 besides biphenyl (Fig. 4) utilizes isopropylbenzene, ethylbenzene, and other monocyclic aromatic compounds (Gonçalves et al. 2006). Biphenyl, isopropylbenzene, and ethylbenzene degradation proceed via analogous reactions, i.e., initial aromatic ring dioxygenation, rearomatization of the *cis*-dihydrodiol to form a dihydroxy aromatic compound, and subsequent *meta* ring cleavage and C–C hydrolysis. Strain RHA1 is a prime example for illustrating the metabolic redundancy of rhodococci, as it contains multiple genes for various degradative steps, distributed on the linear plasmids pRHL1 and pRHL2 and the chromosome (Table 1 lists the catabolic genes identified on the plasmids). However, especially in the case of homologous proteins showing relatively low sequence identities, such multiple "isofunctional" enzymes may well have evolved distinct specificities for different sets of substrates (Gonçalves et al. 2006; McLeod et al. 2006). Even highly similar proteins may show subtle differences in their substrate preferences, as observed for EtbD1 (of pRHL1) and EtbD2 (of pRHL2), which share 97% sequence identity. Both enzymes have been characterized as 2-hydroxy-6-oxohepta-2,4-dienoate hydrolases with low activity toward 2-hydroxy-6-oxophenylhexa-2,4-dienoate (the substrate of BphD), but EtbD2 and EtbD1 convert the latter compound with 11 and 3.3% relative activity, respectively (Yamada et al. 1998). Notably, the expression of both *etbD1* and *etbD2* is upregulated by ethylbenzene as well as biphenyl (Gonçalves et al. 2006), suggesting a physiological role in both catabolic branches.

With respect to the initial step of biphenyl degradation, expression of all sets of ring hydroxylating dioxygenase genes (*bphA, etbA1, etbA2*) likewise is induced by biphenyl, and gene disruption experiments suggested that all these genes are involved in biphenyl degradation (Takeda et al. 2004; Iwasaki et al. 2006; Gonçalves et al. 2006). Two *cis*-dihydrodiol dehydrogenases BphB1 and BphB2, which are encoded on pRHL1 and pRHL2, respectively, catalyze the second step in biphenyl catabolism, i.e., the rearomatization of the *cis*-dihydrodiol to form 2,3-dihydroxybiphenyl. An abundant variety of genes exist for extradiol ring cleavage dioxygenases (*bphC* genes, *etbC* on pRHL2); the EtbC protein shows broad specificity toward 2,3-dihydroxybiphenyl and methylcatechols (Hauschild et al. 1996). The gene *bphC1* is found on pRHL1, whereas *bphC3, C5, C6, C8, C9, C10*, and *C11* are chromosomally encoded. In the fourth step of the biphenyl pathway, at least two hydrolases cleave 2-hydroxy-6-oxophenylhexa-2,4-dienoate (gene products of *bphD1* of pRHL1, *bphD2* located on the chromosome, and possibly *etbD2* of pRHL2). Overall, the transcriptome of strain RHA1 grown on biphenyl and ethylbenzene suggests that at least in this strain, the concept of redundancy not only refers to gene loci, but also to patterns of gene expression and synthesis of catabolic (iso)enzymes (www.rhodococcus.ca; Gonçalves et al. 2006). The cooperative involvement of multiple enzymes with distinct but overlapping substrate

specificities may be a crucial determinant for the superior capability of strain RHA1 to degrade a series of PCB congeners.

The downstream degradation of the 2-hydroxypenta-2,4-dienoate intermediate produced by *meta* ring cleavage and C–C hydrolysis in biphenyl and ethylbenzene degradation (Fig. 4) is catalyzed by gene products of the *bphE, bphF*, and *bphG* genes located on pRHL1 and pRHL2 (Table 1) (Sakai et al. 2003; Gonçalves et al. 2006). In contrast, the *benABCDK* cluster for the degradation of benzoate, the other product of BphD-catalyzed hydrolysis of 2-hydroxy-6-oxophenylhexa-2,4-dienoate (Fig. 4), are located on the chromosome (Kitagawa et al. 2001; www.rhodococcus.ca). Benzoate is degraded via catechol; degradation of the latter likewise is encoded on the chromosome (*cat* gene cluster). Expression of the *ben* and *cat* genes is induced by biphenyl and benzoate (Gonçalves et al. 2006).

Strain RHA1 utilizes phthalate via protocatechuate; however, the proteins involved in protocatechuate degradation are encoded by the chromosomal *pca* gene cluster, whereas phthalate conversion is encoded on the two linear plasmids pRHL1 and pRHL2. Two nearly identical *pad-pat-tpa* gene clusters comprising phthalate degradation genes (*pad*), genes predicted to be involved in uptake of phthalate (*pat*), and presumed terephthalate catabolic genes (*tpa*) are located on both replicons (Table 1). Actually, the *pad-pat-tpa* gene cluster is part of a 32.1-kb duplication which contains 34 putative ORFs, including genes presumed to code for transposases and an integrase (Patrauchan et al. 2005).

In conclusion, the current genomic, transcriptomic, and proteomic data (Patrauchan et al. 2005; Gonçalves et al. 2006; McLeod et al. 2006; www.rhodococcus.ca) suggest that degradation of aromatic compounds by strain RHL1 involves multiple peripheral pathways which are partly, but not exclusively, encoded on its linear plasmids. The intermediates generated are subsequently funneled into a limited set of central (usually chromosomally encoded) pathways that produce TCA cycle intermediates.

3.2.2
Megaplasmids pDK2 and pRHL2, as well as pDK3 and pRHL1, Share Almost Identical Catabolic Gene Clusters

Rhodococcus sp. DK17 was isolated by Kim et al. (2002) by enrichment on *o*-xylene as the sole source of carbon and energy, but it is also able to grow on ethylbenzene, isopropylbenzene, toluene, benzene, phthalate, and terephthalate. Three megaplasmids pDK1, pDK2, and pDK3 were identified in this strain (Table 1). Genes for alkylbenzene dioxygenases (AkbA) *cis*-dihydrodiol dehydrogenase (AkbB), *meta*-cleavage enzyme (AkbC), and downstream catabolic enzymes (AkbDEF) are located on the 330-kb plasmid pDK2 (Kim et al. 2004). Gene organization as well as nucleotide sequences of the *akb* regions are identical to corresponding regions of pRHL2 of *Rhodococ-*

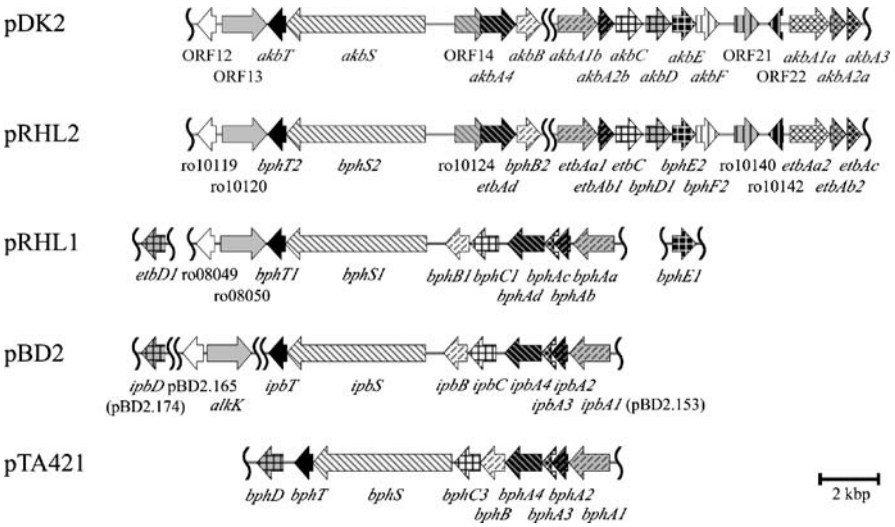


Fig. 5 Organization of genes involved in biphenyl and alkylbenzene degradation. pDK2, *akb* genes located on the linear plasmid pDK2 of *Rhodococcus* sp. DK17 (Kim et al. 2004; AY502075); pRHL2, *bph* genes on pRHL2 of *Rhodococcus* sp. RHA1 (NC_008270); pRHL1, *bph* genes on pRHL1 of strain RHA1 (NC_008296); pBD2, *ipb* genes of linear plasmid pBD2 of *R. erythropolis* BD2 (Stecker et al. 2003; NC_005073.1); pTA421, *bphC3* locus on pTA421 of *R. erythropolis* TA421 (Arai et al. 1998; accession numbers AB014348, D88015, and D88020). The *akbA/etbA/bphA/ipbA* genes code for the oxygenase (A1A2/AaAb), ferredoxin (A3/Ac), and ferredoxin reductase (A4/Ad) components of ring-hydroxylating dioxygenases; *akbB/bphB/ipbB*, *akbC/etbC/bphC*, and *akbD/bphD/ipbD* encode *cis*-dihydrodiol dehydrogenases, *meta*-cleavage dioxygenase, and C–C hydrolases, respectively; *akbE/bphE* and *akbF/bphF* encode the downstream hydratase and aldolase. Genes *akbST/bphST/ipbST* code for putative sensor kinase and response regulator of a two-component signal transduction system. ORFs 12 and 13 of pDK2 and their homologues on pRHL2, pRHL1, and pBD2 may encode an enoyl-CoA hydratase and an acyl-CoA ligase (*alkK*), respectively. The gene products of ro10124 and ro10142 and their homologues on pDK2 are conserved hypothetical proteins; ro10140 of pRHL2 and ORF21 of pDK2 may code for an oxidoreductase (www.rhodococcus.ca; Kim et al. 2004). The *arrows* indicate the size, location, and direction of transcription of the genes and ORFs. Homologous ORFs feature the same pattern. Truncated regions are marked with the symbol ⟆

cus sp. strain RHA1 (Fig. 5) (Kim et al. 2004). In fact, pDK2 shares nearly complete nucleotide sequence identity with pRHL2 of *Rhodococcus* sp. RHA1, except for a 0.9-kb insertion in pRHL2, indicating recent horizontal transfer (Gonçalves et al. 2006).

Plasmid pDK2 and the 750-kb plasmid pDK3 each contain a copy of *oph* and *tph* operons coding for the transformation of phthalate and terephthalate, respectively, to protocatechuate. These operons are separated by a region containing a putative phthalate ester hydrolase gene and putative trans-

porter genes, which is organized like the *patEBCAD* cluster of strain RHA1 (Choi et al. 2005). Plasmids pDK3 of *Rhodococcus* sp. DK17 and pRHL1 from *Rhodococcus* sp. RHA1 actually share two regions of 99% sequence identity, namely, a 73-kb region that includes the genes encoding phthalate and terephthalate metabolism, and another 22-kb segment (Patrauchan et al. 2005).

3.2.3 Multiple Extradiol Dioxygenase Genes Are Distributed on the Genomes of the Biphenyl Degrading Strains *Rhodococcus erythropolis* TA421 and *Rhodococcus globerulus* P6

Growth of *R. erythropolis* TA421 on biphenyl and cometabolic turnover of PCB is associated with the linear 500-kb plasmid pTA421 (Kosono et al. 1997). Similar to *Rhodococcus* sp. strain RHA1 (Sect. 3.2.1), strain TA421 contains a multiplicity of *bphC* genes coding for 2,3-dihydroxybiphenyl 1,2-dioxygenases. Three of the seven *bphC* genes are located on its linear plasmid pTA421 (Table 1; Kosono et al. 1997; Arai et al. 1998). The plasmid-borne *bphC3* is part of the *bph* locus depicted in Fig. 5; its deduced gene product shows 70% identity to BphC1 of pRHL1 and IpbC encoded by the linear plasmid pBD2 (see Sect. 3.2.4). The *bphC4* gene was identified in a *bphA1A2A3BA4C4D* cluster which is also located on the plasmid, suggesting that the extradiol dioxygenases BphC3 and/or BphC4 indeed are involved in biphenyl degradation (Arai et al. 1998). The BphC2 and BphC3 proteins of strain TA421 share 95.3 and 99.7% identity to BphC2 and BphC1, respectively, of *R. globerulus* P6 (Arai et al. 1998), suggesting a recent common ancestor of the corresponding genes. Strain P6 harbors two probably linear plasmids pLP6 and pSP6 (Table 1); however, whilst pSP6 contains the *bphC3* and *bphC4* homologues, a *bphC2* gene is located on both plasmids (Kosono et al. 1997). Eight *bphC* genes were identified in *R. rhodochrous* K37. Two of those (*bphC6*, *bphC8*) are located on a 200-kb plasmid (Table 1), whereas the remaining six appear to be located on the chromosome (Taguchi et al. 2004).

3.2.4 Alkylbenzene Degradation and Co-oxidation of Trichloroethene by *Rhodococcus erythropolis* BD2 is Encoded on pBD2

The complete sequence of the linear conjugative 210,205-bp plasmid pBD2, which confers to *R. erythropolis* strain BD2 the ability to oxidize isopropylbenzene and to cometabolize trichloroethene (Dabrock et al. 1994; Kesseler et al. 1996), has been reported (Stecker et al. 2003). Plasmid pBD2 additionally mediates arsenite and mercury resistance (Dabrock et al. 1994). The gene organization on pBD2 and the high number of transposon-related hy-

pothetical genes suggest a modular composition of the plasmid that presumably arose from horizontal gene transfer events and dynamic rearrangements. The *ipb* genes coding for the components of isopropylbenzene 2,3-dioxygenase (IpbA), *cis*-dihydrodiol dehydrogenase (IpbB), and the *meta* ring cleavage enzyme 3-isopropylcatechol 2,3-dioxygenase (IpbC), and adjacent genes of a two-component signal transduction system (IpbST), are flanked by putative transposase genes and preceded by an IS element. Figure 5 indicates that apart from an insertion between *ipbT* and *alkK* (pBD2.164), the gene organization of the *ipb* locus of pBD2 is very similar to that of the corresponding *bph* locus on pRHL1 of *Rhodococcus* sp. RHA1. Moreover, isopropyl 2,3-dioxygenase and 3-isopropylcatechol 2,3-dioxygenase exhibit 94–100% identity to the gene products of the *bphA* and *bphC1* genes located on pRHL1 (Shimizu et al. 2001; Stecker et al. 2003). The putative genes pBD2.165 and pBD2.164 (*alkK*) on pBD2, presumed to encode an enoyl-CoA hydratase and a medium-chain acyl-CoA ligase, show 98 and 99% identity to ro08049 and ro08050 of pRHL1, respectively. A gene designated *ipbD*, presumed to code for 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase, is located 10 kb downstream of the other *ipb* genes; this gene is also flanked by putative transposase genes and IS sequences. Its deduced product is identical to the EtbD1 protein encoded on pRHL1 of strain RHA1 (Stecker et al. 2003).

3.2.5
Plasmid p1CP of *Rhodococcus opacus* 1CP Codes for Two Distinct Chlorocatechol Pathways

The 740-kb linear plasmid p1CP of *R. opacus* 1CP harbors the genes of the 4-chloro-/3,5-dichlorocatechol degradation pathway (Fig. 6), another set of genes coding for the 3-chlorocatechol pathway (Fig. 6), and at least one maleylacetate reductase gene (Table 1). In contrast, the genes coding for the catechol and protocatechuate branch of the 3-oxoadipate pathway for the degradation of non-chlorinated aromatic compounds are not located on p1CP (König et al. 2004), which is in accordance with the chromosomal location of these genes in most other bacteria. The *clcA2D2B2F* cluster of the 3-chlorocatechol branch is unique in that it comprises the *clcF* gene encoding a 5-chloromuconate dehalogenase, which catalyzes formation of the *cis*-dienelactone (Moiseeva et al. 2002) (Fig. 6). Despite a significant sequence similarity of ClcF to muconolactone isomerases, activity toward muconolactone was not detected, suggesting evolutionary adaptation of the enzyme to the dechlorination reaction. The ClcA2, B2, and D2 proteins of the 3-chlorocatechol pathway share only 40–51% identity with the corresponding Clc proteins of the 4-chloro-/3,5-dichlorocatechol pathway, and thus were suggested to represent independent evolutionary branches (Moiseeva et al. 2002).

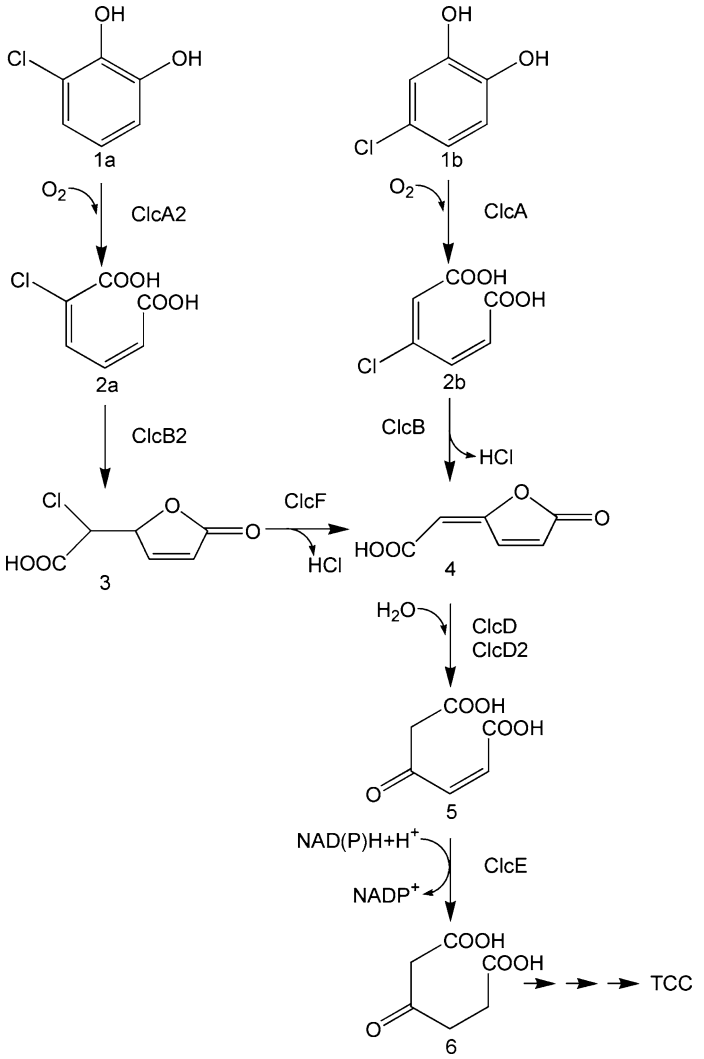


Fig. 6 Pathway of 3-chlorocatechol and 4-chlorocatechol degradation by *R. opacus* 1CP (Moiseeva et al. 2002). **1a**, 3-Chlorocatechol; **1b**, 4-chlorocatechol; **2a**, 2-chloro-*cis,cis*-muconate; **2b**, 3-chloro-*cis,cis*-muconate; **3**, 5-chloromuconolactone; **4**, *cis*-dienelactone (*cis*-4-carboxymethylenebut-2-en-4-olide); **5**, maleylacetate; **6**, 3-oxoadipate. ClcA, ClcA2, chlorocatechol 1,2-dioxygenase; ClcB, ClcB2, chloromuconate cycloisomerase; ClcF, 5-chloromuconolactone dehalogenase; ClcD, ClcD2, dienelactone hydrolase; ClcE, maleylacetate reductase. TCC, tricarboxylic acid cycle

3.3
Linear Plasmids Involved in the Degradation of Polycyclic Aromatic Compounds by *Rhodococcus* sp. Strains

3.3.1
Modular Organization and Different Genomic Localization of Naphthalene Degradation Genes in Strains I24, P200, P400, NCIMB12038, CIR2, TKN14, R7, and SAO101

Rhodococcus sp. I24 is able to utilize naphthalene and toluene as carbon sources, and to co-oxidize indene. The naphthalene inducible genes *nidABCD* (Fig. 7) coding for naphthalene dioxygenase, a dehydrogenase, and a putative aldolase (Treadway et al. 1999) reside on an 80-kb plasmid (AF45237.2; Priefert et al. 2004) (Table 1). Common initial steps in naphthalene degradation by *Rhodococcus* sp. strains NCIMB12038, P200, and P400 likewise involve formation of the *cis*-naphthalene dihydrodiol, catalyzed by naphthalene dioxygenase (NarA), and re-aromatization to dihydroxynaphthalene by *cis*-naphthalene dihydrodiol dehydrogenase (NarB). The *nar* gene clusters have different genomic locations. In *Rhodococcus* sp. NCIMB12038, it is located on the linear plasmid p2SL1 (Table 1). Strain P200 contains a linear 300-kb plasmid termed P20L1 and a circular 150-kb plasmid; however, the *nar* genes appear to be chromosomally located. Strain P400 contains two linear and two circular plasmids (Table 1); copies of the *nar* genes were identified on the 180-kb circular plasmid P40C1 as well as on a smaller linear plasmid (P40L2 or P40L3) (Kulakov et al. 2005). Significant similarity exists between the *nar/nid/rno* genes of strains NCIMB12038, P200, P400, I24, and *Rhocococcus* sp. CIR2. The products of the *narA* and *narB* genes of P200 and P400, for example, show 88–99% identity to the corresponding proteins of NCIMB12038. However, the presence of intervening nonhomologous regions and possible deletions or insertions suggests that *nar* genes have been arranged in several "modules". The central module consists of the *narAaAbB/nidABC/rno(A2A3)A4B* genes, which are present in the same order in all strains (Fig. 7). Putative genes coding for a hydratase-aldolase (*narC*, *nidD*) located downstream of *narAaAbB/nidABC* were identified in strains I24, P200, and P400, whereas NCIMB12038 and CIR2 lack *narC* (but contain a short remnant of its 5′ region).

The naphthalene degradation genes of the linear plasmid p2SL1 of strain NCIMB12038 have highly homologous counterparts in *R. opacus* strain R7 (DQ846881); however, the *rub* and *narR* genes are arranged differently (Fig. 7), and the location of the gene cluster in the genome of strain R7 is not clear. *R. opacus* strain TKN14, which is able to degrade *o*-xylene, contains *nid* genes that are arranged like the *rub1* and *nar* genes of strains P200 and P400, and the corresponding genes of strain I24 (Fig. 7) (Maruyama et al. 2005). However, it is not yet known whether the *nid* locus of strain TKN14 is part of a plasmid or the chromosome.

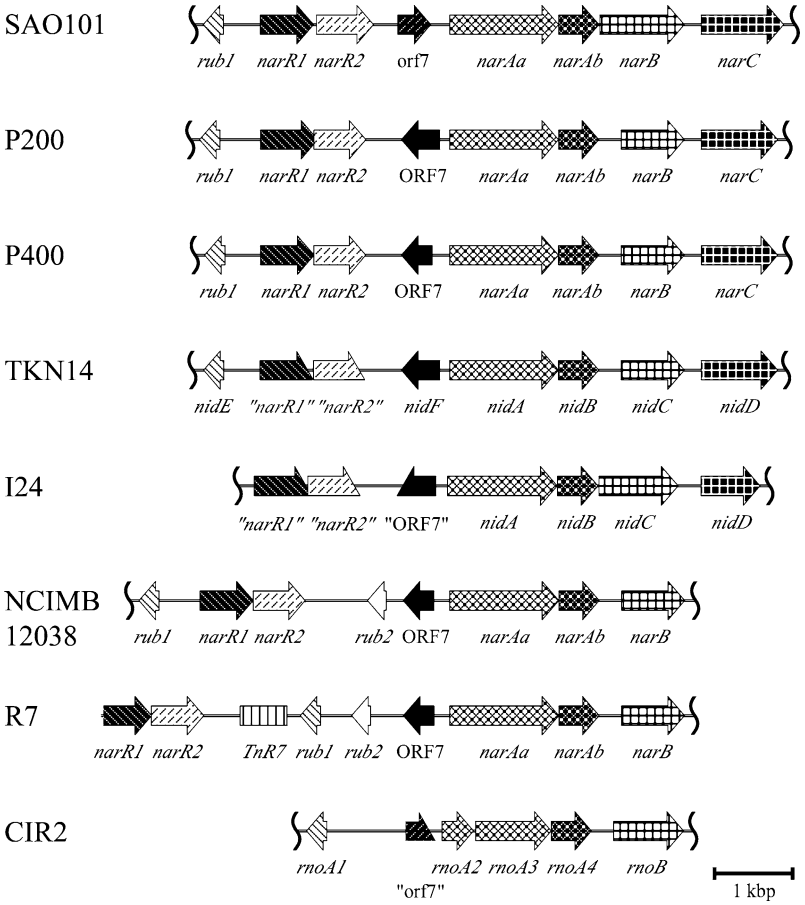


Fig. 7 Organization of gene clusters involved in naphthalene degradation. SAO101, *nar* locus on linear plasmid pWK301 of *R. opacus* strain SAO101 (Kimura et al. 2006; AB110633); P200, chromosomally located *nar* genes of *Rhodococcus* sp. P200 (Kulakov et al. 2005; AY392424); P400, homologues on the circular plasmid P40C1 of *Rhodococcus* sp. strain P400 (Kulakov et al. 2005; AY392423.2); TKN14, *nid* and putative *narR* genes of unknown location from *Rhodococcus* sp. TKN14 (Maruyama et al. 2005; AB206671); I24, homologues located on the 80-kb plasmid of *Rhodococcus* sp. strain I24 (Treadway et al. 1999; Priefert et al. 2004; AF121905); NCIMB12038, *nar* locus of the linear plasmid p2SL1 of *Rhodococcus* sp. NCIMB12038 (Kulakov et al. 2005; AF082663.3); R7, *nar* cluster (unknown location) of *Rhodococcus* sp. strain R7 (DQ846881); CIR2, *rno* genes (unknown location) of *Rhodococcus* sp. CIR2 (Kulakov et al. 2005; AB024936). The *narAaAb/nidAB/rnoA2–4* genes code for naphthalene dioxygenase; *narB/nidC/rnoB* code for *cis*-dihydrodiol dehydrogenases; *narC* and *nidD* encode a hydratase/aldolase; *rub1/nidE/rnoA1* and *rub2* are putative rubredoxin genes; *narR1* and *narR2* genes encode putative regulators of the GntR and NtrC family, respectively (Kulakov et al. 2005). The *arrows* indicate the size, location, and direction of transcription of the ORFs; *arrows with semi-arrowheads* and designations in *quotation marks* indicate potential ORFs. Homologous ORFs feature the same pattern. Truncated regions are marked with the symbol ⟆

R. opacus SAO101 is able to utilize dibenzo-*p*-dioxin and dibenzofuran as sole sources of carbon and energy, and also grows on other aromatic compounds like naphthalene, biphenyl, dibenzothiophene, toluene, phenol, (chloro)benzene, and *p*-nitrophenol (Kimura and Urushigawa 2001; Kitagawa et al. 2004). This organism contains three large linear plasmids pWK301, pWK302, and pWK303 (Table 1) (Kitagawa et al. 2004). Oxidation of dibenzofuran by strain SAO101 proceeds via lateral dioxygenation at the 1,2- and 3,4-positions to form the corresponding *cis*-dihydrodiols (Fig. 8). This ability appears to be due to the relaxed substrate specificity of the oxygenase component (NarA) of its ring-hydroxylating dioxygenase, which catalyzes the oxygenation of naphthalene, dibenzofuran, dibenzo-*p*-dioxin, phenanthrene, and biphenyl. The *nar* cluster (Fig. 7) is located on the 1100-bp linear plasmid pWK301 and is induced in the presence of dibenzofuran (Kimura et al.

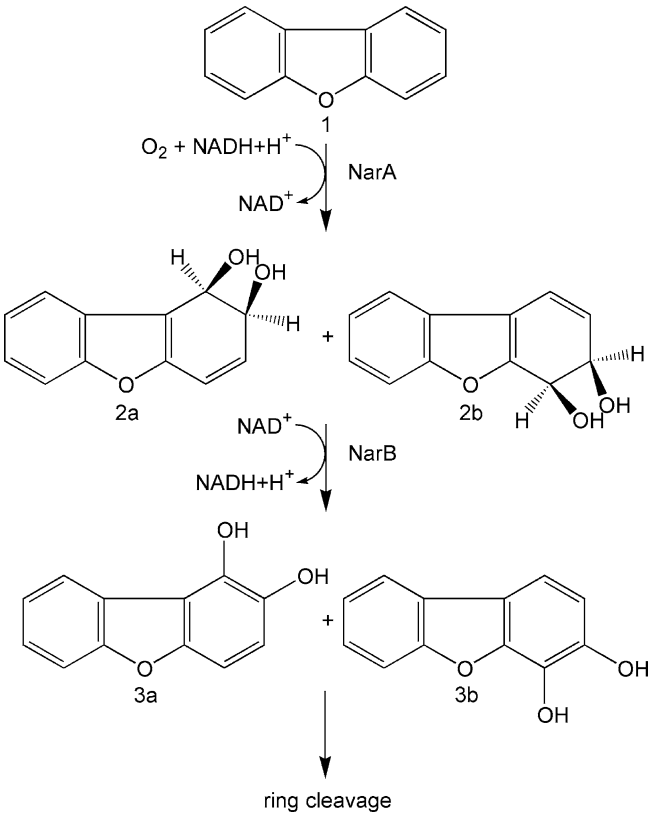


Fig. 8 Initial oxidation of dibenzofuran by *R. opacus* SAO1 (Kimura et al. 2006). **1**, Dibenzofuran; **2a**, *cis*-1,2-dihydro-1,2-dihydroxydibenzofuran; **2b**, *cis*-3,4-dihydro-3,4-dihydroxydibenzofuran; **3a**, 1,2-dihydroxydibenzofuran; **3b**, 3,4-dihydroxydibenzofuran. NarA, ring-hydroxylating dioxygenase; NarB, dihydrodiol dehydrogenase

2006). The deduced NarAaAb proteins show 97–98% sequence identity to the naphthalene-inducible dioxygenase NidAB of *Rhodococcus* sp. strain I24. The gene organization of the catabolic gene cluster of strain SAO101 closely resembles that of gene clusters of *Rhodococcus* sp. P200, P400, TKN14, and I24 (Fig. 7).

R. opacus M213, another naphthalene degrader, was proposed to contain two plasmids pNUO1 and pNUO2 (Table 1). The linear plasmid pNUO1 carries a catechol 2,3-dioxygenase gene *edoD*; however, the role of the pNUO plasmids in naphthalene degradation is at present not known (Uz et al. 2000).

The patterns of homologous regions, and some differences in localization and organization of the genes depicted in Fig. 7, suggest that genetic exchange and genomic rearrangements had occurred within the *nar* gene clusters. The presence of deletions, intervening nonhomologous regions, and (partial) transposase sequences within the *nar* regions is also indicative of DNA shuffling during evolution of the *nar* regions (Kulakov et al. 2005; Kimura et al. 2006).

4
Terrabacter sp.

The genus *Terrabacter* comprises several strains with the ability to oxidize complex aromatic compounds such as dibenzofuran, fluorene, dibenzo-*p*-dioxin, biphenyl, and 1,1-dichloro-2,2-bis(4-chlorophenyl)ethylene (DDE) (Aislabie et al. 1999; Monna et al. 1993; Iida et al. 2002a, Noumura et al. 2004; Zhou et al. 2006). Plasmid-borne genes were shown to be involved in the degradation of aromatic compounds by *Terrabacter* sp. strains YK3 and DBF63; however, the plasmid of strain YK3 is circular (Iida et al. 2002a), whereas strain DBF63 carries two linear plasmids (Nojiri et al. 2002).

4.1
Dibenzofuran and Fluorene Oxidation via Angular Dioxygenation as well as the Protocatechuate Pathway are Encoded on pDBF1 of *Terrabacter* sp. DBF63

Terrabacter sp. strain DBF63 is able to utilize both dibenzofuran and fluorene as sole source of carbon and energy (Monna et al. 1993) and to cometabolize a broad range of chlorinated dibenzo-*p*-dioxins and dibenzofurans (Habe et al. 2001, 2002). Dibenzofuran degradation is initiated by angular dioxygenation, catalyzed by the oxygenase DbfA (with unknown electron transport proteins). The chemically unstable dihydrodiol formed in the enzymatic reaction is spontaneously rearomatized to 2,2′,3-trihydroxybiphenyl, which undergoes DbfB-catalyzed *meta* cleavage. Hydrolysis of the *meta* cleavage

product by DbfC protein yields salicylate, which is metabolized via gentisate or catechol. Since the *dbfA1A2* genes are located on the 160-kb linear plasmid pDBF1 (as well as on its putative derivative plasmid pDBF2; Table 1), whilst the *dbfBC* genes are located on the chromosome, mineralization of dibenzofuran by strain DBF63 requires the presence of both replicons (Nojiri et al. 2002). Remarkably, genes showing more than 98% identity to *dbfA1A2* of *Terrabacter* sp. DBF63 were found in *Rhodococcus* sp. strain YK2 (Iida et al. 2002b), suggesting horizontal DNA transfer between genera.

Plasmid pDBF1 is not only involved in mineralization of dibenzofuran, but also carries all the genes necessary for the degradation of fluorene to intermediates of the tricarboxylic acid cycle (Habe et al. 2005; Table 1). Oxidation of fluorene (Fig. 9) is also initiated by the angular dioxygenase DbfA, which in a monooxygenation reaction yields 9-fluorenol. Subsequent oxidation by the dehydrogenase FlnB results in 9-fluorenone, which is attacked again by DbfA to form 1,1a-dihydroxy-1-hydrofluorene-9-one in an angular dioxygenation (Habe et al. 2004). The latter compound is oxidized by FlnB to 2′-carboxy-2,3-dihydroxybiphenyl, which undergoes *meta* cleavage and hydrolysis; however, chromosomally encoded DbfB and DbfC are not involved in these reactions (Nojiri et al. 2002; Habe et al. 2004). Instead, the heterodimeric extradiol dioxygenase FlnD (the product of *flnD1* and ORF16; Table 1) is responsible for *meta* ring cleavage of 2′-carboxy-2,3-dihydroxybiphenyl, and subsequent hydrolysis of the *meta* cleavage product to form phthalate is catalyzed by FlnE (Habe et al. 2004). Phthalate conversion to protocatechuate is encoded by the *pht* gene cluster, which is located adjacent to the *fln* genes. The pDBF1 plasmid also harbors the *pca* operon coding for protocatechuate catabolism to succinyl-CoA and acetyl-CoA (Habe et al. 2005). Such plasmid localization of 3-oxoadipate pathway genes is remarkable, since it is generally assumed that catabolic plasmids usually encode peripheral pathways, whereas genes of central pathways reside on the chromosome. As mentioned above, the *pca* genes of *R. opacus* 1CP and *Rhodococcus* sp. strain RHA1, for example, are indeed clustered chromosomally (König et al. 2004; Patrauchan et al. 2005); in *Streptomyces coelicolor* A3(2) they are located on the central, stable region of the chromosome (Iwagami et al. 2000).

Two other dibenzofuran degrading actinobacteria, namely *Terrabacter* sp. DFA1 and *Janibacter* sp. DFA10, show the same organization of the *pht* and *dbf-fln* gene clusters as strain DBF63; however, their localization on distinct replicons has not been investigated yet (Noumura et al. 2004). Since several putative transposase genes or IS-like elements were found around the *pca* and *dbf-fln* gene clusters in strain DBF63, Habe et al. (2005) suggested that the *pca* genes were acquired via transposition events and subsequently distributed via horizontal plasmid transfer.

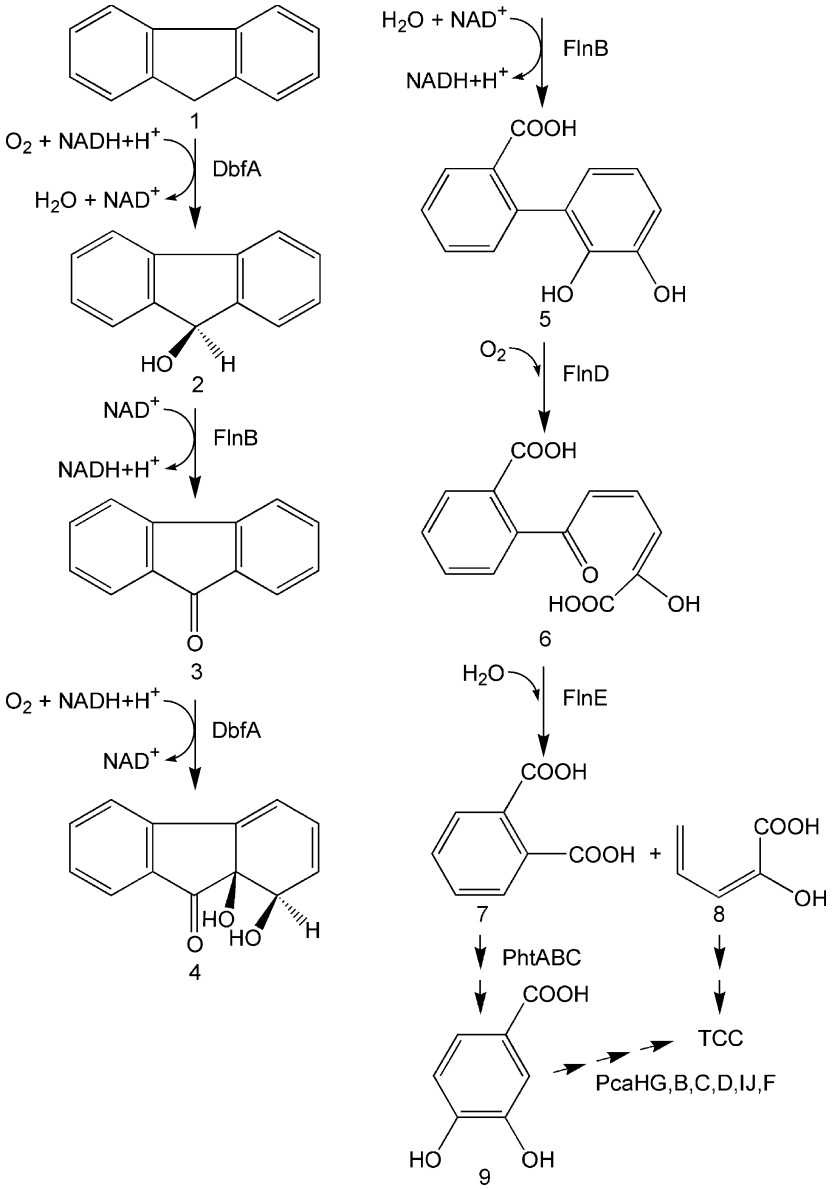


Fig. 9 Degradation of fluorene by *Terrabacter* sp. DBF63 (Habe et al. 2004, 2005). **1**, Fluorene; **2**, 9-fluorenol; **3**, 9-fluorenone; **4**, 1,1a-dihydroxy-1-hydrofluorene-9-one; **5**, 2′-carboxy-2,3-dihydroxybiphenyl; **6**, 2′-carboxy-2-hydroxy-6-oxophenylhexa-2,4-dienoate; **7**, phthalate; **8**, *cis*-2-hydroxypenta-2,4-dienoate; **9**, protocatechuate. DbfA, angular dioxygenase; FlnB, 1,1a-dihydroxy-1-hydro-9-fluorenone dehydrogenase; FlnD, 2′-carboxy-2,3-dihydroxybiphenyl 1,2-dioxygenase; FlnE, 2′-carboxy-2-hydroxy-6-oxophenylhexa-2,4-dienoate hydrolase; PhtABC, enzymes of phthalate conversion to protocatechuate; PcaHG, B, C, D, IJ, F, enzymes of the 3-oxoadipate pathway. TCC, tricarboxylic acid cycle

5
Arthrobacter sp.

Bacteria of the genus *Arthrobacter* are considered to be ubiquitous in soil. They are highly resistant to desiccation and nutrient depletion, and utilize a wide and varied range of natural as well as xenobiotic compounds (Cacciari and Lippi 1987). A number of catabolic traits of *Arthrobacter* strains are encoded on plasmids. A well-known example is the circular, 165-kb plasmid pAO1 of *A. nicotinovorans*, which contains the genes coding for L-nicotine catabolism (Brandsch and Decker 1984; Igloi and Brandsch 2003). Other (presumably circular) catabolic plasmids of different *Arthrobacter* spp. code for the degradation of 2-nitroaromatic compounds (Chauhan et al. 2000; Chauhan and Jain 2000), phthalate (Eaton 2001), phenylurea herbicides (Turnbull et al. 2001), carbaryl and 1-naphthol (Hayatsu et al. 1999), *S*-ethyl-*N*,*N*-dipropylthiocarbamate (Tam et al. 1987), pyridine, dimethylpyridine and hydroxypyridine (Weinberger and Kolenbrander 1979; Agapova et al. 1992), and atrazine (Sajjaphan et al. 2004).

5.1
Linear Plasmid pAL1 of *Arthrobacter nitroguajacolicus* Rü61a Codes for Degradation of 2-Methylquinoline

Arthrobacter nitroguajacolicus Rü61a is able to utilize the *N*-heteroaromatic compound quinaldine (2-methylquinoline) as sole source of carbon and energy (Hund et al. 1990; for a review, see Fetzner 1998b). Quinaldine degradation (Fig. 10) starts with a hydroxylation in position 4 to form 1*H*-4-oxoquinaldine, catalyzed by the molybdenum enzyme quinaldine 4-oxidase (Qox) (Parschat et al. 2003). A monooxygenase-catalyzed hydroxylation at C-3 subsequently produces 1*H*-3-hydroxy-4-oxoquinaldine, which undergoes an unusual 2,4-dioxygenolytic ring cleavage reaction, yielding carbon monoxide and *N*-acetylanthranilate (Frerichs-Deeken et al. 2004; Frerichs-Deeken and Fetzner 2005). *N*-acetylanthranilate is hydrolyzed to anthranilate and acetate by an aryl-acylamidase (Amq) (Kolkenbrock et al. 2006). Note that the genes coding for the enzymes of this pathway module are not related to genes involved in the activation and ring cleavage of aromatic compounds (Parschat et al. 2003).

The genes required for quinaldine conversion to anthranilate are clustered on the linear plasmid pAL1, which shows the typical invertron structure, with proteins attached to its 5′ ends and terminal inverted repeats (Overhage et al. 2005; Parschat et al. 2007). A mutant of strain Rü61a that had lost pAL1 could not convert quinaldine, but was still able to grow on anthranilate, suggesting that the chromosome contains genes for the downstream metabolism. Anthranilate degradation has been assumed to proceed via catechol and the well-known *ortho* cleavage pathway (Hund et al. 1990), which

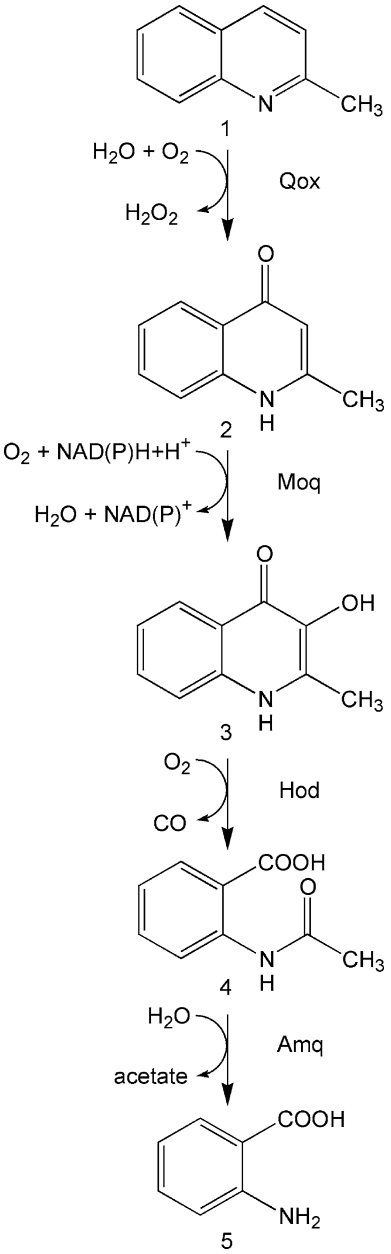


Fig. 10 Quinaldine degradation to anthranilate by *A. nitroguajacolicus* Rü61a (Hund et al. 1990; Fetzner 1998b; Overhage et al. 2005; Parschat et al. 2007). **1**, Quinaldine (2-methylquinoline); **2**, 1*H*-4-oxoquinaldine; **3**, 1*H*-3-hydroxy-4-oxoquinaldine; **4**, *N*-acetylanthranilic acid; **5**, anthranilic acid. Qox, quinaldine 4-oxidase; Moq, 1*H*-4-oxoquinaldine 3-monooxygenase; Hod, 1*H*-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase; Amq, *N*-acetylanthranilate amide hydrolase

probably is chromosomally encoded. Five other quinaldine degrading strains, which all turned out to belong to the genus *Arthrobacter*, were isolated from soil samples by enrichment on quinaldine. Four out of these five isolates contained a pAL1-like linear plasmid carrying genes coding for quinaldine conversion, suggesting horizontal dissemination of this replicon among the genus *Arthrobacter* (Overhage et al. 2005).

6 Concluding Remarks

The comparison of linear catabolic plasmids sequenced so far with related replicons suggests that many linear plasmids are mosaic-like structures that may have resulted from a series of insertions, deletions, transpositions, and recombination events. The observation that putative IS elements or IS-related genetic structures are often associated with catabolic gene clusters supports this hypothesis. Very similar catabolic gene clusters, and thus metabolic pathways, are often shared by phylogenetically different bacteria, suggesting that mobile genetic elements (transferable plasmids, transposons, "genomic islands") have been distributed by horizontal DNA transfer. Horizontal gene transfer likely is a major factor promoting the dispersal of catabolic capabilities and the evolution of bacterial degradation pathways.

Acknowledgements The financial support of our work on linear *Arthrobacter* plasmid pAL1 by the Deutsche Forschungsgemeinschaft (FE 383/11-1) is gratefully acknowledged.

References

Agapova SR, Andreeva AL, Starovoitov II, Vorob'eva LI, Terent'ev PB (1992) Plasmids for biodegradation of 2,6-dimethylpyridine, 2,4-dimethylpyridine, and pyridine in strains of *Arthrobacter*. Mol Gen Mikrobiol Virusol 1992:10–13

Aislabie J, Davison AD, Boul HL, Franzmann PD, Jardine DR, Karuso P (1999) Isolation of *Terrabacter* sp. strain DDE-1, which metabolizes 1,1-dichloro-2,2-bis(4-chlorophenyl)ethylene when induced with biphenyl. Appl Environ Microbiol 65:5607–5611

Arai H, Kosono S, Taguchi K, Maeda M, Song E, Fuji F, Chung SY, Kudo T (1998) Two sets of biphenyl and PCB degradation genes on a linear plasmid in *Rhodococcus erythropolis* TA421. J Ferment Bioeng 86:595–599

Bao K, Cohen SN (2003) Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. Genes Dev 17:774–785

Bergeron H, Labbé D, Turmel C, Lau PCK (1998) Cloning, sequence and expression of a linear plasmid-based and a chromosomal homolog of chloroacetaldehyde dehydrogenase-encoding genes in *Xanthobacter autotrophicus* GJ10. Gene 207:9–18

Brandsch R, Decker K (1984) Isolation and partial characterization of plasmid DNA from *Arthrobacter oxidans*. Arch Microbiol 138:15–17

Cacciari I, Lippi D (1987) Arthrobacters: successful arid soil bacteria. A review. Arid Soil Res Rehabil 1:1–30

Chauhan A, Jain RK (2000) Degradation of *o*-nitrobenzoate via anthranilic acid (*o*-aminobenzoate) by *Arthrobacter protophormiae*: a plasmid-encoded new pathway. Biochem Biophys Res Commun 267:236–244

Chauhan A, Chakraborti AK, Jain RK (2000) Plasmid-encoded degradation of *p*-nitrophenol and 4-nitrocatechol by *Arthrobacter protophormiae*. Biochem Biophys Res Commun 270:733–740

Chen CW (2007) *Streptomyces* Linear Plasmids: Replication and Telomeres. Microbiol Monogr 7

Choi KY, Kim D, Sul WJ, Chae JC, Zylstra GJ, Kim YM, Kim E (2005) Molecular and biochemical analysis of phthalate and terephthalate degradation by *Rhodococcus* sp. strain DK17. FEMS Microbiol Lett 252:207–213

Coleman NV, Spain JC (2003a) Epoxyalkane:coenzyme M transferase in the ethene and vinyl chloride biodegradation pathways of *Mycobacterium* strain JS60. J Bacteriol 185:5536–5545

Coleman NV, Spain JC (2003b) Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading *Mycobacterium* strains. Appl Environ Microbiol 69:6041–6046

Dabrock B, Keßeler M, Averhoff B, Gottschalk G (1994) Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism. Appl Environ Microbiol 60:853–860

Danko AS, Luo M, Bagwell CE, Brigmon RL, Freedman DL (2004) Involvement of linear plasmids in aerobic biodegradation of vinyl chloride. Appl Environ Microbiol 70:6092–6097

Danko AS, Saski CA, Tomkins JP, Freedman DL (2006) Involvement of coenzyme M during aerobic biodegradation of vinyl chloride and ethene by *Pseudomonas putida* strain AJ and *Ochrobactrum* sp. strain TD. Appl Environ Microbiol 72:3756–3758

Eaton RW (2001) Plasmid-encoded phthalate catabolic pathway in *Arthrobacter keyseri* 12B. J Bacteriol 183:3689–3703

Ensign SA, Allen JR (2003) Aliphatic epoxide carboxylation. Annu Rev Biochem 72:55–76

Fetzner S (1998a) Bacterial dehalogenation. Appl Microbiol Biotechnol 50:633–657

Fetzner S (1998b) Bacterial degradation of pyridine, indole, quinoline, and their derivatives under different redox conditions. Appl Microbiol Biotechnol 49:237–250

Francis I, Gevers D, Karimi M, Holsters M, Vereecke D (2007) Linear Plasmids and Phytopathogenicity. Microbiol Monogr 7

Frerichs-Deeken U, Fetzner S (2005) Dioxygenases without requirement for cofactors: identification of amino acid residues involved in substrate binding and catalysis, and testing for rate-limiting steps in the reaction of 1*H*-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase. Curr Microbiol 51:344–352

Frerichs-Deeken U, Ranguelova K, Kappl R, Hüttermann J, Fetzner S (2004) Dioxygenases without requirement for cofactors and their chemical model reaction: compulsory order ternary complex mechanism of 1*H*-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase involving general base catalysis by histidine 251 and single-electron oxidation of the substrate dianion. Biochemistry 43:14 485–14 499

Goethals K, Vereecke D, Jaziri M, van Montagu M, Holsters M (2001) Leafy gall formation by *Rhodococcus fascians*. Annu Rev Phytopathol 39:27–52

Gonçalves ER, Hara H, Miyazawa D, Davies JE, Eltis LD, Mohn WW (2006) Transcriptomic assessment of isozymes in the biphenyl pathway of *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 72:6183–6193

Habe H, Chung JS, Lee JH, Kasuga K, Yoshida T, Nojiri H, Omori T (2001) Degradation of chlorinated dibenzofurans and dibenzo-*p*-dioxins by two types of bacteria having angular dioxygenases with different features. Appl Environ Microbiol 67:3610–3617

Habe H, Ide K, Yotsumoto M, Tsuji H, Yoshida T, Nojiri H, Omori T (2002) Degradation characteristics of a dibenzofuran-degrader *Terrabacter* sp. strain DBF63 toward chlorinated dioxins in soil. Chemosphere 48:201–207

Habe H, Chung JS, Kato H, Ayabe Y, Kasuga K, Yoshida T, Nojiri H, Yamane H, Omori T (2004) Characterization of the upper pathway genes for fluorene metabolism in *Terrabacter* sp. strain DBF63. J Bacteriol 186:5938–5944

Habe H, Chung JS, Ishida A, Kasuga K, Ide K, Takemura T, Nojiri H, Yamane H, Omori T (2005) The fluorene catabolic linear plasmid in *Terrabacter* sp. strain DBF63 carries the β-ketoadipate pathway genes, *pcaRHGBDCFIJ*, also found in proteobacteria. Microbiology 151:3713–3722

Hauschild JE, Masai E, Sugiyama K, Hatta T, Kimbara K, Fukuda M, Yano K (1996) Identification of an alternative 2,3-dihydroxybiphenyl 1,2-dioxygenase in *Rhodococcus* sp. strain RHA1 and cloning of the gene. Appl Environ Microbiol 62:2940–2946

Hayatsu M, Hirano M, Nagata T (1999) Involvement of two plasmids in the degradation of carbaryl by *Arthrobacter* sp. strain RC100. Appl Environ Microbiol 65:1015–1019

Hertwig S (2007) Linear Plasmids and Prophages in Gram Negative Bacteria. Microbiol Monogr 7

Huang CH, Lin YS, Yang YL, Huang SW, Chen CW (1998) The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. Mol Microbiol 28:905–916

Hund HK, de Beyer A, Lingens F (1990) Microbial metabolism of quinoline and related compounds. VI. Degradation of quinaldine by *Arthrobacter* sp. Biol Chem Hoppe-Seyler 371:1005–1008

Igloi GL, Brandsch R (2003) Sequence of the 165-kilobase catabolic plasmid pAO1 from *Arthrobacter nicotinovorans* and identification of a pAO1-dependent nicotine uptake system. J Bacteriol 185:1976–1986

Iida T, Mukouzaka Y, Nakamura K, Kudo T (2002a) Plasmid-borne genes code for an angular dioxygenase involved in dibenzofuran degradation by *Terrabacter* sp. strain YK3. Appl Environ Microbiol 68:3716–3723

Iida T, Mukouzaka Y, Nakamura K, Yamaguchi I, Kudo T (2002b) Isolation and characterization of dibenzofuran-degrading actinomycetes: analysis of multiple extradiol dioxygenase genes in dibenzofuran-degrading *Rhodococcus* species. Biosci Biotechnol Biochem 66:1462–1472

Irvine VA, Kulakov LA, Larkin MJ (2000) The diversity of extradiol dioxygenase (*edo*) genes in cresol degrading rhodococci from a creosote-contaminated site that express a wide range of degradative abilities. Antonie van Leeuwenhoek 78:341–352

Iwagami SG, Yang KQ, Davies J (2000) Characterization of the protocatechuic acid catabolic gene cluster from *Streptomyces* sp. strain 2065. Appl Environ Microbiol 66:1499–1508

Iwasaki T, Miyauchi K, Masai E, Fukuda M (2006) Multiple-subunit genes of the aromatic-ring-hydroxylating dioxygenase play an active role in biphenyl and polychlorinated biphenyl degradation in *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 72:5396–5402

Janssen DB, Scheper A, Dijkhuizen L, Witholt B (1985) Degradation of halogenated aliphatic compounds by *Xanthobacter autotrophicus* GJ10. Appl Environ Microbiol 49:673–677

Janssen DB, Pries F, van der Ploeg J, Kazemier B, Terpstra P, Witholt B (1989) Cloning of 1,2-dichloroethane degradation genes of *Xanthobacter autotrophicus* GJ10 and expression and sequencing of the *dhlA* gene. J Bacteriol 171:6791–6799

Kalkus J, Menne R, Reh M, Schlegel HG (1998) The terminal structures of linear plasmids from *Rhodococcus opacus*. Microbiology 144:1271–1279

Kesseler M, Dabbs ER, Averhoff B, Gottschalk G (1996) Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2. Microbiology 142:3241–3251

Kim D, Kim Y-S, Kim S-K, Kim SW, Zylstra GJ, Kim YM, Kim E (2002) Monocyclic aromatic hydrocarbon degradation by *Rhodococcus* sp. strain DK17. Appl Environ Microbiol 68:3270–3278

Kim D, Chae J-C, Zylstra GJ, Kim Y-S, Kim S-K, Nam MH, Kim YM, Kim E (2004) Identification of a novel dioxygenase involved in metabolism of *o*-xylene, toluene, and ethylbenzene by *Rhodococcus* sp. strain DK17. Appl Environ Microbiol 70:7086–7092

Kimura N, Urushigawa Y (2001) Metabolism of dibenzo-*p*-dioxin and chlorinated dibenzo-*p*-dioxin by a Gram-positive bacterium, *Rhodococcus opacus* SAO101. J Biosci Bioeng 92:138–143

Kimura N, Kitagawa W, Mori T, Nakashima N, Tamura T, Kamagata Y (2006) Genetic and biochemical characterization of the dioxygenase involved in lateral dioxygenation of dibenzofuran from *Rhodococcus opacus* strain SAO101. Appl Microbiol Biotechnol 73:474–484

Kitagawa W, Miyauchi K, Masai E, Fukuda M (2001) Cloning and characterization of benzoate catabolic genes in the Gram-positive polychlorinated biphenyl degrader *Rhodococcus* sp. strain RHA1. J Bacteriol 183:6598–6606

Kitagawa W, Kimura N, Kamagata Y (2004) A novel *p*-nitrophenol degradation gene cluster from a Gram-positive bacterium, *Rhodococcus opacus* SAO101. J Bacteriol 186:4894–4902

Kolkenbrock S, Parschat K, Beermann B, Hinz H-J, Fetzner S (2006) *N*-Acetylanthranilate amidase from *Arthrobacter nitroguajacolicus* Rü61a, an α/β-hydrolase-fold protein active towards aryl-acylamides and -esters, and properties of its cysteine-deficient variant. J Bacteriol 188:8430–8440

König C, Eulberg D, Gröning J, Lakner S, Seibert V, Kaschabek SR, Schlömann M (2004) A linear megaplasmid, p1CP, carrying the genes for chlorocatechol catabolism of *Rhodococcus opacus* 1CP. Microbiology 150:3075–3087

Kosono S, Maeda M, Fuji F, Arai H, Kudo T (1997) Three of the seven *bphC* genes of *Rhodococcus erythropolis* TA421, isolated for a termite ecosystem, are located on an indigenous plasmid associated with biphenyl degradation. Appl Environ Microbiol 63:3282–3285

Krum JG, Ensign SA (2001) Evidence that a linear megaplasmid encodes enzymes of aliphatic alkene and epoxide metabolism and coenzyme M (2-mercaptoethanesulfonate) biosynthesis in *Xanthobacter* strain Py2. J Bacteriol 183:2172–2177

Kulakov LA, Chen S, Allen CCR, Larkin MJ (2005) Web-type evolution of *Rhodococcus* gene clusters associated with utilization of naphthalene. Appl Environ Microbiol 71:1754–1764

Larkin MJ, De Mot R, Kulakov LA, Nagy I (1998) Applied aspects of *Rhodococcus* genetics. Antonie van Leeuwenhoek 74:133–153

Larkin MJ, Kulakov LA, Allen CCR (2005) Biodegradation and *Rhodococcus*—masters of catabolic versatility. Curr Opin Biotechnol 16:282–290

Maruyama T, Ishikura M, Taki H, Shindo K, Kasai H, Haga M, Inomata Y, Misawa N (2005) Isolation and characterization of *o*-xylene oxygenase genes from *Rhodococcus opacus* TKN14. Appl Environ Microbiol 71:7705–7715

Masai E, Sugiyama K, Iwashita N, Shimizu S, Hauschild JE, Hatta T, Kimbara K, Yano K, Fukuda M (1997) The *bphDEF meta*-cleavage pathway genes involved in biphenyl/polychlorinated biphenyl degradation are located on a linear plasmid and separated from the initial *bphACB* genes in *Rhocococcus* sp. strain RHA1. Gene 187:141–149

Mattes TE, Coleman NV, Spain JC, Gossett JM (2005) Physiological and molecular genetic analyses of vinyl chloride and ethene biodegradation in *Nocardioides* sp. strain JS614. Arch Microbiol 183:95–106

McKay DB, Prucha M, Reineke W, Timmis KN, Pieper DH (2003) Substrate specificity and expression of three 2,3-dihydroxybiphenyl 1,2-dioxygenases from *Rhodococcus globerulus* strain P6. J Bacteriol 185:2944–2951

McLeod MP, Warren RL, Hsiao WWL, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, Dosanjh M, Hara H, Petrescu A, Morin RD, Yang G, Stott JM, Schein JE, Shin H, Smailus D, Siddiqui AS, Marra MA, Jones SJM, Holt R, Brinkman FSL, Miyauchi K, Fukuda M, Davies JE, Mohn WW, Eltis LD (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. Proc Natl Acad Sci USA 103:15582–15587

Moiseeva OV, Solyanikova IP, Kaschabek SR, Gröning J, Thiel M, Golovleva LA, Schlömann M (2002) A new modified *ortho* cleavage pathway of 3-chlorocatechol degradation by *Rhodococcus opacus* 1CP: genetic and biochemical evidence. J Bacteriol 184:5282–5292

Monna L, Omori T, Kodama T (1993) Microbial degradation of dibenzofuran, fluorene, and dibenzo-*p*-dioxin by *Staphylococcus auriculans* DBF63. Appl Environ Microbiol 59:285–289

Nojiri H, Kamakura M, Urata M, Tanaka T, Chung JS, Takemura T, Yoshida T, Habe H, Omori T (2002) Dioxin catabolic genes are dispersed on the *Terrabacter* sp. DBF63 genome. Biochem Biophys Res Commun 296:233–240

Nojiri H, Shintani M, Omori T (2004) Divergence of mobile genetic elements involved in the distribution of xenobiotic–catabolic capacity. Appl Microbiol Biotechnol 64:154–174

Noumura T, Habe H, Widada J, Chung JS, Yoshida T, Nojiri H, Omori T (2004) Genetic characterization of the dibenzofuran-degrading Actinobacteria carrying the *dbfA1A2* gene homologues isolated from activated sludge. FEMS Microbiol Lett 239:147–155

Overhage J, Sielker S, Homburg S, Parschat K, Fetzner S (2005) Identification of large linear plasmids in *Arthrobacter* spp. encoding the degradation of quinaldine to anthranilate. Microbiology 151:491–500

Parschat K, Hauer B, Kappl R, Kraft R, Hüttermann J, Fetzner S (2003) Gene cluster of *Arthrobacter ilicis* Rü61a involved in the degradation of quinaldine to anthranilate. Characterization and functional expression of the quinaldine 4-oxidase genes *qoxLMS*. J Biol Chem 278:27483–27494

Parschat K, Overhage J, Strittmatter AW, Henne A, Gottschalk G, Fetzner S (2007) Complete nucleotide sequence of the 113kb linear catabolic plasmid pAL1 of *Arthrobacter nitroguajacolicus* Rü61a, and transcriptional analysis of genes involved in quinaldine degradation. J Bacteriol 189:3855–3867

Patrauchan MA, Florizone C, Dosanjh M, Mohn WW, Davies J, Eltis LD (2005) Catabolism of benzoate and phthalate in *Rhodococcus* sp. strain RHA1: redundancies and convergence. J Bacteriol 187:4050–4063

Priefert H, O'Brien XM, Lessard PA, Dexter AF, Choi EE, Tomic S, Nagpal G, Cho JJ, Agosto M, Yang L, Treadway SL, Tamashiro L, Wallace M, Sinskey AJ (2004) Indene

bioconversion by a toluene inducible dioxygenase of *Rhodococcus* sp. I24. Appl Microbiol Biotechnol 65:168–176

Ridder IS, Rozeboom HJ, Kalk KH, Janssen DB, Dijkstra BW (1997) Three-dimensional structure of L-2-haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 complexed with the substrate-analogue formate. J Biol Chem 272:33015–33022

Saeki H, Furuhashi K (1994) Cloning and characterization of a *Nocardia corallina* B-276 gene cluster encoding alkene monooxygenase. J Ferment Bioeng 78:399–406

Saeki H, Akira M, Furuhashi K, Averhoff B, Gottschalk G (1999) Degradation of trichloroethene by a linear-plasmid-encoded alkene monooxygenase in *Rhodococcus corallinus* (*Nocardia corallina*) B-276. Microbiology 145:1721–1730

Sajjaphan K, Shapir N, Wackett LP, Palmer M, Blackmon B, Tomkins J, Sadowsky MJ (2004) *Arthrobacter aurescens* TC1 atrazine catabolism genes *trzN*, *atzB*, and *atzC* are linked on a 160-kilobase region and are functional in *Escherichia coli*. Appl Environ Microbiol 70:4402–4407

Sakaguchi K (1990) Invertrons, a class of structurally and functionally related genetic elements that includes linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. Microbiol Rev 54:66–74

Sakai M, Miyauchi K, Kato N, Masai E, Fukuda M (2003) 2-Hydroxypenta-2,4-dienoate metabolic pathway genes in a strong polychlorinated biphenyl degrader, *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 69:427–433

Sekine M, Tanikawa S, Omata S, Saito M, Fujisawa T, Tsukatani N, Tajima T, Sekigawa T, Kosugi H, Matsuo Y, Nishiko R, Imamura K, Ito M, Narita H, Tago S, Fujita N, Harayama S (2006) Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. Environ Microbiol 8:334–346

Shimizu S, Kobayashi H, Masai E, Fukuda M (2001) Characterization of the 450-kb linear plasmid in a polychlorinated biphenyl degrader, *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 67:2021–2028

Stecker C, Johann A, Herzberg C, Averhoff B, Gottschalk G (2003) Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. J Bacteriol 185:5269–5274

Swaving J, Weijers CA, van Ooyen AJ, de Bont JA (1995) Complementation of *Xanthobacter* Py2 mutants defective in epoxyalkane degradation, and expression and nucleotide sequence of the complementing DNA fragment. Microbiology 141:477–484

Taguchi K, Motoyama M, Kudo T (2004) Multiplicity of 2,3-dihydroxybiphenyl dioxygenase genes in the Gram-positive polychlorinated biphenyl degrading bacterium *Rhodococcus rhodochrous* K37. Biosci Biotechnol Biochem 68:787–795

Takeda H, Hara N, Sakai M, Yamada A, Miyauchi K, Masai E, Fukuda M (2004) Biphenyl-inducible promoters in a polychlorinated biphenyl-degrading bacterium, *Rhodococcus* sp. RHA1. Biosci Biotechnol Biochem 68:1249–1258

Tam AC, Behki RM, Khan SU (1987) Isolation and characterization of an *S*-ethyl-*N*,*N*-dipropylthiocarbamate-degrading *Arthrobacter* strain and evidence for plasmid-associated *S*-ethyl-*N*,*N*-dipropylthiocarbamate degradation. Appl Environ Microbiol 53:1088–1093

Tardif G, Greer CW, Labbé D, Lau PCK (1991) Involvement of a large plasmid in the degradation of 1,2-dichloroethane by *Xanthobacter autotrophicus*. Appl Environ Microbiol 57:1853–1857

Treadway SL, Yanagimachi KS, Lankenau E, Lessard PA, Stephanopoulos G, Sinskey AJ (1999) Isolation and characterization of indene bioconversion genes from *Rhodococcus* strain I24. Appl Microbiol Biotechnol 51:786–793

Turnbull GA, Ousley M, Walker A, Shaw E, Morgan JAW (2001) Degradation of substituted phenylurea herbicides by *Arthrobacter globiformis* strain D47 and characterization of a plasmid-associated hydrolase gene, *puhA*. Appl Environ Microbiol 67:2270–2275

Uz I, Duan YP, Ogram A (2000) Characterization of the naphthalene-degrading bacterium, *Rhodococcus opacus* M213. FEMS Microbiol Lett 185:231–238

Vaillancourt FH, Haro M-A, Drouin NM, Karim Z, Maaroufi H, Eltis LD (2003) Characterization of extradiol dioxygenases from a polychlorinated biphenyl-degrading strain that possess higher specificities for chlorinated metabolites. J Bacteriol 185:1253–1260

van Beilen JB, Smits THM, Whyte LG, Schorcht S, Rothlisberger M, Plaggemeier T, Engesser K-H, Witholt B (2002) Alkane hydroxylase homologues in Gram-positive strains. Environ Microbiol 4:676–682

van der Geize R, Dijkhuizen L (2004) Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. Curr Opin Microbiol 7:255–261

van der Ploeg J, van Hall G, Janssen DB (1991) Characterization of the haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 and sequencing of the *dhlB* gene. J Bacteriol 173:7925–7933

van der Ploeg J, Smidt MP, Landa AS, Janssen DB (1994) Identification of chloroacetaldehyde dehydrogenase involved in 1,2-dichloroethane degradation. Appl Environ Microbiol 60:1599–1605

Verschueren KHG, Seljée F, Rozeboom HJ, Kalk KH, Dijkstra BW (1993) Crystallographic analysis of the catalytic mechanism of haloalkane dehalogenase. Nature 363:693–698

Warren R, Hsiao WWL, Kudo H, Myhre M, Dosanjh M, Petrescu A, Kobayashi H, Shimizu S, Miyauchi K, Masai E, Yang G, Stott JM, Schein JE, Shin H, Khattra J, Smailus D, Butterfield YS, Siddiqui A, Holt R, Marra MA, Jones SJM, Mohn WW, Brinkman FSL, Fukuda M, Davies J, Eltis LD (2004) Functional characterization of a catabolic plasmid from polychlorinated-biphenyl-degrading *Rhodococcus* sp. strain RHA1. J Bacteriol 186:7783–7795

Weinberger M, Kolenbrander PE (1979) Plasmid-determined 2-hydroxypyridine utilization by *Arthrobacter crystallopoietes*. Can J Microbiol 25:329–334

Whyte LG, Smits THM, Labbé D, Witholt B, Greer CW, van Beilen JB (2002) Gene cloning and characterization of multiple alkane hydroxylase systems in *Rhodococcus* strains Q15 and NRRL B-16531. Appl Environ Microbiol 68:5933–5942

Yamada A, Kishi H, Sugiyama K, Hatta T, Nakamura K, Masai E, Fukuda M (1998) Two nearly identical aromatic compound hydrolase genes in a strong polychlorinated biphenyl degrader, *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 64:2006–2012

Zhou HW, Guo CL, Wong YS, Tam NFY (2006) Genetic diversity of dioxygenase genes in polycyclic aromatic hydrocarbon-degrading bacteria isolated from mangrove sediments. FEMS Microbiol Lett 262:148–157

Zhou N-Y, Jenkins A, Chan Kwo Chion CKN, Leak DJ (1999) The alkene monooxygenase from *Xanthobacter* strain Py2 is closely related to aromatic monooxygenases and catalyzes aromatic monohydroxylation of benzene, toluene, and phenol. Appl Environ Microbiol 65:1589–1595

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_092/Published online: 26 May 2007

Linear Plasmids and Phytopathogenicity

Isolde Francis[1,2] · Dirk Gevers[1,3] · Mansour Karimi[1,2] ·
Marcelle Holsters[1,2,4] (✉) · Danny Vereecke[1,2]

[1]Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Ghent, Belgium

[2]Department of Molecular Genetics, Ghent University, 9052 Ghent, Belgium

[3]Laboratory of Microbiology, Ghent University,
K.L. Ledeganckstraat 35, 9000 Ghent, Belgium
marcelle.holsters@psb.ugent.be

[4]To whom correspondence should be addressed: Department of Plant Systems Biology,
VIB, Universiteit Ghent, Technologiepark 927, 9052 Ghent, Belgium

1 Introduction 100

2 *Rhodococcus*, a Metabolically Diverse Genus of the Actinomycetes 101

3 The Linear Plasmid pFiD188 of *R. fascians* Strain D188 is Closely Related to Linear Plasmids of Other *Rhodococcus* Strains 103

4 Virulence Determinants on the Linear Plasmid pFiD188 105

5 Cross Talk Between the Linear Plasmid pFiD188 and the Chromosome . . 108

6 An Overview of the Infection Process: Putting the Players in Place 109

References 111

Abstract The occurrence of plasmids is typically correlated with the acquisition of specialized traits that can be very diverse. Production of secondary metabolites, achievement of complex catabolic reactions, and adaptation to particular habitats are just a few examples of functions that are often encoded by circular or linear extrachromosomal replicons. Although virulence is another classical characteristic that is frequently associated with the presence of plasmids, until now there is only one known case in which the phytopathogenic capacity is encoded by a linear plasmid. The actinomycete *Rhodococcus fascians* is a wide-spectrum plant pathogen that provokes shooty neoplastic outgrowths on its hosts. The key virulence determinants of this bacterium are located on a linear replicon of the invertron type. Even though the sequence determination of the linear plasmid of *R. fascians* strain D188 (pFiD188) is not completed, a first comparative analysis shows that is has probably a common origin with catabolic linear plasmids of other *Rhodococcus* strains. Whereas the colinear regions encode plasmid maintenance functions, the three unique regions of pFiD188 are involved in specific aspects of the interaction.

Abbreviations

att	Attenuated
fas	Fasciation

HGT	Horizontal gene transfer
hyp	Hypervirulent
iaa	Indole-3-acetic acid
nrp	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
ORF	Open reading frame
PAI	Pathogenicity island
stk	Serine threonine kinase
TIR	Terminal inverted repeat
vic	Virulence on the chromosome

1 Introduction

With the quote "*Natura non facit saltum*", Darwin's theory states that species do not evolve with leaps, but with small steps by vertical inheritance from parent to offspring with periodic selection. Currently, there is a general agreement that also lateral or horizontal gene transfer (HGT) are important evolutionary mechanisms (Kurland et al. 2003; Lawrence and Hendrickson 2003; Gal-Mor and Finlay 2006). HGT is defined as any process in which an organism transfers genetic material to another cell that is not its offspring, via natural transformation, conjugation, or transduction (Gal-Mor and Finlay 2006). HGT can enable rapid acquisition of new traits that improve fitness under specific environmental conditions and even occupation of new ecological niches. Because pathogenicity can be considered as an adaptation to a specific niche, the host, it is not surprising that virulence factors of bacteria that are pathogenic to animals or plants are often encoded on transmissible genetic elements, such as transposons, plasmids, bacteriophages, and pathogenicity islands (Frost et al. 2005).

The term pathogenicity island (PAI), introduced by Hacker et al. (1990), specifies a region in the chromosome that contains one or commonly many virulence genes present in pathogenic bacteria but absent in not or less pathogenic bacteria of the same or related species. Within PAIs the genes have often a G+C content and codon usage that differ from those of the rest of the chromosome. PAIs are usually flanked by specific genetic elements, such as direct repeats, insertion sequence elements, and transfer RNA genes, and frequently possess genes coding for genetic mobility factors, such as integrases, transposases, and origins of replication. These features indicate that these particular genetic elements are or have previously been able to spread among bacterial populations and contribute to microbial evolution (Hacker et al. 1997; Gal-Mor and Finlay 2006). The recent availability of multiple genomic sequences and the use of various comparative genomic approaches have revealed that PAIs are widespread and can be found in Gram-negative and Gram-positive human, mammalian, and plant pathogens (Hacker and

Kaper 2000; Oelschlaeger and Hacker 2004; Gal-Mor and Finlay 2006). PAIs of phytopathogenic bacteria are found in diverse genera, such as *Streptomyces* (Kers et al. 2005), *Pseudomonas* (Alfano et al. 2000; Araki et al. 2006), *Xanthomonas* (Kim et al. 2003), and *Erwinia* (Oh and Beer 2005). The generation of new pathogenic species through mobilizable PAIs is illustrated by the recent and independent emergence of two new phytopathogenic *Streptomyces* species, *S. acidiscabies* and *S. turgidiscabies* (Kers et al. 2005; Loria et al. 2006).

Plasmids are stable self-replicating extrachromosomal elements that usually do not contain genes for essential cellular functions. The occurrence of virulence genes on circular plasmids and their role in dissemination of virulence traits has been recognized for decades (Vivian et al. 2001). However, it is becoming increasingly clear that virulence determinants are often clustered on the plasmids in typical PAIs, as in *Pantoa agglomerans* (Barash and Manulis 2005) and *Pseudomonas syringae* pv. *phaseolicola* (Oguiza et al. 2004).

Since the discovery by Hayakawa et al. (1979) of the first bacterial linear plasmid in *S. rochei*, many linear double-stranded DNA plasmids of various sizes have been isolated from *Streptomyces* species (Netolitzky et al. 1995; Mochizuki et al. 2003; Chater and Kinashi, in this volume; see also Chen, in this volume) and other bacteria, such as *Clavibacter michiganensis* subspecies *sepedonicus* (Brown et al. 2002) and the actinomycetes *Mycobacterium* (Picardeau and Vincent 1997; Le Dantec et al. 2001; Kobryn, in this volume), *Rhodococcus* (Crespi et al. 1992; Stecker et al. 2003; Sekine et al. 2006), and *Planobispora* (Polo et al. 1998). Two types of linear plasmids exist: the so-called hairpin plasmids with covalently closed ends, and the invertron-type linear plasmids with terminal inverted repeats (TIRs) and proteins covalently linked to the 5′ termini. The latter type forms the largest group of extrachromosomal linear replicons and is predominantly found in actinomycetes (Meinhardt et al. 1997; Chater and Kinashi, in this volume; see also Chen, in this volume). These linear plasmids confer advantageous abilities on their hosts, but until now there is only one example of a plant pathogenic bacterium whose virulence genes are located on a linear plasmid. This phytopathogen, the subject of this chapter, is *Rhodococcus fascians*, an actinomycete that provokes leafy gall formation on host plants.

2
Rhodococcus, a Metabolically Diverse Genus of the Actinomycetes

The genus *Rhodococcus* belongs to the *Corynebacterinae*, a group of mycolic acid-containing organisms within the *Actinomycetales*. The suborder currently includes the genera *Tsukamurella*, *Dietzia*, *Corynebacterium*, *Williamsia*, *Turicella*, *Mycobacterium*, *Gordonia*, *Skermania*, *Rhodococcus*, and *Nocardia* (Gürtler et al. 2004). Rhodococci are aerobic, Gram-positive, nonmotile, asporogenous, GC-rich, nocardioform bacteria, of which some are able to

exhibit mycelial growth with fragmentation into rod-shaped or coccoid elements. Rhodococci are common throughout nature and have been isolated from very diverse habitats, such as soils, rocks, boreholes, groundwater, marine sediments, animal dung, insect guts, and diseased animals and plants (Bell et al. 1998). This widespread occurrence reflects the great ecological and biotechnological importance of rhodococci because of their unique enzymatic capabilities and extensive catabolic diversity, such as biodegradation of hydrophobic natural compounds and xenobiotics, including polychlorinated biphenyls. These characteristic metabolic functions are often encoded by genes on plasmids (van der Geize and Dijkhuizen 2004; Larkin et al. 2005; Fetzner et al., in this volume).

Besides the metabolically useful rhodococci, there are two pathogenic species, *R. equi* and *R. fascians*. The animal pathogen *R. equi* causes primarily chronic bronchopneumonia and enteritis in foals. Since 1967, it has also been reported to infect immunocompromised humans, which makes it an emerging threat in the current era of HIV infection (Weinstock and Brown 2002). The virulence genes of *R. equi*, including the virulence-associated protein genes, are located on a circular plasmid pREAT701 within a PAI of 27.5 kb (Takai et al. 2000).

R. fascians is a phytopathogen that causes leafy galls on a broad range of dicotyledonous as well as monocotyledonous plants through secretion of signal molecules that interfere with the hormone balance and, hence, with the developmental pathways of the host (Goethals et al. 2001; Vereecke et al.

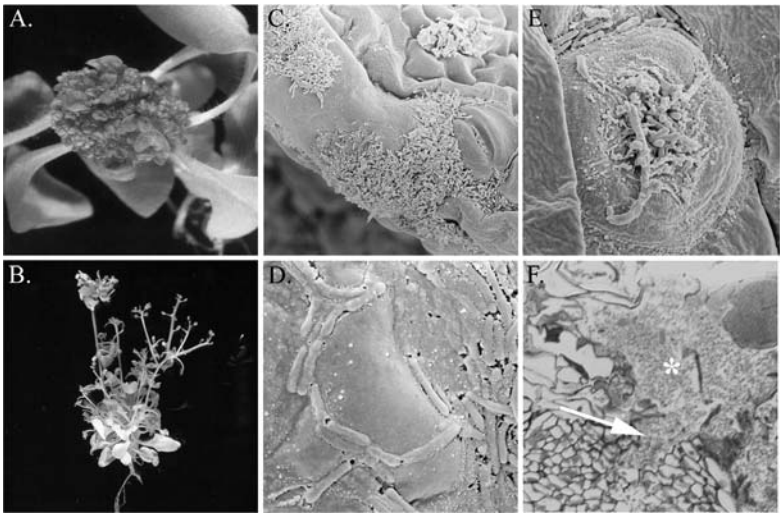


Fig. 1 Aspects of the *R. fascians*–plant interaction. **A** Leafy gall on tobacco. **B** Bushy phenotype of infected *Arabidopsis thaliana*. **C**, **D** Epiphytic colonies embedded in a protective layer. **E** Invasion of stomata. **F** Ingression site (*arrow*) under an epiphytic colony (*)

2002a) (Fig. 1). *R. fascians* initially colonizes the plant epiphytically without causing disease. Although invasion through stomata and wounds has been observed, upon an unknown trigger the bacteria penetrate the plant tissue by locally destroying the underlying cells forming ingression sites (Fig. 1). Whereas the epiphytic population directs the onset of symptom development (Cornelis et al. 2001), the endophytic population controls the sustainability of the disease. The continuous presence of metabolically active bacteria is required for maintenance of the leafy gall structure (Vereecke et al. 2000). Genes implicated in leafy gall formation are located on a large conjugative linear plasmid (Crespi et al. 1992) that does not contain genes essential for bacterial growth because it can be lost without any obvious effect on bacterial viability (Crespi et al. 1992). The sequence of the linear plasmid of *R. fascians* strain D188, designated pFiD188 (for fasciation induction), is currently being determined. As estimated by pulsed-field electrophoresis and restriction analysis (Pisabarro et al. 1998) the plasmid size is 200 kb, of which 175 kb has already been sequenced (unpublished data). Below, we will glance at the structure and composition of pFiD188.

3 The Linear Plasmid pFiD188 of *R. fascians* Strain D188 is Closely Related to Linear Plasmids of Other *Rhodococcus* Strains

Functional analysis of the linear replicon pFiD188 has shown that it carries covalently attached proteins on the 5′-terminal ends. However, no extensive TIRs, typical for invertrons, could be detected in pFiD188 (unpublished data). This lack of long TIRs is a common characteristic of several rhodococcal linear plasmids (Kalkus et al. 1998; Stecker et al. 2003; Sekine et al. 2006).

Homology searches with the available sequences reveal that pFiD188 has no homology to the circular virulence plasmid of *R. equi*, but has a limited conservation with genes on linear plasmids of other actinomycetes, such as SCP1 of *S. coelicolor* A3(2) (Bentley et al. 2004), plasmid 1 of *Mycobacterium* species MCS (accession CP000385), and pRHL1 and pRHL3 of *Rhodococcus* sp. strain RHA1 (Warren et al. 2004; McLeod et al. 2006). However, pFiD188 is strikingly similar to the linear plasmids pBD2 and pREL1 of *R. erythropolis* strains BD2 and PR4 (Stecker et al. 2003; Sekine et al. 2006), and pRHL2 of *Rhodococcus* sp. strain RHA1 (Shimizu et al. 2001; McLeod et al. 2006; Fetzner et al., in this volume), suggesting a common origin (Fig. 2). The two *R. erythropolis* plasmids exhibit six regions of extensive homology (R1 through R6; Sekine et al. 2006). After comparative analysis of the four linear replicons with the Artemis comparison tool (Carver et al. 2005), four colinear regions (labeled R1, R2, R3, and R6) and three regions unique for each plasmid (labeled U1, U2, and U3) were found. The conserved regions probably code for plasmid replication, maintenance, and partitioning functions. However, because

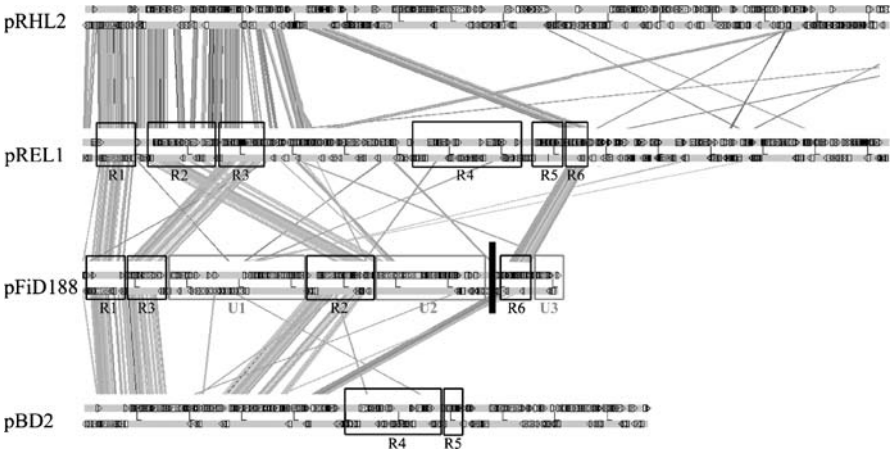


Fig. 2 Sequence comparison of the linear plasmids. The Artemis comparison tool (Carver et al. 2005) was used to plot protein similarities (tblastx) between pFiD188 (*R. fascians*), pBD2 and pREL1 (*R. erythropolis* strain BD2 and PR4, respectively), and pRHL2 (*Rhodococcus* sp. strain RHA1). *Lines* drawn between two adjacent linearized replicons show the location of homologous genes; *darker lines* represent higher-quality matches. Conserved (R) and unique (U) regions are indicated. The *black bar* flanking region R6 in pFiD188 depicts the site of the sequence gap

the regions R2, R3, and R6 mostly hold genes that encode hypothetical proteins, it is impossible to predict their role in these processes. On the other hand, region R1 comprises putative genes for the telomere-binding and terminal proteins and, therefore, most probably functions in replication of the linear plasmids (Bao and Cohen 2003; see also Chen, in this volume). R4, the most elaborate region of homology between pBD2 and pREL1, and region R5 are not conserved on pFiD188 nor on pRHL2. The gene products encoded by these two regions are involved in lipoprotein processing, and heavy metal and arsenate resistance (Sekine et al. 2006). In *R. fascians* strain D188, heavy metal resistance (cadmium) is carried on a separate replicon, the circular plasmid pD188 (Desomer et al. 1988). Interestingly, whereas one end of the four linear replicons is very conserved, the other end is unique for each plasmid (Fig. 2) and represents genes involved in the characteristic features of the particular strains. For plasmids pBD2 and pREL1, the unique end encodes specific metabolic functions, isopropylbenzene catabolism (Stecker et al. 2003) and alkane degradation (Sekine et al. 2006), respectively. For pFiD188, this region is possibly involved in virulence, whereas for pRHL2 it mainly contains hypothetical proteins. Besides the unique end, each of the four replicons has two additional nonconserved regions (Fig. 2). In pFiD188, the unique region U1 consists of four loci, three of which (*hyp*, *att*, and *fas*) have been identified as virulence loci through random insertion mutagenesis (Crespi et al. 1992); region U2 carries the *nrp* locus, and the terminal region

U3 harbors the *stk* locus. Although the latter unique regions have not been extensively characterized, they plausibly also encode functions that are important during the interaction of *R. fascians* with its host. In the next section, the genes located in the three unique regions of pFiD188 will be discussed in detail.

4 Virulence Determinants on the Linear Plasmid pFiD188

Two of the four loci of region U1, *fas* and *att*, are fundamental for virulence and have been extensively studied (Goethals et al. 2001). The *fas* locus codes for the most important pathogenicity factor of *R. fascians* and mutations lead to a nonvirulent phenotype. The locus contains an operon of six open reading frames (ORFs) that encode the machinery for the synthesis of a cytokinin-like signal molecule essential for symptom development (Crespi et al. 1994; Fig. 3). The gene product of *orf*1 is homologous to P450-type cytochrome monooxygenases that play a central role in biosynthetic and biodegradative pathways in actinomycetes (McLean et al. 2005). The accompanying electron transport complex is encoded by *orf*2 and *orf*3. The *orf*2 gene product is a bifunctional protein: the amino-terminal part shows homology to ferredoxins, whereas the carboxy-terminal region is similar to the α subunit of pyruvate dehydrogenases. The β subunit of this enzyme is encoded by the gene product of *orf*3. The products of these three genes constitute most probably the machinery for a single enzymatic reaction directed by the P450 monooxygenase. The central gene *orf*4 codes for an isopentenyl transferase involved in the first committed step toward cytokinin biosynthesis (Crespi et al. 1992; Sakakibara 2005). The *orf*5 and *orf*6 gene products show similarity to cytokinin dehydrogenases, and glutathione-(*S*)-transferases and lysine decarboxylases, respectively.

The central role of the *fas* operon in the pathology is reflected by the strict correlation between virulence and the presence of *orf*1 and *orf*4 in a large collection of *R. fascians* isolates (Stange et al. 1996; Miller et al. 2006). Moreover, the *fas* operon has been identified in the chromosomal PAI of *S. turgidis-*

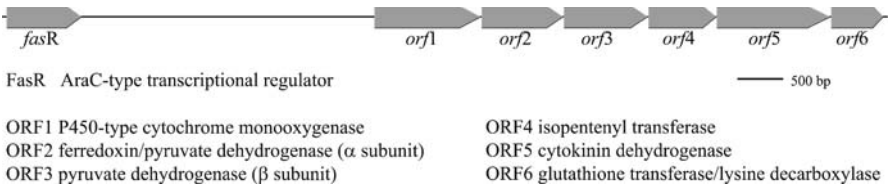


Fig. 3 Schematic representation of the *fas* operon. The organization of the genes and their deduced polypeptides are indicated

cabies, a sporulating actinomycete and causal agent of scab disease (Kers et al. 2005). The typical disease symptoms caused by this pathogen are not related to fasciation. Under laboratory conditions, however, *S. turgidiscabies* is able to induce leafy galls on tobacco (*Nicotiana tabacum*) that are indistinguishable from *R. fascians*-induced hyperplasia, supporting the important role of the *fas* genes in this symptomatology (Loria et al. 2006). Although *fas* transcription is constitutive, translation of the *fas* messenger RNAs is highly responsive to several environmental signals. In the early steps of the interaction, it is absolutely dependent on the activity of the Att proteins that produce an autoregulatory compound (Temmerman et al. 2000; Maes et al. 2001; Cornelis et al. 2002). Upstream of the *fas* genes lies *fas*R that encodes a key transcriptional regulator, a member of the AraC family (Temmerman et al. 2000; Martin and Rosner 2001) that is essential for *fas* gene expression because the *fas*R mutant is nonvirulent. The working model is that FasR in combination with the autoregulatory compound controls the expression of a translational activator that will mediate translation of the constitutively produced *fas* messenger RNAs (Temmerman et al. 2000).

The actual structure of the *att*-encoded autoregulatory compound remains to be elucidated, but sequence analysis of the nine ORFs of the *att* operon (*att*XABCDEFGH) revealed that the gene products of *att*A, *att*B, and *att*H are involved in arginine biosynthesis, whereas *att*D, *att*E, and *att*F are homologous to proteins implicated in the formation of a β-lactam ring. Hence, the autoregulatory compound is hypothesized to be an antibiotic-like molecule. Mutants in the *att* locus display an attenuated virulence phenotype, confirming the need for the Att autoregulatory compound to express appropriately *fas* that, in turn, is required for full virulence (Maes et al. 2001; Cornelis et al. 2002). Based on experimental data, the Att compound is proposed to be produced at a low constitutive level that steadily increases as the bacteria increase in number or get confined on the plant surface. At a certain threshold level, *att* gene expression is induced by AttR, a LysR-type transcriptional regulator (Maes et al. 2001; Schell 1993), with a steep increase in the concentration of the Att compound as a consequence. This positive autoregulation is needed for the direct or indirect onset of expression of other virulence-associated genes, such as *fas*. The *att* genes are expressed principally in bacteria located on the plant surface at early stages of infection, whereas expression of the *fas* locus occurs throughout infection (Cornelis et al. 2002). Besides its role as autoregulatory molecule, the compound produced by the Att proteins seems to be involved in the penetration of plant tissues. Although *att* mutants are capable of inducing mild symptoms on their host, unlike the wild-type strain, they require wounding to do so (Maes et al. 2001). Whether this defect is an indirect result of the poor induction of an unidentified penetration locus or whether the Att compound itself is directly involved in breaking through the plant cell wall is currently unknown. Whatever the case, the *att* operon obviously plays a key role in the

shift from an epiphytic plant-associated to an endophytic plant-pathogenic lifestyle.

For the two other loci of region U1 far less functional information is currently available. A *hyp* mutant exhibits a hypervirulent phenotype characterized by the early formation of larger leafy galls on tobacco, suggesting that the Hyp proteins balance virulence by either modifying the bacterial signals to a less active form or by controlling the expression level of the virulence genes (Crespi et al. 1992). Sequence analysis of the *hyp* locus favors the latter hypothesis because of the presence of a gene encoding a homolog of the D-E-A-D box family of RNA helicases (Vereecke et al. 1996; Temmerman 2000). These proteins play an important role in the regulation of messenger RNA expression and translation initiation (Linder 2006). Finally, located between the *hyp* and the *att* locus lies a gene cluster possibly involved in the biosynthesis of γ-butyrolactone (unpublished data). In Gram-positive bacteria, γ-butyrolactones function as signaling molecules that regulate antibiotic production and morphological differentiation (Takano 2006). Whether this signaling molecule plays a role during the interaction of *R. fascians* and its host remains to be determined.

Mutation of the *nrp* locus located in the unique region U2 also leads to an attenuated virulence (Maes 2001). Sequence analysis of the *nrp* locus revealed that the six ORFs probably code for a nonribosomal peptide synthetase (NRPS) composed of five modules, suggesting the production of a pentapeptide. NRPSs are responsible for the production of a multitude of biologically active compounds, including signal molecules and phytotoxins (Finking and Marahiel 2004). Interestingly, the dipeptidic thaxtomin secreted by phytopathogenic *Streptomyces* species inhibits cellulose biosynthesis by the host and causes cell hypertrophy in expanding plant tissues, thus facilitating the penetration of the plant by the bacteria (Kers et al. 2005; Loria et al. 2006). By analogy, we are currently investigating whether the peptide produced by the NRPS of *R. fascians* functions as a toxin produced in the early steps of the infection process to enable access to inner plant tissues.

The terminal unique region U3 contains the *stk* locus and harbors a five-gene operon (unpublished data). The central gene putatively encodes a bifunctional protein; the carboxy terminus is homologous to penicillin-binding proteins, and the amino terminus to RodA/FtsW cell shape/cell division proteins of the SEDS family (acronym for shape, elongation, division, and sporulation) (Noirclerc-Savoye et al. 2003). The other genes of the operon are similar to proteins involved in phosphate-dependent signal transduction and regulation pathways. Remarkably, this operon is conserved in all actinobacteria sequenced to date, including *Mycobacterium tuberculosis*, *M. leprae*, *Corynebacterium glutamicum*, and *S. coelicolor*, pointing to an essential role in the cell's biology (Boitel et al. 2003; Dasgupta et al. 2006). In these organisms, the region is located on the chromosome and is vital for bacterial growth through regulation of cell division and/or cell elongation. By

analogy, in *R. fascians* a functionally conserved operon is predicted to be located on the chromosome. Hence, the homologous copy on the linear plasmid, which is dispensable for bacterial growth, probably results from a gene duplication event or has been acquired through HGT. Because in *M. tuberculosis* and *Streptococcus pyogenes* these proteins are involved in bacterial morphology and septation during cell division (Kang et al. 2005; Dasgupta et al. 2006; Jin and Pancholi 2006), we presently hypothesize that the *stk* locus of *R. fascians* is involved in the directional growth of the bacterium towards the plant.

5
Cross Talk Between the Linear Plasmid pFiD188 and the Chromosome

The earliest finding on the possible strategy utilized by *R. fascians* to alter plant development was the notion that some bacterial strains were able to degrade auxin (Lacey 1948). The opinion was that a decreased auxin concentration in the infected tissues would shift the auxin/cytokinin balance in favor of shoot formation (Kemp 1978). However, more recent studies of the phenotype of infected plants have revealed that auxin might play an important role in symptom development (Vereecke et al. 2000). Analysis of secreted indole compounds has indeed demonstrated that *R. fascians* strain D188 carries auxin biosynthetic genes on its chromosome. Although the indole-3-acetic acid (*iaa*) genes have not been identified yet, biochemical analyses show that the kinetics of auxin biosynthesis are controlled by pFiD188 (Vandeputte et al. 2005). Based on the available data, a dual role for auxin during the interaction is postulated, first as colonization factor and subsequently as a virulence determinant.

Persistent survival inside the plant tissues, required for maintenance of the leafy gall structure, necessitates the alteration of the bacterial metabolism. The expression of a malate synthase gene *vic*A, located in the *vic* (virulence on the chromosome) locus, is upregulated in the presence of plant extracts and controlled by the linear plasmid. Malate synthase functions in the glyoxylate shunt of the Krebs cycle and catalyzes the condensation of acetyl coenzyme A and glyoxylate to form malate. Together with isocitrate lyase, it circumvents the loss of two carbons during the tricarboxylic acid cycle, when the bacterium grows on C_2 compounds as the sole carbon source. The functionality of the *vic* locus is imperative for endophytic survival, because mutations render the bacteria unable to grow on the nutrients available inside the plant, with a diminished virulence as a result (Vereecke et al. 2002a,b).

The above data show that although the main virulence determinants are encoded by the linear plasmid pFiD188, the chromosome plays an important role in the adaptation to the epiphytic and endophytic lifestyle. All aspects of

the infection process require fine-tuned expression of key genes, which is accomplished by the interplay of regulatory proteins and compounds encoded both by the chromosome and the linear plasmid, allowing the integration of a wide range of signals and an equilibration of the interaction.

The distribution of virulence-associated genes over chromosome and plasmid occurs in other phytopathogens as well. In *Clavibacter michiganensis* subspecies *michiganensis*, for instance, the genes directly responsible for induction of disease in host plants are located on two transmissible circular plasmids, pCM1 and pCM2. Full virulence is only observed when both plasmids are present. However, genes required for the other aspects of the plant–microbe interaction, such as host recognition, infection, suppression of host defense, and effective colonization, map on the chromosome (Meletzus et al. 1993; Gartemann et al. 2003). Similarly, in *Agrobacterium tumefaciens*, the tumor-inducing plasmid carries the main virulence machinery. Nevertheless, the very first step of the interaction with the host, chemotaxis, is accomplished by the cooperation of both plasmid- and chromosome-borne proteins. Moreover, attachment to the host cell, a vital step in pathogenicity, is also mediated by multiple chromosomal gene products (McCullen and Binns 2006). These observations underline the fact that pathogenicity acquired by plasmids is a novel adaptive trait that broadens the ecological niche of soilborne organisms.

6
An Overview of the Infection Process: Putting the Players in Place

R. fascians is a ubiquitous well-adapted soil bacterium, but it is an equally well-adapted epiphyte. A first crucial step to gain access into this new habitat is to reach the plant surface from its location in the soil. As a nonmotile organism, it is currently unknown how *R. fascians* accomplishes this aspect. Perhaps the *stk* locus enables the bacteria to exhibit directional (hyphal) growth toward the plant. Of course, it cannot be ruled out that the bacteria encounter the plant by chance via, e.g., plant growth, raindrops, insect vectors, or other soil-dwelling organisms. As soon as *R. fascians* contacts a host plant, it will form large epiphytic colonies embedded in a protective slime layer of bacterial origin. A constitutive level of secreted auxin may aid in the acquisition of nutrients or in the suppression of plant defense. The bacteria enter the plant via ingression sites that necessitate breaching through the cuticula and the epidermal cell layer. This local breakdown of the plant's outer barrier is possibly facilitated by a phytotoxin produced by the NRPS encoded by the *nrp* locus. Expression of the virulence genes is initially controlled by the production of the Att compound that gradually builds up under specific conditions favorable for infection. When a threshold concentration is reached, a positive autoregulatory loop is activated,

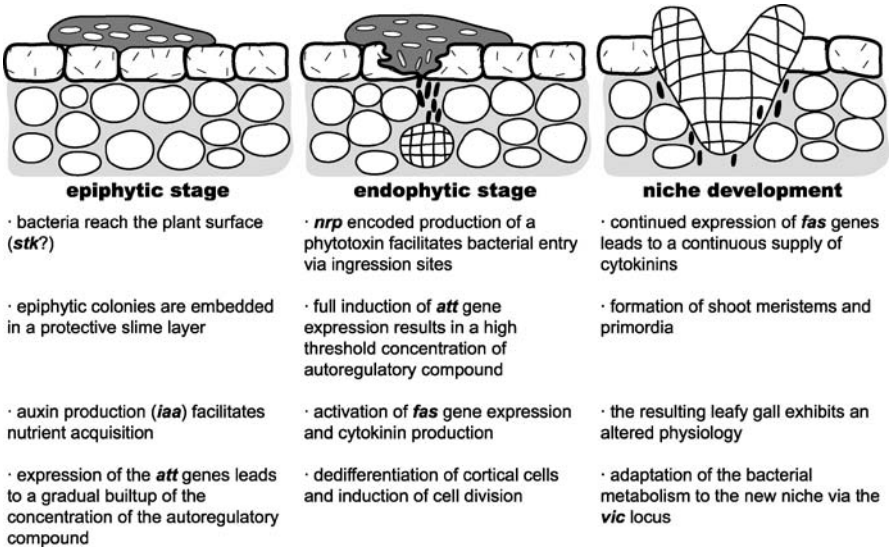


Fig. 4 Overview of the infection process of *R. fascians*. *Dark gray*, slime layer; *light gray*, apoplast; *dashes*, epidermal cells; *white*, cortical cells; *white* bacteria, express both the *att* and the *fas* genes; *gray* bacteria, stop expressing the *att* genes; and *black* bacteria, express the *fas* genes

leading to full induction of *att* gene expression and activation of *fas* gene expression. Whereas the expression of the *att* genes is restricted to the epiphytic bacteria, the synthesis of the *fas*-dependent cytokinin-like signal molecules continues throughout the interaction. Expression of the *fas* genes in bacteria located in the epidermis coincides with the dedifferentiation of parenchyma cells to shoot meristems. The persistent delivery of morphogenic substances by the bacteria is imperative for the formation and maintenance of the leafy gall structure. In the leafy gall, physiological conditions are different from those of normal nondiseased plants. The spectrum of compounds that prevail in a leafy gall is believed to strongly differ from that of uninfected plants. Bacteria have most probably coevolved to utilize these molecules as specific nutrients via the chromosomal *vic* locus. Hence, the induction of leafy gall formation leads to a specific niche for *R. fascians* that will metabolically colonize this newly formed organ. Different aspects of the interaction are illustrated in Fig. 1, whereas a schematic overview is given in Fig. 4.

Acknowledgements The authors wish to thank Martine De Cock for help in preparing the manuscript. This work was supported by a research grant from the Research Foundation–Flanders (G0338.04). I.F. is indebted to the Institute for the Promotion of Innovation by Science and Technology in Flanders for a predoctoral fellowship. D.G. is a postdoctoral researcher of the Research Foundation–Flanders.

References

Alfano JR, Charkowski AO, Deng W-L, Badel JL, Petnicki-Ocwieja T, van Dijk K, Collmer A (2000) The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. Proc Natl Acad Sci USA 97:4856–4861

Araki H, Tian D, Gross EM, Jakob K, Halldorsdottir SS, Kreitman M, Bergelson J (2006) Presence/absence polymorphism for alternative pathogenicity islands in *Pseudomonas viridiflava*, a pathogen of *Arabidopsis*. Proc Natl Acad Sci USA 103:5887–5892

Bao K, Cohen SN (2006) Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. Genes Dev 17:774–785

Barash I, Manulis S (2005) Hrp-dependent biotrophic mechanism of virulence: how has it evolved in tumorigenic bacteria? Phytoparasitica 33:317–324

Bell KS, Philp JC, Aw DWJ, Christofi N (1998) The genus *Rhodococcus*. J Appl Microbiol 85:195–210

Bentley SD, Brown S, Murphy LD, Harris DE, Quail MA, Parkhill J, Barrell BG, McCormick JR, Santamaria RI, Losick R, Yamasaki M, Kinashi H, Chen CW, Chandra G, Jakimowicz D, Kieser HM, Kieser T, Chater KF (2004) SCP1, a 356 023 bp linear plasmid adapted to the ecology and developmental biology of its host, *Streptomyces coelicolor* A3(2). Mol Microbiol 51:1615–1628

Boitel B, Ortiz-Lombardía M, Durán R, Pompeo F, Cole ST, Cerveñansky C, Alzari PM (2003) PknB kinase activity is regulated by phosphorylation in two Thr residues and dephosphorylation by PstP, the cognate phospho-Ser/Thr phosphatase, in *Mycobacterium tuberculosis*. Mol Microbiol 49:1493–1508

Brown SE, Knudson DL, Ishimaru CA (2002) Linear plasmid in the genome of *Clavibacter michiganensis* subsp. *sepedonicus*. J Bacteriol 184:2841–2844

Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J (2005) ACT: the Artemis comparison tool. Bioinformatics 21:3422–3423

Chater KF, Kinashi H (2007) *Streptomyces* Linear Plasmids: Their Discovery, Functions, Interactions with other Replicons, and Evolutionary Significance (in this volume). Springer, Heidelberg

Chen CW (2007) *Streptomyces* Linear Plasmids: Replication and Telemores (in this volume). Springer, Heidelberg

Cornelis K, Ritsema T, Nijsse J, Holsters M, Goethals K, Jaziri M (2001) The plant pathogen *Rhodococcus fascians* colonizes the exterior and interior of the aerial parts of plants. Mol Plant-Microbe Interact 14:599–608

Cornelis K, Maes T, Jaziri M, Holsters M, Goethals K (2002) Virulence genes of the phytopathogen *Rhodococcus fascians* show specific spatial and temporal expression patterns during plant infection. Mol Plant-Microbe Interact 15:398–403

Crespi M, Messens E, Caplan AB, Van Montagu M, Desomer J (1992) Fasciation induction by the phytopathogen *Rhodococcus fascians* depends upon a linear plasmid encoding a cytokinin synthase gene. EMBO J 11:795–804

Crespi M, Vereecke D, Temmerman W, Van Montagu M, Desomer J (1994) The *fas* operon of *Rhodococcus fascians* encodes new genes required for efficient fasciation of host plants. J Bacteriol 176:2492–2501

Dasgupta A, Datta P, Kundu M, Basu J (2006) The serine/threonine kinase PknB of *Mycobacterium tuberculosis* phosphorylates PBPA, a penicillin-binding protein required for cell division. Microbiology 152:493–504

Desomer J, Dhaese P, Van Montagu M (1988) Conjugative transfer of cadmium resistance plasmids in *Rhodococcus fascians* strains. J Bacteriol 170:2401–2405
Fetzner S, Kolkenbrock S, Parschat K (2007) Catabolic Linear Plasmids (in this volume). Springer, Heidelberg
Finking R, Marahiel MA (2004) Biosynthesis of nonribosomal peptides. Annu Rev Microbiol 58:453–488
Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732
Gal-Mor O, Finlay BB (2006) Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell Microbiol 8:1707–1719
Gartemann K-H, Kirchner O, Engemann J, Gräfen I, Eichenlaub R, Burger A (2003) *Clavibacter michiganensis* subsp. *michiganensis*: first steps in the understanding of virulence of a Gram-positive phytopathogenic bacterium. J Biotechnol 106:179–191
Goethals K, Vereecke D, Jaziri M, Van Montagu M, Holsters M (2001) Leafy gall formation by *Rhodococcus fascians*. Annu Rev Phytopathol 39:27–52
Gürtler V, Mayall BC, Seviour R (2004) Can whole genome analysis refine the taxonomy of the genus *Rhodococcus*? FEMS Microbiol Rev 28:377–403
Hacker J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol 54:641–679
Hacker J, Bender L, Ott M, Wingender J, Lund B, Marre R, Goebel W (1990) Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. Microb Pathog 8:213–225
Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol Microbiol 23:1089–1097
Hayakawa T, Otake N, Yonehara H, Tanaka T, Sakaguchi K (1979) Isolation and characterization of plasmids from *Streptomyces*. J Antibiot 32:1348–1350
Jin H, Pancholi V (2006) Identification and biochemical characterization of a eukaryotic-type serine/threonine kinase and its cognate phosphatase in *Streptomyces pyogenes*: their biological functions and substrate identification. J Mol Biol 357:1351–1372
Kalkus J, Menne R, Reh M, Schlegel HG (1998) The terminal structures of linear plasmids from *Rhodococcus opacus*. Microbiology 144:1271–1279
Kang C-M, Abbott DW, Park ST, Dascher CC, Cantley LC, Husson RN (2005) The *Mycobacterium tuberculosis* serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. Genes Dev 19:1692–1704
Kemp DR (1978) Indole-3-ylacetic acid metabolism of *Corynebacterium fascians*. In: Loutit MW, Miles JAR (eds) Microbial ecology. Springer, Berlin, pp 341–345
Kers JA, Cameron KD, Joshi MV, Bukhalid RA, Morello JE, Wach MJ, Gibson DM, Loria R (2005) A large, mobile pathogenicity island confers plant pathogenicity on *Streptomyces* species. Mol Microbiol 55:1025–1033
Kim J-G, Park BK, Yoo C-H, Jeon E, Oh J, Hwang I (2003) Characterization of the *Xanthomonas axonopodis* pv. glycines Hrp pathogenicity island. J Bacteriol 185:3155–3166
Kobryn K (2007) The Linear Hairpin Replicons of *Borrelia burgdorferi* (in this volume). Springer, Heidelberg
Kurland CG, Canback B, Berg OG (2003) Horizontal gene transfer: a critical view. Proc Natl Acad Sci USA 100:9658–9662
Lacey MS (1948) Studies on *Bacterium fascians*. V. Further observations on the pathological and physiological reactions of *Bact. fascians*. Ann Appl Biol 35:572–581

Larkin MJ, Kulakov LA, Allen CCR (2005) Biodegradation and *Rhodococcus*—masters of catabolic versatility. Curr Opin Biotechnol 16:282–290

Lawrence JG, Hendrickson H (2003) Lateral gene transfer: when will adolescence end? Mol Microbiol 50:739–749

Le Dantec C, Winter N, Gicquel B, Vincent V, Picardeau M (2001) Genomic sequence and transcriptional analysis of a 23-kilobase mycobacterial linear plasmid: evidence for horizontal transfer and identification of plasmid maintenance systems. J Bacteriol 183:2157–2164

Linder P (2006) Dead-box proteins: a family affair—active and passive players in RNP remodeling. Nucleic Acids Res 34:4168–4180

Loria R, Kers J, Joshi M (2006) Evolution of plant pathogenicity in *Streptomyces*. Annu Rev Phytopathol 44:467–487

Maes T (2001) The *att* locus of the plant pathogen *Rhodococcus fascians*. PhD dissertation, Ghent University, Belgium

Maes T, Vereecke D, Ritsema T, Cornelis K, Ngo Thi Thu H, Van Montagu M, Holsters M, Goethals K (2001) The *att* locus of *Rhodococcus fascians* strain D188 is essential for full virulence on tobacco through the production of an autoregulatory compound. Mol Microbiol 42:13–28

Martin RG, Rosner JL (2001) The AraC transcriptional activators. Curr Opin Microbiol 4:132–137

McCullen CA, Binns AN (2006) *Agrobacterium tumefaciens* and plant cell interactions and activities required for interkingdom macromolecular transfer. Annu Rev Cell Dev Biol 22:101–127

McLean KJ, Sabri M, Marshall KR, Lawson RJ, Lewis DG, Clift D, Balding PR, Dunford AJ, Warman AJ, McVey JP, Quinn A-M, Sutcliffe MJ, Scrutton NS, Munro AW (2005) Biodiversity of cytochrome P450 redox systems. Biochem Soc Trans 33:796–801

McLeod MP, Warren RL, Hsiao WWL, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, Dosanjh M, Hara H, Petrescu A, Morin RD, Yang G, Stott JM, Schein JE, Shin H, Smailus D, Siddiqui AS, Marra MA, Jones SJM, Holt R, Brinkman FSL, Miyauchi K, Fukuda M, Davies JE, Mohn WW, Eltis LD (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insight into a catabolic powerhouse. Proc Natl Acad Sci USA 103:15582–15587

Meinhardt F, Schaffrath R, Larsen M (1997) Microbial linear plasmids. Appl Microbiol Biotechnol 47:329–336

Meletzus D, Bermpohl A, Dreier J, Eichenlaub R (1993) Evidence for plasmid-encoded virulence factors in the phytopathogenic bacterium *Clavibacter michiganensis* supsp. *michiganensis* NCPPB382. J Bacteriol 175:2131–2136

Miller ML, Collins K, Kraus J, Putnam ML (2006) PCR detection of pathogenic *Rhodococcus fascians* and *Agrobacterium tumefaciens* from herbaceous perennials. Phytopathology 96:S79

Mochizuki S, Hiratsu K, Suwa M, Ishii T, Sugino F, Yamada K, Kinashi H (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. Mol Microbiol 48:1501–1510

Netolitzky DJ, Wu X, Jensen SE, Roy KL (1995) Giant linear plasmids of β-lactam antibiotic producing *Streptomyces*. FEMS Microbiol Lett 131:27–34

Noirclerc-Savoye M, Morlot C, Gérard P, Vernet T, Zapun A (2003) Expression and purification of FtsW and RodA from *Streptococcus pneumoniae*, two membrane proteins involved in cell division and cell growth, respectively. Protein Expr Purif 30:18–25

Oelschlaeger TA, Hacker J (2004) Impact of pathogenicity islands in bacterial diagnostics. APMIS 112:930–936

Oguiza JA, Rico A, Rivas LA, Sutra L, Vivian A, Murillo J (2004) *Pseudomonas syringae* pv. phaseolicola can be separated into two genetic lineages distinguished by the possession of the phaseolotoxin biosynthetic cluster. Microbiology 150:473–482

Oh C-S, Beer SV (2005) Molecular genetics of *Erwinia amylovora* involved in the development of fire blight. FEMS Microbiol Lett 253:185–192

Picardeau M, Vincent V (1997) Characterization of large linear plasmids in mycobacteria. J Bacteriol 179:2753–2756

Pisabarro A, Correia A, Martín JF (1998) Pulsed-field gel electrophoresis analysis of the genome of *Rhodococcus fascians*: genome size and linear and circular replicon composition in virulent and avirulent strains. Curr Microbiol 36:302–308

Polo S, Guerini O, Sosio M, Dehò G (1998) Identification of two linear plasmids in the actinomycete *Planobispora rosea*. Microbiology 144:2819–2825

Sakakibara H (2005) Cytokinin biosynthesis and regulation. Vitam Horm 72:271–287

Schell MA (1993) Molecular biology of the LysR family of transcriptional regulators. Annu Rev Microbiol 47:597–626

Sekine M, Tanikawa S, Omata S, Saito M, Fujisawa T, Tsukatani N, Tajima T, Sekigawa T, Kosugi H, Matsuo Y, Nishiko R, Imamura K, Ito M, Narita H, Tago S, Fujita N, Harayama S (2006) Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. Environ Microbiol 8:334–346

Shimizu S, Kobayashi H, Masai E, Fukuda M (2001) Characterization of the –450 kb linear plasmid in a polychlorinated biphenyl degrader, *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 67:2021–2028

Stange RR, Jeffares D, Young C, Scott DB, Eason JR, Jameson PE (1996) PCR amplification of the *fas-1* gene for detection of virulent strains of *Rhodococcus fascians*. Plant Pathol 45:407–417

Stecker C, Johann A, Herzberg C, Averhoff B, Gottschalk G (2003) Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. J Bacteriol 185:5269–5274

Takai S, Hines SA, Sekizaki T, Nicholson VM, Alperin DA, Osaki M, Takamatsu D, Nakamura M, Suzuki K, Ogino N, Kakuda T, Dan H, Prescott JF (2000) DNA sequence and comparison of virulence plasmids from *Rhodococcus equi* ATCC 33701 and 103. Infect Immun 68:6840–6847

Takano E (2006) γ-Butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. Curr Opin Microbiol 9:287–294

Temmerman W (2000) Role of the *fas* locus in leafy gall development by the phytopathogen *Rhodococcus fascians*. PhD thesis, Ghent University, Belgium

Temmerman W, Vereecke D, Dreesen R, Van Montagu M, Holsters M, Goethals K (2000) Leafy gall formation is controlled by *fas*R, an AraC-type regulatory gene, in *Rhodococcus fascians*. J Bacteriol 182:5832–5840

van der Geize R, Dijkhuizen L (2004) Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. Curr Opin Microbiol 7:255–261

Vandeputte O, Öden S, Mol A, Vereecke D, Goethals K, El Jaziri M, Prinsen E (2005) Biosynthesis of auxin by the Gram-positive phytopathogen *Rhodococcus fascians* is controlled by compounds specific to infected plant tissues. Appl Environ Microbiol 71:1169–1177

Vereecke D, Temmerman W, Maes T, Van Montagu M, Goethals K (1996) Molecular analysis of the virulence determinants of the phytopathogen *Rhodococcus fascians*. Med Fac Landbouwwet Univ Gent 61/2a:231–240

Vereecke D, Burssens S, Simón-Mateo C, Inzé D, Van Montagu M, Goethals K, Jaziri M (2000) The *Rhodococcus fascians*–plant interaction: morphological traits and biotechnological applications. Planta 210:241–251

Vereecke D, Cornelis K, Temmerman W, Jaziri M, Van Montagu M, Holsters M, Goethals K (2002a) Chromosomal locus that affects the pathogenicity of *Rhodococcus fascians*. J Bacteriol 184:1112–1120

Vereecke D, Cornelis K, Temmerman W, Holsters M, Goethals K (2002b) Versatile persistence pathways for pathogens of animals and plants. Trends Microbiol 10:485–488

Vivian A, Murillo J, Jackson RW (2001) The roles of plasmids in phytopathogenic bacteria: mobile arsenals? Microbiology 147:763–780

Warren R, Hsiao WWL, Kudo H, Myhre M, Dosanjh M, Petrescu A, Kobayashi H, Shimizu S, Miyauchi K, Masai E, Yang G, Stott JM, Schein JE, Shin H, Khattra J, Smailus D, Butterfield YS, Siddiqui A, Holt R, Marra MA, Jones SJM, Mohn WW, Brinkman DSL, Fukuda M, Davies J, Eltis LD (2004) Functional characterization of a catabolic plasmid from polychlorinated-biphenyl-degrading *Rhodococcus* sp. strain RHA1. J Bacteriol 186:7783–7795

Weinstock DM, Brown AE (2002) *Rhodococcus equi*: an emerging pathogen. Clin Infect Dis 34:1379–1385

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_093/Published online: 2 June 2007

The Linear Hairpin Replicons of *Borrelia burgdorferi*

Kerri Kobryn

Dép. de Microbiologie et d'Infectiologie, Faculté de Médecine, Université de Sherbrooke, 3001, 12e Ave N., Sherbrooke, QC J1H 5N4, Canada
Kerri.Kobryn@USherbrooke.ca

1 Introduction 118

2 Lyme Borreliosis 118

3 The *Borrelia burgdorferi* Genome 119

4 DNA Replication 120
4.1 Mode of Replication and Origins of Replication 120
4.2 Autonomous Replication and Plasmid Compatibility Functions 122

5 Telomere Resolution 124
5.1 In Vivo Demonstration of Telomere Resolution 124
5.2 The Telomere Resolvase (ResT) 125
5.3 Substrate Requirements for Telomere Resolution 127
5.4 Active Site Components of ResT 128
5.4.1 The Tyrosine Recombinase-Like Component of the Active Site 129
5.4.2 The Hairpin-Binding Module 130
5.5 Reaction Reversal by ResT 131
5.6 ResT's Relationship to Topoisomerases and Site-Specific Recombinases . . 132
5.7 Regulation of ResT Activity In Vivo 133

6 Perspectives 134

References 135

Abstract The genomes of the spirochetes of the genus *Borrelia* are unique among prokaryotes; all member species possess a linear chromosome as well as linear plasmids terminated by covalently closed DNA hairpins, or hairpin telomeres. The usual "end-replication problem" for linear replicons is overcome by the hairpin telomeres presenting to the cellular replication machinery an uninterrupted DNA chain. DNA replication initiates internally and proceeds bidirectionally. Replication through the hairpin turnaround at the telomeres produces replication intermediates that possess inverted repeat replicated telomere junctions, which act as the substrates for a specialized DNA breakage and reunion enzyme referred to as a telomere resolvase. The telomere resolvase converts each replicated telomere into two hairpin telomeres, thereby liberating linear daughter chromosomes from the replicated intermediates. The telomere resolution reaction has mechanistic similarities to those catalysed by type IB topoisomerases and tyrosine recombinases. Details of the replication and telomere resolution processes have been revealed by recent molecular biology and biochemical studies in *Borrelia burgdorferi*, the causative agent of Lyme disease.

1 Introduction

Prokaryotes with genomes composed of linear genetic elements terminated by hairpin (hp) telomeres are typified by the spirochetes of the genus *Borrelia*. Amongst this genus are important human pathogens (Barbour 2001; Schwan et al. 1999) that cause Lyme borreliosis and relapsing fever maladies (Burgdorfer et al. 1982; Dworkin et al. 2002). The most extensively studied member of the species that cause Lyme borreliosis is the North American agent *Borrelia burgdorferi*; the best-characterized relapsing fever spirochete is the European *Borrelia hermsii*. Both Lyme borreliosis and relapsing fever are zoonotic diseases with complex transmission cycles between their mammalian and arthropod hosts. More recently, the plant pathogen and agrobiotechnology tool, *Agrobacterium tumefaciens*, has also been discovered to harbour a linear hairpin chromosome, in addition to a circular chromosome (Goodner et al. 2001).

This chapter will concentrate on *B. burgdorferi*, because the prototype B31 MI strain has been sequenced in its entirety and the enzyme responsible for the formation of the hp telomeres has been described. I will review what has been learned about the replication and segregation of the numerous genetic elements of this organism, as well as summarize what is known about the enzyme involved in formation of the hp telomeres.

2 Lyme Borreliosis

The spirochete bacterium *B. burgdorferi* is a causative agent of Lyme disease in North America (Burgdorfer et al. 1982). Lyme disease is a common vector-transmitted zoonotic infection in North America. *Peromyscus leucopus*, the white-footed mouse, is the most important mammalian reservoir species in eastern North America, while the *Ixodes scapularis* tick is the arthropod vector responsible for horizontal transmission of *B. burgdorferi* (Donahue et al. 1987; LoGiudice et al. 2003). Aside from the white-footed mouse, many other small mammals and even birds can serve as hosts for *B. burgdorferi* (Pal and Fikrig 2003). Initial tick colonization occurs with the larval blood meal in late summer. The spirochetes then survive in the tick gut over the winter. When the tick undergoes a moult the following spring to produce the nymphal stage, *B. burgdorferi* migrates to the salivary glands during the nymphal blood meal. This is when the spirochete is transmitted to the next mammalian host (Steere et al. 2004). Human infections are not part of the transmission cycle. Untreated, Lyme disease can cause serious cardiac and neurological complications; many patients diagnosed late in the course of an infection develop Lyme arthritis (Steere 2001; Steere et al. 1994). Withdrawal

of the transmission-blocking vaccine in 2002 has restricted preventative options (Hanson and Edelman 2003).

B. burgdorferi is an extracellular pathogen that establishes infection by migration through various host tissues, adhesion to host cells and by attenuation and evasion of the immune response (Singh and Girschick 2004; Steere et al. 2004). Infection persistence, and presumably host range, is aided by the *B. burgdorferi* genome encoding multiple but related versions of surface exposed lipoproteins. Many examples of pathogenicity and infective cycle determinants are encoded on the plasmid component of the genome. A fascinating example is the VlsE surface lipoprotein that undergoes extensive antigenic variation mediated by programmed gene conversion. The plasmid-borne telomeric expression site for the VlsE undergoes combinatorial recombination with a diverged and unexpressed contiguous array of pseudogene cassettes located internally on the same plasmid (lp28-1; see Zhang et al. 1997; Zhang and Norris 1998).

An elaborated version of this antigenic variation system gives rise to the eponymous characteristic of the disease that the relapsing fever spirochetes cause (Barbour and Restrepo 2000; Dai et al. 2006). A succinct comparison of the characteristics and role of the antigenic variation systems from *B. burgdorferi* and *B. hermsii* is available in a recent microcommentary (Norris 2006). It is intriguing that in both systems the expression locus is located proximal to a telomere, an attribute that often pertains to the expression loci of parasite contingency gene systems (Barry et al. 2003). The significance of this in *Borrelia* is presently unknown, although it is known that *B. burgdorferi* telomeres are intrinsically recombinogenic (Casjens et al. 2000; Chaconas et al. 2001) and in *B. hermsii* a telomeric location of the expression site may be important for its transcription (Barbour et al. 2000).

3 The *Borrelia burgdorferi* Genome

What has attracted molecular biologists to the study of *B. burgdorferi* is the unusual nature of its genome. Its most outstanding feature is its highly segmented or fragmented nature; the genome of the prototype B31 MI strain is composed of a small (911 kb) linear chromosome, as well as at least 12 linear and nine circular extrachromosomal elements that, collectively, account for 40% of the coding potential of the genome (Casjens et al. 2000; Fraser et al. 1997). This is the greatest number of genetic elements ever found to coexist within a bacterial cell.

In general terms, there appears to be a division of labour between the chromosome and the extrachromosomal elements or "plasmids". Genes encoded on the chromosome tend to be housekeeping genes present in single copy while the genes on the plasmids tend to be *Borrelia*-specific, of unknown

function or involved in the infective cycle, and present in multiple, related copies. Many of these paralogous gene families (PFs) show evidence of undergoing mutational decay. The linear plasmids, especially near the telomeres, appear to be undergoing recent and ongoing genetic exchange (Casjens et al. 2000; Huang et al. 2004b).

The linear replicons are terminated by covalently closed DNA hairpins, referred to as telomeres. *B. burgdorferi* was the first bacterial species to be found with a linear chromosome with hp telomeres (Barbour and Garon 1987; Baril et al. 1989; Ferdows and Barbour 1989). The unique nature of the genome raises several intriguing questions. What is the mode of replication of the linear genetic elements? What is the evolutionary origin of the linearity? Does genome linearity contribute to genome evolution and function? How are so many genetic elements faithfully segregated?

4
DNA Replication

4.1
Mode of Replication and Origins of Replication

The study of DNA replication in *B. burgdorferi* has just recently commenced. What has been established is the position of the replication origin for the linear chromosome and the minimal complement of open reading frames (ORFs) required for autonomous replication of several examples of the circular and linear replicons.

A key study physically mapped the origin of replication of the linear chromosome to a small 2-kb region almost exactly in the centre of the chromosome (Picardeau et al. 1999). In combination with examination of AT- and CG-skew to identify a candidate origin region, a technique called nascent DNA strand analysis was employed. By further analysis of the switch point of the polarity of CG-skew and sequence analysis for origin-like characteristics, the authors conclude that the origin most likely lies in the intergenic region between *dnaA* and *dnaN*. Replication from this origin proceeds bidirectionally. A further important point is that the hp telomeres show no origin activity in the nascent DNA strand assay, eliminating the possibility of their use as replication origins (Picardeau et al. 1999).

The discovery of a specific binding site for the Hbb protein in the *dnaA–dnaN* intergenic region lends further support to the designation of this area as the origin (Kobryn et al. 2000). Hbb is the *B. burgdorferi* homologue of the nucleoid proteins HU and IHF (Tilly et al. 1996). These accessory DNA bending proteins often act in conjunction with DnaA in the open complex formation step of origin firing in other bacterial systems (Hwang and Kornberg 1992; Kornberg 1992; Roth et al. 1994). Hbb binding to its specific site

introduces a sharp U-bend between two flanking direct repeats and induces $KMnO_4$ sensitivity of several T nucleotides outwith the Hbb binding site (Kobryn et al. 2000). These are all biochemical features consistent with a role of Hbb binding to this site in replication initiation.

The bidirectional internal origin and the demonstration of the lack of origin activity of the hp telomeres strongly argued against replication models for the chromosome that do not invoke a DNA breakage and reunion event (telomere resolution) at the replicated telomeres (*rTels*; see Casjens 1999 for a review of replication models for the linear replicons). The lack of hp telomere origin activity also argues against models similar to those proposed for poxvirus replication, which invoke site-specific nicking at one telomere followed by strand displacement synthesis and telomere resolution to process the resulting linear head-to-head, tail-to-tail concatamers into unit length viral genomes terminated by hp telomeres (Fig. 1; see Kobryn and Chaconas 2001, and references therein).

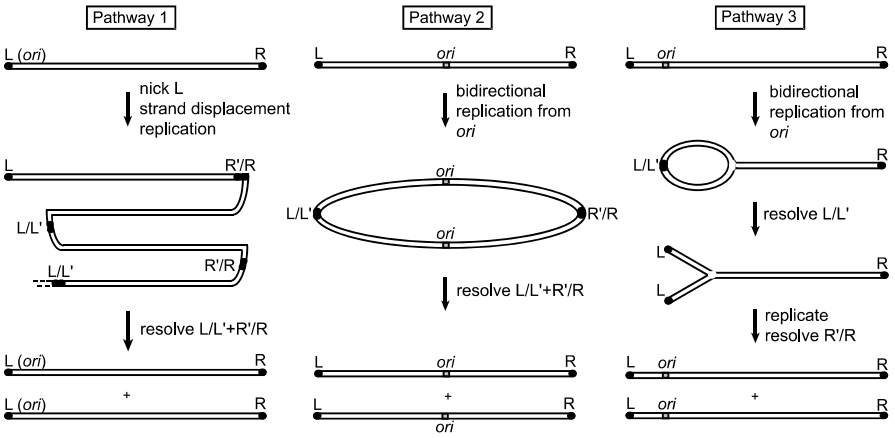


Fig. 1 Replication strategies for linear replicons with hp telomeres. Three replication pathways are shown, each employing a telomere resolution step to liberate unit length linear daughters from replication intermediates. Pathway 1 shows a strategy in which the hp telomere is used as a replication origin; hp nicking at one hp telomere followed by strand displacement synthesis (several rounds of initiation and synthesis ensue) produces arrays of head-to-head, tail-to-tail, linear concatameric genomes. This strategy has been proposed for the Poxviridae (Du and Traktman 1996; Traktman 1996). Pathways 2 and 3 represent strategies employed when a bidirectional internal origin of replication is used. Pathway 2 represents the outcome of replication if both hp telomeres must be replicated before telomere resolution can ensue; an inverted repeat circular dimer intermediate would result. Pathway 3 represents the alternative, in which each replicated telomere junction is resolved as it is formed. For replicons with an asymmetrically disposed origin, Y-shaped replication intermediates would result after one *rTel* had been resolved. Phage N15 has been demonstrated to employ pathway 3 to maintain lysogeny (Ravin et al. 2003). It is not currently known whether *B. burgdorferi* uses pathway 2 or 3 or a mixture of the two pathways depending on the linear replicon

In a subsequent paper, the AT- and CG-skew analysis, which was validated for use in *B. burgdorferi* in the initial study, was further applied to analysis of the remaining linear replicons as well as to some of the circular plasmids (Picardeau et al. 2000). The DNA strand compositional asymmetry of the plasmids is less pronounced than that of the chromosome, but when the skew values are plotted cumulatively against locus position all the linear replicons possess a single point where the polarity of the skew switches; these points represent putative bidirectional origins. The findings suggest bidirectional origins for all the *B. burgdorferi* replicons, linear and circular. In other eubacterial systems typically only the chromosome utilizes bidirectional θ-like replication, while natural plasmids use unidirectional θ-like replication (del Solar et al. 1998).

The chromosomal origin maps to the centre of the chromosome but the CG-skew analysis suggests an asymmetrically disposed origin on some of the linear plasmids (Picardeau et al. 2000). This raises the interesting issue of when the replicated telomere junctions are processed relative to one another. If, as is the case for phage N15, one *rTel* is resolved before the other is even formed, then Y-shaped replication intermediates would result on the linear replicons with off-centre origins (Ravin et al. 2003; see Hertwig 2007 (this volume) for a review of the linear prophage of N15 and other closely related elements). If replication of both hp telomeres must be completed prior to resolution, then inverted repeat circular dimer intermediates would result (Fig. 1). Under conditions of moderate to fast growth bacteria uncouple replication initiation from cell division (i.e. S phase and G2/M overlap). If this also holds true for *B. burgdorferi* then even more complex replication intermediates would result. The modes of cellular regulation of the telomere resolution process would be predicted to be quite different for systems that utilized pathway 2 vs 3 depicted in Fig. 1.

4.2 Autonomous Replication and Plasmid Compatibility Functions

The remainder of the work on *B. burgdorferi* DNA replication has focused on defining the ORFs and other DNA regions required for autonomous replication of several examples of the circular, as well as linear, replicons. This work has been pursued to develop several shuttle vectors for genetic manipulation of the spirochete. Another motivation for pursuing these studies has been the recognition that *B. burgdorferi* represents a superb model system in which to study issues of plasmid biology. The reasons that it constitutes such a good system are twofold: the plasmids represent essential genetic elements for the infective cycle of this pathogen, and the sheer number of co-resident plasmids should make it an ideal model in which to study plasmid replication, partition and compatibility (see Stewart et al. 2005).

The plasmids have been reported to be maintained in the same copy number as the chromosome (Hinnebusch and Barbour 1992; Morrison et al. 1999).

The initial experimental evidence delineating the genetic elements required for plasmid maintenance came from the development of the first *B. burgdorferi–Escherichia coli* shuttle vector pBSV2 (Stewart et al. 2001). A small 3.3-kb region from circular plasmid cp9 containing three ORFs confers replication proficiency in the spirochete of a pBluescript-based vector modified to carry a Kan^R cassette functional in both *E. coli* and *B. burgdorferi*. This region of cp9 was tested because it contains ORFs that belong to a cluster of PF members (PF32, 49, 50 and 57/62) that are found clustered together, in various combinations, on all the other plasmids, leading to the prediction that these ORFs would represent functions required for replication and partition (Casjens et al. 2000).

Further experimental evidence testing this prediction has come from studies utilizing similar regions from cp32, cp26, lp25, lp28-1 and lp17 to construct further shuttle vectors (Beaurepaire and Chaconas 2005; Byram et al. 2004; Eggers et al. 2002; Stewart et al. 2003). The minimum in "*cis*" requirements for autonomous replication differ from one example to another, but in some cases can be reduced to either a PF57 or PF62 gene member and a small amount of nearby flanking sequence (Beaurepaire and Chaconas 2005; Eggers et al. 2002). A PF57 or a member of the homologous PF62 family is always represented in the minimal required regions, leading to the suggestion that they may encode alternative versions of a plasmid initiator or Rep function (Beaurepaire and Chaconas 2005). Bioinformatics sheds no light on this issue since the PF49, 50, 57 and 62 gene families have no similarity to ORFs found outside the genus *Borrelia* (Casjens et al. 2000). Similarly, sequence comparisons of the known and candidate origin sequences identified show no compelling homology to each other or to characterized bacterial origins (Stewart et al. 2003). An exception to this lack of homology to other bacterial proteins are the PF32 paralogues, which have been noted to be similar to ParA of phage P1 (Zuckert and Meyer 1996). ParA, in conjunction with ParB, which *B. burgdorferi* seems to lack a homologue of, is involved in segregation of the P1 prophage genome (Ebersbach and Gerdes 2005; Li et al. 2004). In a cp32-derived shuttle vector the PF32 paralogue mediates plasmid incompatibility (Eggers et al. 2002). However, the PF32 paralogue on lp17 can be deleted without an effect on incompatibility or plasmid stability (Beaurepaire and Chaconas 2005).

An interesting observation that has come out of the studies defining DNA regions required for autonomous replication of the plasmids is that some such areas can function in both circular and linear contexts. The minimal region from cp9 can function in a linear context, albeit with reduced copy number (Chaconas et al. 2001). Conversely, the minimal regions for replication from the linear plasmids lp25 and lp28-1 can function in a circular context (Stewart et al. 2003). On the other hand, the corresponding minimal region of lp17 demonstrates a bias for function in its native linear context; circular versions require a lp17 helper (Beaurepaire and Chaconas 2005, 2006). Additionally,

the spontaneous recovery of a stable circular version of a linear plasmid in *B. hermsii* further underscores the unity of the replication functions utilized by the linear and circular replicons (Ferdows et al. 1996). Where a bias for maintenance of one topological state over another has been noted, it appears that a supercoiling effect on the transcription of plasmid maintenance genes plays a role in this bias (Beaurepaire and Chaconas 2006).

Dissection of the determinants of plasmid stability and (in)compatibility has, to date, been confounded by the possibility of the donation in *trans* of some of these functions from other plasmids and by the presence on two linear plasmids of putative restriction/modification systems (Kawabata et al. 2004; Lawrenz et al. 2002). As for the mechanism of maintaining compatibility of so many genetic elements within the same bacterial cell, one wonders whether it might be related to the simple parameter of cell length. A typical bacterium like *E. coli* is $\sim$2 μm in length. Plasmid compatibility in other bacterial species may be determined, at least in part, by compatible plasmids residing at unique subcellular "addresses" along the long axis of the cell (Ho et al. 2002; Pogliano 2002). *Borrelia* spirochetes are unusually long and thin cells (up to $\sim 25 \times 0.5$ μm) potentially affording many more unique addresses for multiple genetic elements along the cell axis.

5
Telomere Resolution

5.1
In Vivo Demonstration of Telomere Resolution

The discovery of a bidirectional, internal origin of replication for the *B. burgdorferi* chromosome, coupled with the absence of detectable initiation from the hp telomeres, strongly implied that replication would eventually proceed around the hairpin turns to produce inverted repeat replicated telomere (*rTel*) junctions. A significant advance in our understanding of the mode of replication of the linear genetic elements came from the subsequent demonstration in vivo that such *rTel* junctions are processed by a DNA breakage and reunion reaction, referred to as telomere resolution, to produce two hp telomeres (Fig. 1; Chaconas et al. 2001). In this study it was demonstrated that integration of a synthetic *rTel* into lp17, near the left telomere, creates a substrate for a cellular telomere resolvase activity that acts at the *rTel* to produce a deletion derivative of lp17 containing a bona fide hp telomere derived from the synthetic junction. The part of lp17 deleted from the transient lp17 integrant species is lost from the cell, presumably due to the lack of replication functions on the resulting fragment, providing further evidence that hp telomeres cannot serve as origins of replication. More evidence that the synthetic *rTel*s are resolved into hp telomeres comes from

an experiment which showed that cloning of a synthetic *rTel* into a circular shuttle vector (pBSV2) replicating in *B. burgdorferi* is sufficient to convert the shuttle vector into a linear replicon terminated by hp telomeres (Chaconas et al. 2001). Telomere resolution in vivo is sequence specific; a mock *rTel* junction with the same inverted repeat symmetry and sequence composition of a real substrate, but that is of non-telomeric sequence, is not a substrate for the telomere resolvase. Another noteworthy result is that integration of an *rTel* into lp17 is accompanied by a 7- to 15-fold stimulation of the recovery of transformants, implying that either *rTel* junctions or hp telomeres are intrinsically recombinogenic. This effect is dependent upon the *rTel* being resolvable, suggesting that the stimulation of recombination is a consequence of the action of the telomere resolving machinery (Chaconas et al. 2001).

5.2 The Telomere Resolvase (ResT)

Coincident with the work describing a telomere resolution step in vivo in *B. burgdorferi*, the telomere resolvase from the *E. coli* phage N15 was discovered (Deneke et al. 2000). It had been noted previously that the BBB03 locus of *B. burgdorferi* had an area of sequence homology with TelN, the protein later confirmed as the N15 telomere resolvase (Rybchin and Svarchevsky 1999). See Hertwig 2007 (this volume) for a review of telomere resolution in phage N15 and other related elements.

The *Borrelia* enzyme that performs telomere resolution was confirmed to be the product of the BBB03 locus of the circular plasmid (cp26) and was named ResT (for Resolvase of Telomeres; Kobryn and Chaconas 2002). The activity of ResT was assayed in vitro on plasmids carrying *rTel* junctions established to be in vivo substrates. Provided that the negative supercoils of the purified plasmids are removed by prior nicking, relaxation or linearization, ResT resolves these *rTel*s into two covalently closed hp telomeres. The reaction is conservative, site-specific and requires no accessory factors, divalent metal ions or high-energy cofactors.

ResT, TelN and the other confirmed or suspected telomere resolvases share sequence motifs in common with the tyrosine recombinase family of site-specific recombinases and with the type IB topoisomerases (see Fig. 2; Deneke et al. 2004; Kobryn and Chaconas 2002). Both of these enzyme families share a common chemical mechanism of DNA phosphoryl transfer that proceeds by a two-step transesterification. The first step, which stores the DNA bond energy, involves the formation, from cleavage of the DNA by an active site tyrosine residue, of a covalent enzyme–DNA intermediate. The hydroxyl group of the cleaved DNA strand executes the second transesterification by attacking the phosphotyrosyl bond of the enzyme–DNA complex. This reseals or transfers the DNA strand (see Mizuuchi 2002 for a general discussion of

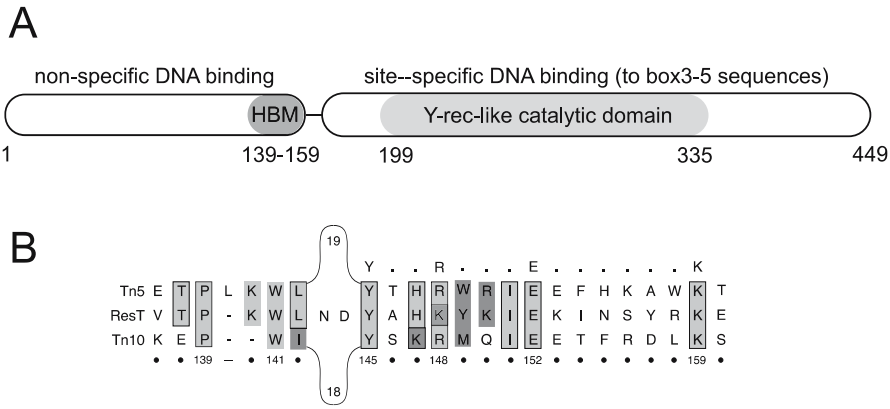


Fig. 2 The characterized components of ResT. **A** The hairpin-binding module (HBM) and the tyrosine recombinase-like catalytic region overlaid on a domain representation of ResT as defined by partial protease digestion (Tourand et al. 2007). The *numbers* represent amino acid numbers. **B** Sequence alignment of ResT with the transposases of Tn5 and Tn10 highlighting the HBM of ResT. The YREK residue conserved in IS4 elements (Rezsohazy et al. 1993) is noted above the alignment. The hydrophobic pocket residues are located to the *left* of the loops indicating insertions in the transposases, and those of the YREK segment to the *right*. *Light shading* indicates sequence identity and similarity is shown by *dark shading*. The *numbering* below the schema indicates ResT amino acid numbers. Adapted from Bankhead and Chaconas (2004) with permission from the National Academy of Sciences USA

chemical mechanisms). The reactions of both enzyme families are characterized by formation of a 3′-phosphotyrosyl enzyme–DNA intermediate and 5′-OH ends on the cleaved DNA (Gopaul and Duyne 1999; Grainge and Jayaram 1999; Hallet and Sherratt 1997).

To investigate whether the sequence relationship of part of ResT to the catalytic domains of the tyrosine recombinases and topoisomerases indicated that telomere resolution proceeds through a similar mechanism, evidence of a covalent ResT–DNA intermediate was sought. Such a complex is formed and has the expected 3′ polarity. The overall reaction and formation of the covalent ResT–DNA complex is dependent upon the presence of the OH group of the tyrosine residue predicted from the sequence alignment to be the active site nucleophile (Y335). Mutation of this residue to phenylalanine abolishes both. The position of the cleavage sites on the *rTel* junction map to positions 3 nt from the axis of symmetry. The resulting 6-bp 5′ stagger between the cleavage sites in the *rTel* falls within the 6–8-bp range found for all currently characterized tyrosine recombinases. One can infer from the simple buffer requirements of the reaction, along with the observation of a 3′-ResT–DNA complex that has the kinetic and reactive properties of a reaction intermediate, that telomere resolution catalysed by ResT also proceeds via a two-step transesterification (Kobryn and Chaconas 2002).

5.3 Substrate Requirements for Telomere Resolution

Next, the substrate requirements for telomere resolution in vitro were examined (Tourand et al. 2003). This work was guided by an alignment of all the telomeres sequenced at that time from *B. burgdorferi*, *B. afzelii* and *B. hermsii* (Casjens 1999, and references therein). From that alignment it is apparent that all the sequenced telomeres share five boxes of homology, accounting for 14 bp of the terminal 23 bp. Starting from this foundation the minimal size of an *rTel* derived from the left telomere of plasmid lp17 (Hinnebusch and Barbour 1991) was determined to be 38 bp from 19 bp in the unreplicated hp telomere form. Systematic introduction of transversions at each of the 14 conserved base pairs within the model lp17 *rTel* identifies box 3 as especially important for resolution (see Fig. 3). Box 3 is located 15–19 bp from the symmetry axis of the *rTel* and is in the position that normally directs sequence-specific binding of tyrosine recombinases to their respective sites. Full-length ResT does not bind its substrate in a gel-shift assay (other than by aggregation) but the C-terminal domain, as defined by partial protease digestion (aa∼160–449), contains the tyrosine recombinase-like domain and the part of ResT required for site recognition. Box 3–5 sequences are sufficient for the site-specific gel shift

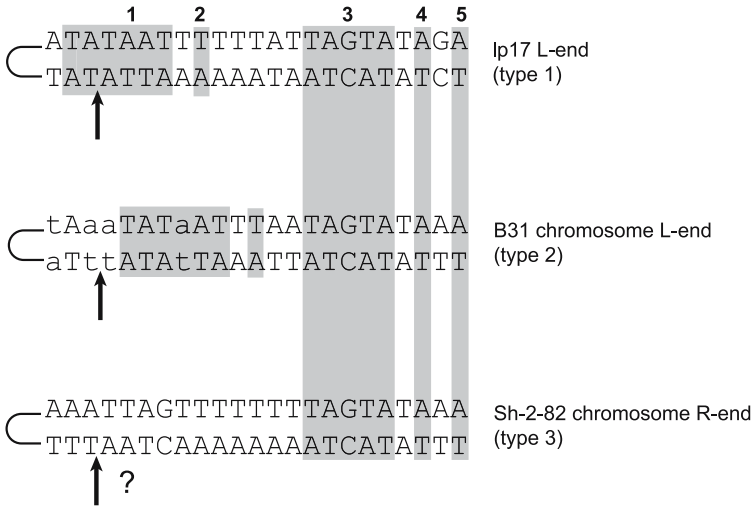


Fig. 3 The three types of hp telomeres in *B. burgdorferi*. A representative of each type is shown and the conserved sequence boxes are shaded and numbered. Positions with known sequence variability within box 1 or near the hp turnaround are indicated in *lower case*. The *arrows* indicate the position of cleavage by ResT. This is not yet determined for the type 3 telomere, so the expected position of cleavage is indicated with a "?" mark. Adapted from Chaconas (2005) with permission from Blackwell Publishing

seen with the isolated C-terminal domain (Tourand and Chaconas 2007; Jayaram 2007).

Another feature highlighted by the alignment of the known hp telomeres is the existence of two types of hp telomere that differ by the spacing of the box 1 sequence relative to the symmetry axis and to boxes 3–5. A third telomere type without any discernable box 1 has been described (Huang et al. 2004b) and is designated as a type 3 telomere (see Fig. 3 for a summary of the telomere types). The existence of multiple telomere types raises the possibility that there may be a requirement for more than one telomere resolving activity. Types 1 and 2 and many different substrates in which the disposition of box 1 relative to boxes 3–5 and the symmetry axis were constructed and assayed for resolution by ResT in vitro. ResT was found to efficiently process only the type 1 and type 2 spacing variants (Tourand et al. 2003). This result presented somewhat of a conundrum since in the type 1 *rTel* the scissile phosphates lay within box 1, while for the type 2 substrate the position of the cleavage sites would have to move from being 6 to 12 bp apart if box 1 sequences defined the position of cleavage (see Fig. 3). The position of the cleavage sites in the type 2 and a further poorly resolved spacing variant map to 6 bp apart around the symmetry axis of the *rTel*, indicating that site-specific binding of ResT to boxes 3–5, whose position relative to the symmetry axis remains fixed, positions the active sites for catalysis. Nonetheless, replacement of box 1 sequences in either the type 1 or type 2 substrates leads to a defect of activity in vitro. Mutation of the same sequence, which occurs in the otherwise unrelated sequence of the substrate for the N15 resolvase, TelN, also results in defective resolution; this indicates that the box 1 sequence is important for resolution in both systems (Deneke et al. 2002). Nevertheless, a type 3 *rTel* is also resolved in vitro, indicating that mutation of the telomere to the particular sequence found at the corresponding box 1 position is tolerated (Tourand et al. 2006). It may prove to be the case that any sequence around the cleavage site with the appropriate helical parameters for cross-axis complex formation and DNA hairpin formation (see Sect. 5.7) may be tolerated and that box 1 sequences are conserved at some telomeres as part of a promoter element (Chaconas 2005). Clearly, more hp telomeres must be sequenced and the role of the box 1 sequences given further study.

5.4 Active Site Components of ResT

ResT shares a common chemical mechanism with the tyrosine recombinases and the type IB topoisomerases. As noted in Fig. 2, part of the C-terminal domain of ResT can be aligned with the catalytic domains of these proteins. However, there is also a contribution from part of the N-terminal domain for activating catalysis by a "hairpin-binding module" that seems to pre-

distort the DNA between the scissile phosphates to activate DNA cleavage and to facilitate rapid hairpin formation after the cleavage step (Bankhead and Chaconas 2004). It is the interplay between these two different active site components that accounts for the unique reaction of telomere resolution. I will examine each active site component in turn.

5.4.1 The Tyrosine Recombinase-Like Component of the Active Site

For both the tyrosine recombinases and the type IB topoisomerases the attacking nucleophile for DNA cleavage is an active site tyrosine. In each enzyme family there is also a constellation of active site residues involved in activating cleavage of the phosphodiester bond in the DNA and of the phosphotyrosyl bond in the enzyme–DNA complex, through transition state stabilization and facilitation of leaving group expulsion. For the tyrosine recombinases it consists of a conserved pentad RKHR(H/W), while for the type IB topoisomerases it is a conserved tetrad of RKR(H/N) (Perry et al. 2006; Shuman 1998; Tian et al. 2004; Van Duyne 2001).

Changing the putative active site tyrosine (Y335) of ResT to phenylalanine produced a catalytically dead protein (Kobryn and Chaconas 2002). The assignment of Y335 as the active site nucleophile was confirmed by trapping the phosphotyrosyl ResT–DNA intermediate with a biotinylated *rTel* suicide substrate. The resulting complex was subjected to tryptic digestion, the phosphopeptide–DNA fragment purified with streptavidin-conjugated magnetic beads, the DNA component of the complex P1 nuclease-digested and the resulting purified phosphopeptide sequenced by tandem mass spectroscopy (Deneke et al. 2004).

Systematic mutagenesis of putative active site residues based on alignment of the telomere resolvases with the tyrosine recombinases, type IB topoisomerases and themselves was performed by alanine scanning, except in the case of tyrosines which were changed to phenylalanines. As expected from the alignment mutation of R199, K224, Y293 and R296, it produces a resolvase still proficient for DNA binding but defective for telomere resolution. The last amino acid of the pentad (H/W) is H324 of ResT which appears to be dispensable for the reaction. This side chain is generally thought to play a catalytic role in the tyrosine recombinases but there is one example, the yeast Flp recombinase, in which the corresponding W330 plays a purely structural role (Chen and Rice 2003).

ResT R199 and K224 are side chains that correspond to the general acid catalyst (R199 donates a proton to K224 which donates the proton to the leaving group) involved in leaving group expulsion for the tyrosine recombinases and type IB topoisomerases, for which this contention has been tested (Ghosh et al. 2005; Krogh and Shuman 2000, 2002). While these residues do appear to be important for catalysis, they fail to show convincing evi-

dence (at least the mutations to alanine tested) of being directly involved in proton donation to the DNA leaving group during cleavage. None of the other mutants tested in this study met the criteria of a general acid catalyst. Mutagenesis of two additional residues conserved amongst the telomere resolvases, but with no counterpart in the recombinase or topoisomerase families (R198 and Y299), suggest their possible involvement in catalysis (Deneke et al. 2004).

5.4.2 The Hairpin-Binding Module

An area of ResT near the end of the N-terminal domain has sequence similarity to an important motif present in bacterial cut-and-paste transposases (see Fig. 2). Cut-and-paste transposases mobilize transposons like Tn5 or Tn10 via complete excision of the element from the chromosome by the introduction of double-strand breaks into the transposon ends and by their subsequent transfer to new DNA sites (Haniford 2002; Reznikoff 2003). Extensive biochemical and structural work has identified the "YREK motif", in conjunction with a constellation of active site residues typical of the RNaseH family of integrases/nucleases, as being responsible for transposon excision (Allingham et al. 2001; Ason and Reznikoff 2002; Davies et al. 2000; Lovell et al. 2002; Naumann and Reznikoff 2002). These transposases break four DNA strands utilizing only two active sites. This is accomplished by transition through a DNA hairpin intermediate (Kennedy et al. 1998, 2000) at each transposon end followed by hydrolytic opening of the hairpins (which are formed on the transposon side of the Tn/chromosome junctions). Creating the DNA distortion required for hairpin formation is the role of the YREK motif (Allingham et al. 2001; Ason and Reznikoff 2002; Davies et al. 2000; Rice and Baker 2001). The YREK motif has two separable components: a hydrophobic pocket that forms stacking interactions with the penultimate base of the transposon end which is flipped out of the double helix, contributing most of the distortion required for changing the trajectory of the DNA strands to form the hairpin, and the YREK residues that make stabilizing contacts with the distorted DNA about to be hairpinned.

Mutation of the corresponding hydrophobic pocket in ResT (P139A, W141A and the double mutant) leads to a severely compromised resolvase; mutation of the YKEK motif produces a cold-sensitive defect (Bankhead and Chaconas 2004). Importantly, mutants in the hairpin-binding module of ResT, as this region has come to be known, show a defect at the DNA cleavage step, unlike the transposase mutants that manifest their failure to make DNA hairpins as a strand transfer defect (Allingham et al. 2001; Bankhead and Chaconas 2004). A pre-cleavage distortion of the *rTel* DNA between the scissile phosphates by the hairpin-binding module is inferred from the ob-

servation that introduction of a heteroduplex into the central 2 bp of the *rTel* leads to a rescue of the cleavage defect of these mutants (DNA hairpinning then proceeds normally). The rationale is that the small region of heteroduplex around the symmetry axis of the *rTel* mimics the pre-cleavage action of this part of the protein.

5.5 Reaction Reversal by ResT

Like its bacteriophage counterparts (Huang et al. 2004a, ResT appears not to turn over during the course of the in vitro reaction. This raised the possibility of very slow product release or product inhibition of resolution. In the course of investigating this issue ResT was found to be able to interact with its hp telomere products. Unexpectedly, ResT is also able to cleave its product and to fuse hp telomeres together in a chemical reversal of the resolution reaction (Kobryn and Chaconas 2005). The properties of the reverse reaction suggest that reaction reversal occurs after product release from the resolution reaction, rather than in the context of a post-resolution synaptic complex of the hp telomeres. Temperature plays a key role in regulating the extent of reaction reversal observed. Low-temperature incubation (8 °C) suppresses the forward reaction more than the reverse reaction, allowing for an accumulation of the reversal product (an *rTel*). This effect is likely mediated by the presence in the hp telomeres of the pre-distorted structure near the cleavage site (i.e. the hairpin turnaround). Formation of this structure by the hairpin-binding module confers cold sensitivity to the cleavage step of the resolution (forward) reaction.

At first sight, telomere fusion seems like it should be heavily disfavoured in vitro due to the entropic cost of bringing two hp telomeres together for their fusion. However, the presence in the product of the hp imparts a higher energy state to the hp telomeres. This finding, combined with the fact that ResT binds to the hp products with equal affinity to that for the *rTel*, as opposed to higher affinity, somewhat offsets the large bias the reaction would normally display against reversal (Kobryn and Chaconas 2005). The ability of ResT to cleave and fuse its hp telomere products raises the issue of how these activities are restrained in vivo. Frequent hp telomere cleavage or telomere fusions followed by mutation or deletion of the resulting junctions would lead to high levels of inviability or genome instability (i.e. duplications, translocations etc.). The linear replicons of *B. burgdorferi* appear to be undergoing just these kinds of events, especially near the telomeres (Casjens et al. 2000; Huang et al. 2004b). It has been proposed that action of the telomere resolvase on its product may explain these events (for a discussion of this issue, see Chaconas 2005; Kobryn and Chaconas 2005).

5.6 ResT's Relationship to Topoisomerases and Site-Specific Recombinases

Topoisomerases, site-specific recombinases and telomere resolvases share a common chemical mechanism; however, the gross nature of their reactions differ because the number of DNA strands cleaved and the identity of the strands they are joined to differ for each class of enzyme. Topoisomerases cleave one (types I and III) or two strands (types II and IV) which are rejoined after rotation or strand passage to change the levels of cellular DNA supercoiling or catenation (Champoux 2001; Shuman 1998; Wang 1996). Site-specific recombinases (both serine and tyrosine classes) cleave four DNA strands and join them to new partners to produce recombinant products. Telomere resolution is a unique reaction whose complexity lies midway in this continuum. Here, two DNA strands are cleaved and rejoined to the opposing complementary strand. The fact that strand transfer occurs to a new partner strand makes the resolution reaction somewhat akin to a recombination reaction in which strand transfer occurs within the *rTel* substrate rather than to a recombination partner duplex (Kobryn et al. 2005). These differences in the global nature of the reactions catalysed are accompanied by different requirements for the coordination of active sites and for the timing of the cleavage and strand joining steps on each strand of the target site.

It was, therefore, of interest to investigate these issues for both the telomere resolution and fusion reactions in order to determine the depth of the apparent relationship of ResT to the type IB topoisomerases, on the one hand, and to the tyrosine recombinases on the other. In contrast to the type IB topoisomerases which act as monomers, ResT is not active on isolated half-sites or hp telomeres. ResT–ResT communication through, at the very least, an in-line synapse of such sites mimicking the structure of the *rTel* substrate seems to be required to activate DNA cleavage (Kobryn et al. 2005).

Tyrosine recombinases catalyse reactions in which two sites are brought together followed by execution of pairwise DNA cleavage and strand transfer reactions of the equivalent strand in each duplex to form a Holliday junction intermediate. The junction is then isomerized to activate the remaining pair of active sites for cleavage and strand transfer of the other pair of strands to produce the recombinant product (Van Duyne 2002). In contrast to this, ResT is inferred to catalyse nearly simultaneous cleavage and strand transfer reactions on the two halves of its substrate. Despite this, for both steps in both the forward and reverse reactions, the chemical steps can be uncoupled from each other. This unusual combination of characteristics is consistent with the pre-cleavage action of the hairpin-binding module of ResT imposing a commitment to completion of the cleavage and strand transfer cycle (Kobryn et al. 2005).

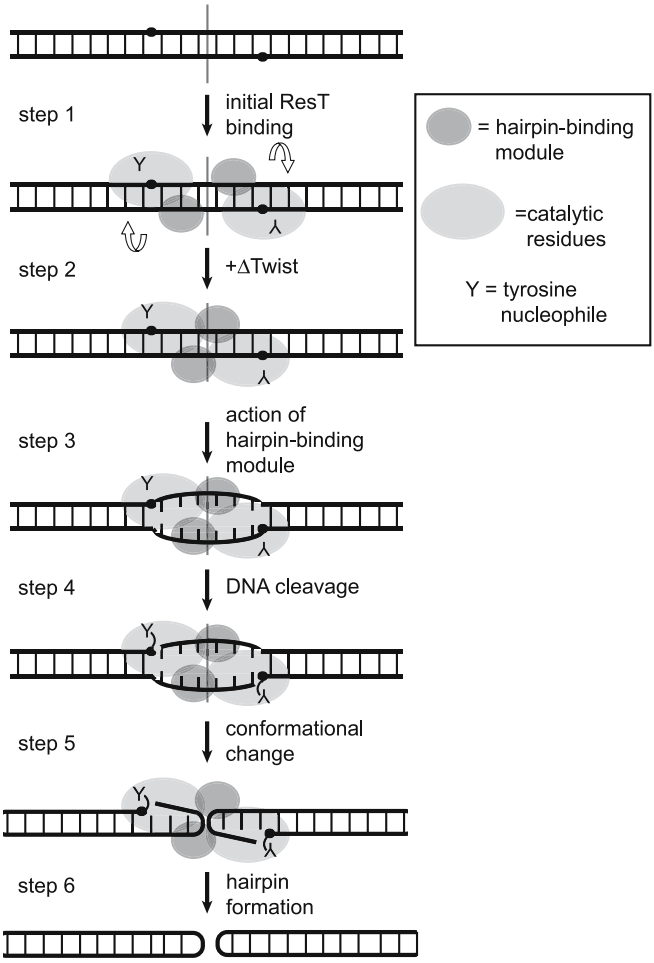
The pattern of active site residues and the reaction characteristics just described above are suggestive of a situation in which part of an ancestral cut-

and-paste transposase was fused to part of an ancestral tyrosine recombinase resulting in this new activity of telomere resolution. Conversion or co-option of either a plasmid/chromosome dimer resolving enzyme (i.e. XerC/(D)) or a bacteriophage integrase (e.g. λ integrase) in this manner in a bacterium with a circular genome could result in the genome's linearization and fragmentation. Discovery of further ancestral properties in ResT and its reactions from either partner in the proposed fusion would lend further support to this hypothesis.

5.7 Regulation of ResT Activity In Vivo

As expected from its central role in the replication of the linear replicons in *B. burgdorferi*, ResT is an essential cellular activity. Efforts to knock out ResT always result in merodiploidy (Byram et al. 2004) unless a homologue from *B. hermsii* is first provided to the cell (Tourand et al. 2006). Despite the successful replacement of *B. burgdorferi* ResT with *B. hermsii* ResT and the normal growth and genome maintenance in such strains, *B. hermsii* ResT was found to be unable to process the type 2 *rTel*s in vitro. Closer examination of some of the inactive substrates assayed earlier in vitro for *B. burgdorferi* ResT also revealed that they are functional in the context of the spirochete (Tourand et al. 2006). This suggests the action of either a redundant function (notwithstanding the essential nature of ResT) or, more likely, the participation of one or more accessory factors in vivo.

Recent results suggest a completely unexpected candidate: positive DNA supercoiling (Bankhead et al. 2006). In vitro results reveal a strong stimulation of telomere resolution by *B. burgdorferi* ResT and a complete rescue of the resolution defect of *B. hermsii* ResT on a type 2 *rTel*. Negative supercoiling has an opposing effect; a progressive inhibition of resolution occurs with increasing levels of negative supercoiling. Biochemical studies indicate that the stimulation from the positive DNA supercoiling is conferred by promoting a very early reaction step: formation of ResT–ResT interactions in a cross-axis complex inferred to exist from the studies discussed in Sect. 5.6 (see Fig. 4 for a schematic summary cartoon of the telomere resolution reaction). The results are most consistent with this interaction being promoted by the twist component of the positive supercoiling (Bankhead et al. 2006). Negative supercoiling plays an important role in many aspects of DNA metabolism in bacteria, including replication, transcription, packaging, recombination and transposition (Kanaar and Cozzarelli 1992). Until now positive DNA supercoiling has only been viewed in eubacteria as a by-product of processes such as transcription and replication that must be removed by cellular topoisomerases. These results raise the intriguing possibility that the energy of positive DNA supercoiling per se may be harnessed to activate/control telomere resolution, linking telomere resolution to some process that produces



positive twist or supercoiling in the spirochete (e.g. replication or transcription converging on, as yet, unresolved replicated telomere junctions). Future studies will be required to test this hypothesis, since it is also possible that the positive DNA supercoiling in vitro is merely substituting for an unidentified accessory factor. Figure 4 presents a summary model of the telomere resolution reaction catalysed by ResT.

6 Perspectives

One fully expects that continued study of telomere biology and plasmid replication/segregation at the molecular biology and biochemical level will

◀ **Fig. 4** Schematic summary of the telomere resolution reaction catalysed by ResT. Initial binding of ResT (step 1) to the *rTel* in a substrate lacking positive supercoiling is non-cooperative (Bankhead et al. 2006). Introduction of positive twist by some resolution-coupled cellular process in vivo or by thermal fluctuation or reverse gyrase in vitro facilitates ResT–ResT communication and the formation of a cross-axis complex (step 2). No specific architecture for the interactions in the cross-axis complex should be inferred by the schematic depiction presented. In step 3 of the reaction the hairpin-binding module (HBM) induces a distortion of the DNA between the cleavage sites. This is represented as strand separation between the cleavage sites due to the fact that heteroduplexed substrates rescue HBM mutants. The actual structural nature of the distortion induced has yet to be characterized. The action of the HBM licenses cleavage of both DNA strands (step 4) and facilitates the conformational change (step 5) required for hairpin formation (step 6). Once the DNA strands are cleaved, steps 5 and 6 follow without the possibility of resealing strands in the parental configuration (Kobryn et al. 2005). This approach to hairpin formation may be imposed by dissipation of the positive twist used to build the cross-axis complex. The hp telomeres are depicted after product release. The time frame for this step of the reaction is unknown

continue to shed light on such diverse topics as the origin and role of genome linearity, genome evolution and pathogenesis in the fascinating spirochetes of the genus *Borrelia*.

Acknowledgements KK is supported by a grant from the Canadian Institutes of Health Research (CIHR). I would like to thank George Chaconas for communication of unpublished results and Brendan Bell for critical reading of the manuscript.

References

Allingham JS, Wardle SJ, Haniford DB (2001) Determinants for hairpin formation in Tn10 transposition. EMBO J 20:2931–2942

Ason B, Reznikoff WS (2002) Mutational analysis of the base flipping event found in Tn5 transposition. J Biol Chem 277:11284–11291

Bankhead T, Chaconas G (2004) Mixing active site components: a recipe for the unique enzymatic activity of a telomere resolvase. Proc Natl Acad Sci USA 101:13768–13773

Bankhead T, Kobryn K, Chaconas G (2006) Unexpected twist: harnessing the energy in positive supercoils to control telomere resolution. Mol Microbiol 62:895–905

Barbour AG (2001) *Borrelia*: a diverse and ubiquitous genus of tick-borne pathogens. In: Scheld MW, Craig WA, Hughes JM (eds) Emerging infections 5. American Society for Microbiology, Washington DC, pp 153–173

Barbour AG, Garon CF (1987) Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. Science 237:409–411

Barbour AG, Restrepo BI (2000) Antigenic variation in vector-borne pathogens. Emerg Infect Dis 6:449–457

Barbour AG, Carter CJ, Sohaskey CD (2000) Surface protein variation by expression site switching in the relapsing fever agent *Borrelia hermsii*. Infect Immun 68:7114–7121

Baril C, Richaud C, Baranton G, Saint Girons IS (1989) Linear chromosome of *Borrelia burgdorferi*. Res Microbiol 140:507–516

Barry JD, Ginger ML, Burton P, McCulloch R (2003) Why are parasite contingency genes often associated with telomeres? Int J Parasitol 33:29–45

Beaurepaire C, Chaconas G (2005) Mapping of essential replication functions of the linear plasmid lp17 of *B. burgdorferi* by targeted deletion walking. Mol Microbiol 57(1):132–142

Beaurepaire C, Chaconas G (2006) Topology-dependent transcription in linear and circular plasmids of the segmented genome of *B. burgdorferi*. Mol Microbiol 63(2):443–453

Burgdorfer W, Barbour AG, Hayes SF, Benach JL, Grunwaldt E, Davis JP (1982) Lyme disease: a tick-borne spirochetosis? Science 216:1317–1319

Byram R, Stewart PE, Rosa P (2004) The essential nature of the ubiquitous 26-kilobase circular replicon of *Borrelia burgdorferi*. J Bacteriol 186:3561–3569

Casjens S (1999) Evolution of the linear DNA replicons of the *Borrelia* spirochetes. Curr Opin Microbiol 2:529–534

Casjens S, Palmer N, Van Vugt R, Huang WH, Stevenson B, Rosa P, Lathigra R, Sutton G, Peterson J, Dodson RJ et al (2000) A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. Mol Microbiol 35:490–516

Chaconas G (2005) Hairpin telomeres and genome plasticity in Borrelia: all mixed up in the end. Mol Microbiol 58:625–635

Chaconas G, Stewart PE, Tilly K, Bono JL, Rosa P (2001) Telomere resolution in the Lyme disease spirochete. EMBO J 20:3229–3237

Champoux JJ (2001) DNA topoisomerases: structure, function, mechanism. Annu Rev Biochem 70:369–413

Chen Y, Rice PA (2003) The role of the conserved Trp330 in Flp-mediated recombination. Functional and structural analysis. J Biol Chem 278:24800–24807

Dai Q, Restrepo BI, Porcella SF, Raffel SJ, Schwan TG, Barbour AG (2006) Antigenic variation by *Borrelia hermsii* occurs through recombination between extragenic repetitive elements on linear plasmids. Mol Microbiol 60:1329–1343

Davies DR, Goryshin IY, Reznikoff WS, Rayment I (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. Science 289:77–85

del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M, Diaz-Orejas R (1998) Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev 62:434–464

Deneke J, Ziegelin G, Lurz R, Lanka E (2000) The protelomerase of temperate *Escherichia coli* phage N15 has cleaving–joining activity. Proc Natl Acad Sci USA 97:7721–7726

Deneke J, Ziegelin G, Lurz R, Lanka E (2002) Phage N15 telomere resolution: target requirements for recognition and processing by the protelomerase. J Biol Chem 277:10410–10419

Deneke J, Burgin AB, Wilson SL, Chaconas G (2004) Catalytic residues of the telomere resolvase ResT: a pattern similar to, but distinct from, tyrosine recombinases and type IB topoisomerases. J Biol Chem 279:53699–53706

Donahue JG, Piesman J, Spielman A (1987) Reservoir competence of white-footed mice for Lyme disease spirochetes. Am J Trop Med Hyg 36:92–96

Du S, Traktman P (1996) Vaccinia virus DNA replication: two hundred base pairs of telomeric sequence confer optimal replication efficiency on minichromosome templates. Proc Natl Acad Sci USA 93:9693–9698

Dworkin MS, Schwan TG, Anderson DE Jr (2002) Tick-borne relapsing fever in North America. Med Clin North Am 86:417–433

Ebersbach G, Gerdes K (2005) Plasmid segregation mechanisms. Annu Rev Genet 39:453–479

Eggers CH, Caimano MJ, Clawson ML, Miller WG, Samuels DS, Radolf JD (2002) Identification of loci critical for replication and compatibility of a *Borrelia burgdorferi* cp32-based shuttle vector for the expression of fluorescent reporters in the Lyme disease spirochete. Mol Microbiol 43:281–295

Ferdows MS, Barbour AG (1989) Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. Proc Natl Acad Sci USA 86:5969–5973

Ferdows MS, Serwer P, Griess GA, Norris SJ, Barbour AG (1996) Conversion of a linear to a circular plasmid in the relapsing fever agent *Borrelia hermsii*. J Bacteriol 178:793–800

Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK et al (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature 390:580–586

Ghosh K, Lau CK, Gupta K, Van Duyne GD (2005) Preferential synapsis of loxP sites drives ordered strand exchange in Cre-loxP site-specific recombination. Nat Chem Biol 1:275–282

Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Qurollo B, Goldman BS, Cao Y, Askenazi M, Halling C et al. (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science 294:2323–2328

Gopaul DN, Duyne GD (1999) Structure and mechanism in site-specific recombination. Curr Opin Struct Biol 9:14–20

Grainge I, Jayaram M (1999) The integrase family of recombinase: organization and function of the active site. Mol Microbiol 33:449–456

Hallet B, Sherratt DJ (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. FEMS Microbiol Rev 21:157–178

Haniford DB (2002) Transposon Tn10. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM, Washington DC, pp 457–483

Hanson MS, Edelman R (2003) Progress and controversy surrounding vaccines against Lyme disease. Expert Rev Vaccines 2:683–703

Hertwig S (2007) Linear Plasmids and Prophages in Gram Negative Bacteria. Microbiol Monogr (in this volume)

Hinnebusch J, Barbour AG (1991) Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. J Bacteriol 173:7233–7239

Hinnebusch J, Barbour AG (1992) Linear- and circular-plasmid copy numbers in *Borrelia burgdorferi*. J Bacteriol 174:5251–5257

Ho TQ, Zhong Z, Aung S, Pogliano J (2002) Compatible bacterial plasmids are targeted to independent cellular locations in *Escherichia coli*. EMBO J 21:1864–1872

Hwang DS, Kornberg A (1992) Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. J Biol Chem 267:23083–23086

Huang WM, Joss L, Hsieh T, Casjens S (2004a) Protelomerase uses a topoisomerase IB/Y-recombinase type mechanism to generate DNA hairpin ends. J Mol Biol 337:77–92

Huang WM, Robertson M, Aron J, Casjens S (2004b) Telomere exchange between linear replicons of *Borrelia burgdorferi*. J Bacteriol 186:4134–4141

Jayaram M (2007) Split target specificity of ResT: a design for protein delivery, site selectivity and regulation of enzyme activity? Mol Micobiol 64(3):575–579

Kanaar R, Cozzarelli N (1992) Roles of supercoiled DNA structure in DNA transactions. Curr Opin Struct Biol 2:369–379

Kawabata H, Norris SJ, Watanabe H (2004) BBE02 disruption mutants of *Borrelia burgdorferi* B31 have a highly transformable, infectious phenotype. Infect Immun 72:7147–7154

Kennedy AK, Guhathakurta A, Kleckner N, Haniford DB (1998) Tn10 transposition via a DNA hairpin intermediate. Cell 95:125–134

Kennedy AK, Haniford DB, Mizuuchi K (2000) Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. Cell 101:295–305

Kobryn K, Chaconas G (2001) The circle is broken: telomere resolution in linear replicons. Curr Opin Microbiol 4:558–564

Kobryn K, Chaconas G (2002) ResT, a telomere resolvase encoded by the Lyme disease spirochete. Mol Cell 9:195–201

Kobryn K, Chaconas G (2005) Fusion of hairpin telomeres by the *B. burgdorferi* telomere resolvase ResT: implications for shaping a genome in flux. Mol Cell 17:783–791

Kobryn K, Naigamwalla DZ, Chaconas G (2000) Site-specific DNA binding and bending by the *Borrelia burgdorferi* Hbb protein. Mol Microbiol 37:145–155

Kobryn K, Burgin AB, Chaconas G (2005) Uncoupling the chemical steps of telomere resolution by ResT. J Biol Chem 280:26788–26795

Kornberg A, Baker TA (1992) DNA replication, 2nd edn. Freeman, New York

Krogh BO, Shuman S (2000) Catalytic mechanism of DNA topoisomerase IB. Mol Cell 5:1035–1041

Krogh BO, Shuman S (2002) Proton relay mechanism of general acid catalysis by DNA topoisomerase IB. J Biol Chem 277:5711–5714

Lawrenz MB, Kawabata H, Purser JE, Norris SJ (2002) Decreased electroporation efficiency in *Borrelia burgdorferi* containing linear plasmids lp25 and lp56: impact on transformation of infectious *B. burgdorferi*. Infect Immun 70:4798–4804

Li Y, Dabrazhynetskaya A, Youngren B, Austin S (2004) The role of Par proteins in the active segregation of the P1 plasmid. Mol Microbiol 53:93–102

LoGiudice K, Ostfeld RS, Schmidt KA, Keesing F (2003) The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. Proc Natl Acad Sci USA 100:567–571

Lovell S, Goryshin IY, Reznikoff WR, Rayment I (2002) Two-metal active site binding of a Tn5 transposase synaptic complex. Nat Struct Biol 9:278–281

Mizuuchi K, Baker TA (2002) Chemical mechanisms for mobilizing DNA. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) Mobile DNA II. ASM, Washington DC, pp 12–23

Morrison TB, Ma Y, Weis JH, Weis JJ (1999) Rapid and sensitive quantification of *Borrelia burgdorferi*-infected mouse tissues by continuous fluorescent monitoring of PCR. J Clin Microbiol 37:987–992

Naumann TA, Reznikoff WS (2002) Tn5 transposase active site mutants. J Biol Chem 277:17623–17629

Norris SJ (2006) Antigenic variation with a twist—the Borrelia story. Mol Microbiol 60:1319–1322

Pal U, Fikrig E (2003) Adaptation of *Borrelia burgdorferi* in the vector and vertebrate host. Microbes Infect 5:659–666

Perry K, Hwang Y, Bushman FD, Van Duyne GD (2006) Structural basis for specificity in the poxvirus topoisomerase. Mol Cell 23:343–354

Picardeau M, Lobry JR, Hinnebusch BJ (1999) Physical mapping of an origin of bidirectional replication at the centre of the *Borrelia burgdorferi* linear chromosome. Mol Microbiol 32:437–445

Picardeau M, Lobry JR, Hinnebusch BJ (2000) Analyzing DNA strand compositional asymmetry to identify candidate replication origins of *Borrelia burgdorferi* linear and circular plasmids. Genome Res 10:1594–1604

Pogliano J (2002) Dynamic cellular location of bacterial plasmids. Curr Opin Microbiol 5:586–590

Ravin NV, Kuprianov VV, Gilcrease EB, Casjens SR (2003) Bidirectional replication from an internal ori site of the linear N15 plasmid prophage. Nucleic Acids Res 31:6552–6560

Reznikoff WS (2003) Tn5 as a model for understanding DNA transposition. Mol Microbiol 47:1199–1206

Rezsohazy R, Hallet B, Delcour J, Mahillon J (1993) The IS4 family of insertion sequences: evidence for a conserved transposase motif. Mol Microbiol 9:1283–1295

Rice PA, Baker TA (2001) Comparative architecture of transposase and integrase complexes. Nat Struct Biol 8:302–307

Roth A, Urmoneit B, Messer W (1994) Functions of histone-like proteins in the initiation of DNA replication at oriC of *Escherichia coli*. Biochimie 76:917–923

Rybchin VN, Svarchevsky AN (1999) The plasmid prophage N15: a linear DNA with covalently closed ends. Mol Microbiol 33:895–903

Schwan TG, Burgdorfer W, Rosa P (1999) *Borrelia*. In: Murray PR, Baron EJ, Pfaller MA, Tenover FC, Yolken RH (eds) Manual of clinical microbiology. ASM, Washington DC, pp 746–758

Shuman S (1998) Vaccinia virus DNA topoisomerase: a model eukaryotic type IB enzyme. Biochim Biophys Acta 1400:321–337

Singh SK, Girschick HJ (2004) Molecular survival strategies of the Lyme disease spirochete *Borrelia burgdorferi*. Lancet Infect Dis 4:575–583

Steere AC (2001) Lyme disease. N Engl J Med 345:115–125

Steere AC, Levin RE, Molloy PJ, Kalish RA, Abraham JH, Liu NY, Schmid CH (1994) Treatment of Lyme arthritis. Arthritis Rheum 37:878–888

Steere AC, Coburn J, Glickstein L (2004) The emergence of Lyme disease. J Clin Invest 113:1093–1101

Stewart PE, Thalken R, Bono JL, Rosa P (2001) Isolation of a circular plasmid region sufficient for autonomous replication and transformation of infectious *Borrelia burgdorferi*. Mol Microbiol 39:714–721

Stewart PE, Chaconas G, Rosa P (2003) Conservation of plasmid maintenance functions between linear and circular plasmids in *Borrelia burgdorferi*. J Bacteriol 185:3202–3209

Stewart PE, Byram R, Grimm D, Tilly K, Rosa P (2005) The plasmids of *Borrelia burgdorferi*: essential genetic elements of a pathogen. Plasmid 53:1–13

Tian L, Claeboe CD, Hecht SM, Shuman S (2004) Remote phosphate contacts trigger assembly of the active site of DNA topoisomerase IB. Structure (Camb) 12:31–40

Tilly K, Fuhrman J, Campbell J, Samuels DS (1996) Isolation of *Borrelia burgdorferi* genes encoding homologues of DNA-binding protein HU and ribosomal protein S20. Microbiology 142:2471–2479

Tourand Y, Kobryn K, Chaconas G (2003) Sequence-specific recognition but position-dependent cleavage of two distinct telomeres by the *Borrelia burgdorferi* telomere resolvase, ResT. Mol Microbiol 48:901–911

Tourand Y, Bankhead T, Wilson SL, Putteet-Driver AD, Barbour AG, Byram R, Rosa P, Chaconas G (2006) Differential telomere processing by Borrelia telomere resolvases in vitro but not in vivo. J Bacteriol 188:7378–7386

Tourand Y, Lee L, Chaconas G (2007) Telomere resolution by *Borrelia burgdorferi* ResT through the collaborative efforts of tethered DNA binding domains. Mol Microbiol 64(3):580–590

Traktman P (1996) Poxvirus DNA replication. In: De Pamphlis ML (ed) DNA replication in eukaryotic cells. Cold Spring Harbor Laboratory Press, Plainview, NY, pp 775–798

Van Duyne GD (2001) A structural view of cre-loxp site-specific recombination. Annu Rev Biophys Biomol Struct 30:87–104

Van Duyne GD (2002) A structural view of tyrosine recombinase site-specific recombination. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitiz AM (eds) ASM, Washington DC, pp 93–117

Wang JC (1996) DNA topoisomerases. Annu Rev Biochem 65:635–692

Zhang JR, Norris SJ (1998) Genetic variation of the *Borrelia burgdorferi* gene vlsE involves cassette-specific, segmental gene conversion. Infect Immun 66:3698–3704

Zhang JR, Hardham JM, Barbour AG, Norris SJ (1997) Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. Cell 89:275–285

Zuckert WR, Meyer J (1996) Circular and linear plasmids of Lyme disease spirochetes have extensive homology: characterization of a repeated DNA element. J Bacteriol 178:2287–2298

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_094/Published online: 20 June 2007

Linear Plasmids and Prophages in Gram-Negative Bacteria

Stefan Hertwig

Bundesinstitut für Risikobewertung,
Abteilung 4 – Biologische Sicherheit, 12277 Berlin, Germany
stefan.hertwig@bfr.bund.de

1 Incidence of Linear Plasmids in Gram-Negative Bacteria 142

2 Linear Plasmid Prophages with Covalently Closed Ends 142
2.1 Overall Organization of the Genomes 144

3 Protelomerases are Functional Analogues of Lambdoid Phage Integrases and Belong to the Family of Site-Specific Tyrosine Recombinases 145
3.1 Simple In Vitro and In Vivo Assays to Follow Protelomerase Activity . . . 147
3.2 The Telomere Resolution Site (*telRL*) Maps Adjacent to the Protelomerase Gene 147
3.2.1 The Most Efficient Binding Substrate for TelN is *tos* 150
3.2.2 Two TelN Monomers Bind to a Segment of ∼ 50 bp at *telRL* 152
3.2.3 *TelO* is Essential for the Processing of the Telomere Resolution Site 154
3.2.4 The *telRL* Site is Cleaved by Staggered Cuts Six Base Pairs Apart Around the Centre of Dyad Symmetry 155
3.2.5 TelN Binds Covalently to the 3′-Phosphoryl of the Cleaved Strands 156
3.3 Protelomerases are also Involved in Linear Plasmid Replication and Maintenance 157

4 Perspectives 160

References 160

Abstract Among the known linear plasmids in Gram-negative bacteria, there are three well-characterized phages (N15, PY54 and ϕKO2) whose prophages replicate as linear plasmids with covalently closed hairpin ends (telomeres). Close to the left telomere, a gene (*tel*) resides encoding the protelomerase. This tyrosine recombinase-like enzyme has cleaving–joining activity and is responsible for the generation of the telomeres by processing a palindromic telomere resolution site (*telRL*) located adjacent to the *tel* gene. The protelomerase exerts its activity as cleaving–joining enzyme in a concerted action. Functional analyses with the N15 protelomerase TelN demonstrated that two TelN molecules recognize *telRL* and bind to the target site. Within the centre of *telRL* (*telO*), the double-stranded DNA is cut by staggered cleavages six base pairs apart centred around the axis of dyad symmetry of the target site. TelN is transiently attached to the 3′ ends of the cleaving sites by covalent binding. Then the new phosphodiester bond with the complementary strand is generated in a transesterification step. Protelomerases are multifunctional enzymes. Besides telomere formation, TelN was shown to be implicated in linear plasmid replication and in the regulation of plasmid partitioning.

1
Incidence of Linear Plasmids in Gram-Negative Bacteria

During the last decade, an increasing number of publications have appeared about the presence of linear plasmids in Gram-negative bacteria. The first report dates back to 1967 when Ravin described N15, a linear plasmid prophage of *Escherichia coli* (Ravin and Golub 1967). Linear plasmids were also found in the spirochete genus *Borrelia* (Plasterk et al. 1985). Since then a lot of work has been done on the characterization of *Borrelia* plasmids (Kobryn 2007, in this volume). Linear plasmids have also been identified in other Gram-negative bacteria, namely *Agrobacterium tumefaciens* (Jumas-Bilak et al. 1998), *Ochrobactrum* spp., *Pseudomonas putida* (Danko et al. 2004) and in several *Xanthobacter* strains (Krum and Ensign 2001; Tardif et al. 1991). The plasmids of these organisms are apparently associated with the biodegradation of organic compounds, e.g. vinyl chloride, 1,2-dichloroethane and propylene (Fetzner et al. 2007, in this volume). Using pulsed-field gel electrophoresis (PFGE) techniques like contour-clamped homogeneous-field electrophoresis (CHEFE), the linearity of the plasmids has been documented and sizes between 90 and 320 kb have been determined. Unfortunately, only partial sequence data are available and details about the structure and the termini of these plasmids are not yet known.

Much more is known about the three linear plasmid prophages N15, PY54 and ϕKO2 that were isolated from enterobacteria. The complete DNA sequences have been determined and functional analyses have been performed for various genes. Hence the following sections will focus on the characteristics of these unusual forms of phage plasmids.

2
Linear Plasmid Prophages with Covalently Closed Ends

The temperate *E. coli* phage N15 has been known since 1964 when Victor Ravin isolated the phage from sewage in Moscow, Russia. It is the only linear plasmid found in *E. coli*. Phage PY54 was recovered after mitomycin C treatment of a non-pathogenic *Yersinia enterocolitica* strain isolated from food in Germany (Popp et al. 2000). The ϕKO2 plasmid was detected in a nickel-resistant *Klebsiella oxytoca* strain isolated from the mineral oil–water coolant–lubricant emulsion of a metal working machine in Sweden (Stoppel et al. 1995). Despite their completely different origins, the three temperate phages share some striking similarities. They exhibit a lambda-like morphology with a hexagonal head and a long, non-contractile tail. N15 and PY54 have a rather narrow host range exclusively infecting *E. coli* and some *Y. enterocolitica* strains, respectively, while ϕKO2 could not be propagated on indicator bacteria as yet. It is not clear whether this phage is defective or

if it has an unusual host specificity. The genomes of the three phages are composed of linear double-stranded DNA, which is 46.3 kb (N15, PY54) and 51.6 kb (ϕKO2) in size. Contrary to N15, which has 12-bp overhangs at the 5′ ends of the genome, very similar to lambda, the PY54 and ϕKO2 cohesive ends are 10 bp long and protruding at the 3′ ends.

At the lysogenic stage, the prophages replicate as linear low copy-number plasmids with covalently closed hairpin termini. This was inferred from two

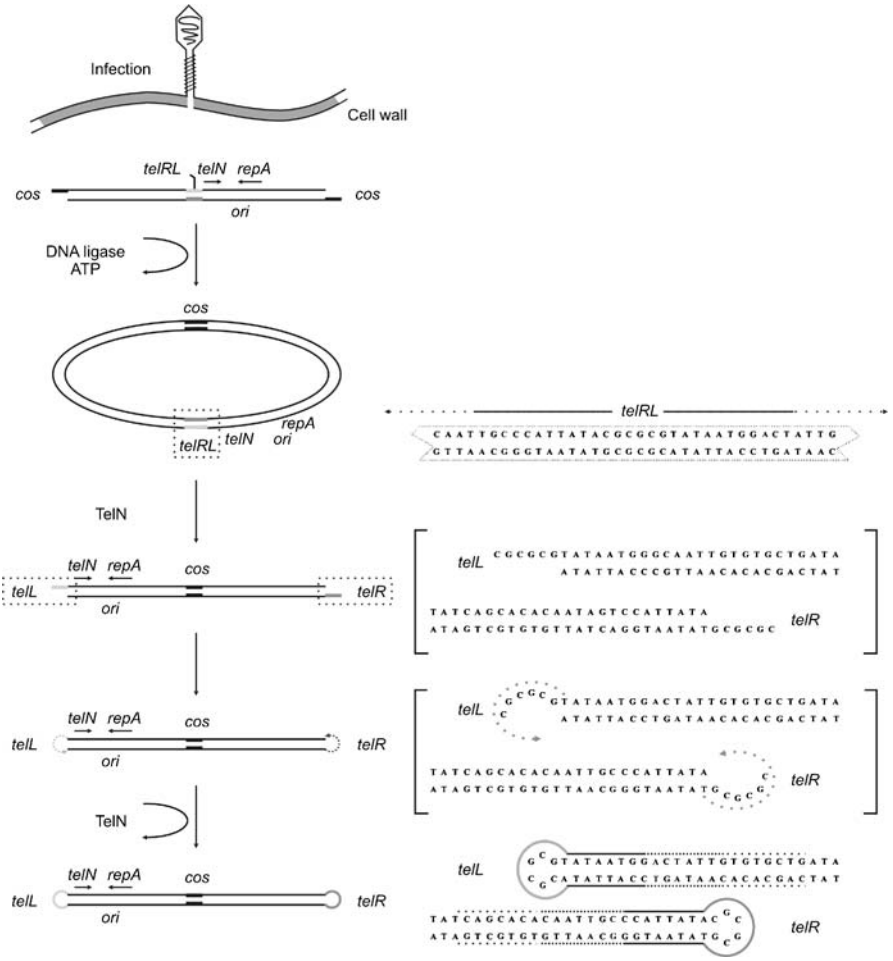


Fig. 1 Conversion of the linear N15 phage genome into the linear plasmid prophage. Upon infection the phage genome circularizes by its cohesive ends (*cos*). The circular molecule is then subjected to telomere resolution. The telomere resolution site (*telRL*) located opposite to *cos* is processed by the protelomerase TelN, resulting in a linear plasmid with covalently closed ends (see text for details). For better understanding, sequences shown in *brackets* depict probable intermediates of the reaction

observations: (1) terminal, but not internal, restriction fragments renature rapidly after heat denaturation and quick cooling; and (2) treatment of plasmid DNA with the S1 nuclease, which cuts single-stranded DNA, abolishes this effect (Svarchevsky and Rybchin 1984). The size and sequence of each plasmid are the same as those of the respective phage genome but both molecules are approximately 50% circularly permuted, which means that the end sequences of the genomes are located in the middle of the plasmids and vice versa. Figure 1 illustrates the conversion of the N15 linear genome into the linear plasmid. Upon infection of a host cell, the phage genome circularizes by its cohesive ends. Following this step, a phage-encoded enzyme termed protelomerase (prokaryotic telomerase, see Sect. 3) cleaves by staggered cuts a palindromic target sequence (*telRL*) located opposite to *cos*. The overhanging DNA strands are self-complementary and therefore able to fold back. Two new phosphodiester bonds are formed, resulting in the linear plasmid with covalently closed ends.

The exact mechanism of lytic replication at which the linear plasmid is converted into the linear genome has not yet been elucidated. However, it is assumed that similar to λ, lytic growth starts with circular molecules that are used for rolling circle replication leading to concatemeres. Single genomes are then generated by site-specific cleavage at *cos* and packed into capsids.

2.1 Overall Organization of the Genomes

The nucleotide sequences of N15, PY54 and ϕKO2 have been determined (Casjens et al. 2004; Hertwig et al. 2003b; Ravin et al. 2000). Sequence analyses disclosed 59 (N15), 67 (PY54) and 64 (ϕKO2) putative open reading frames (ORFs). Similarities to known proteins and functional analyses allowed assignments for a number of gene products. The genes of each phage are situated in two arms separated by the *tel* site and are arranged in a similar order (Fig. 2). Genes coding for head and tail assembly proteins are located in the left arm. In addition, the left arm of the three phages carries genes for partitioning proteins. Close to the *tel* site, the right arm harbours the protelomerase gene *tel* essential for the generation of the plasmid hairpin ends. Rightward a large gene resides encoding a multifunctional replication protein (RepA). Adjacent to *repA* the genomes contain the primary immunity region comprising genes for the prophage repressor CB, the Cro repressor and the Q transcription antiterminator. Further to the right genes probably associated with DNA methylation and host cell lysis have been identified. However, for a significant number of genes of each phage a functional prediction could not be made, even though homologous counterparts at similar positions have been detected in the other phage genomes.

Overall the left arm of the PY54 genome is more closely related to ϕKO2 than to N15, whereas N15 and ϕKO2 show the strongest homologies in the

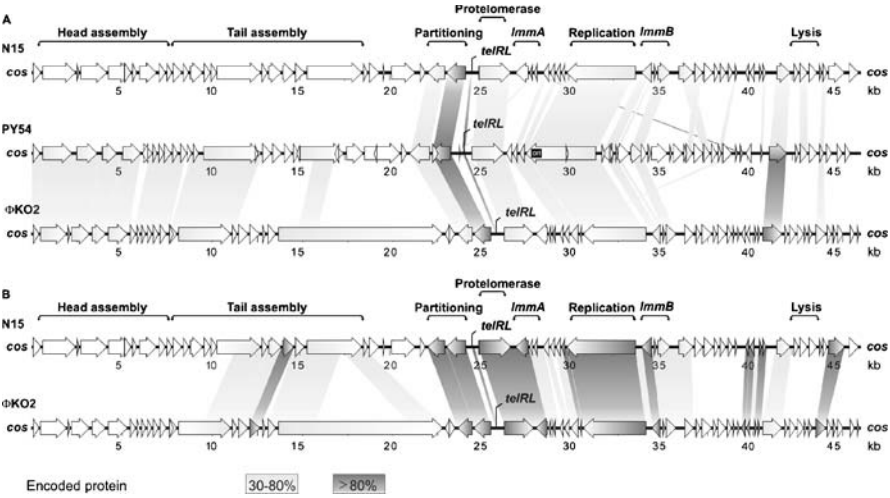


Fig. 2 Organization of the phage N15, PY54 and ϕKO2 genomes. Putative genes are indicated by *arrows*. Functional assignments are shown above the genes. The positions of the telomere resolution sites (*telRL*) and the origin of PY54 are indicated. *Grey bars* connect homologous regions of the genomes. **a** Homologies between PY54 and the other two phages. **b** Relationship between N15 and ϕKO2

right arm. As the genomes of all three phages are mosaically related to λ and as some of their genes are highly homologous to each other or to those of lambda-like phages, it has been suggested to include N15, PY54 and ϕKO2 as a subgroup into the lambdoid phage family (Casjens et al. 2004). This family comprises a number of similar phages (e.g. λ, P22) that are genetic mosaics sharing a common order and organization of genetic functions along their genomes (Juhala et al. 2000). Some members of this family can easily recombine with each other to produce functional hybrids. Horizontal gene exchange is therefore suggested to be a major component of evolution for these viruses (Hendrix et al. 2003).

3 Protelomerases are Functional Analogues of Lambdoid Phage Integrases and Belong to the Family of Site-Specific Tyrosine Recombinases

At the lysogenic stage, temperate bacteriophages are generally integrated into the bacterial chromosome. The integration is promoted by phage-encoded enzymes called integrases, which mediate unidirectional site-specific recombination between the DNA recognition sequences of the phage and its host. Integrases may be grouped into two major families, the tyrosine recombinases and the serine recombinases, based on their mode of catalysis (Groth

and Calos 2004). Members of the tyrosine family are, e.g., the integrases of the lambdoid phages λ, P22 and HK022. Contrary to most temperate phages, the prophages of N15, PY54 and ϕKO2 are not integrated into the bacterial chromosome but replicate as linear plasmids with covalently closed ends

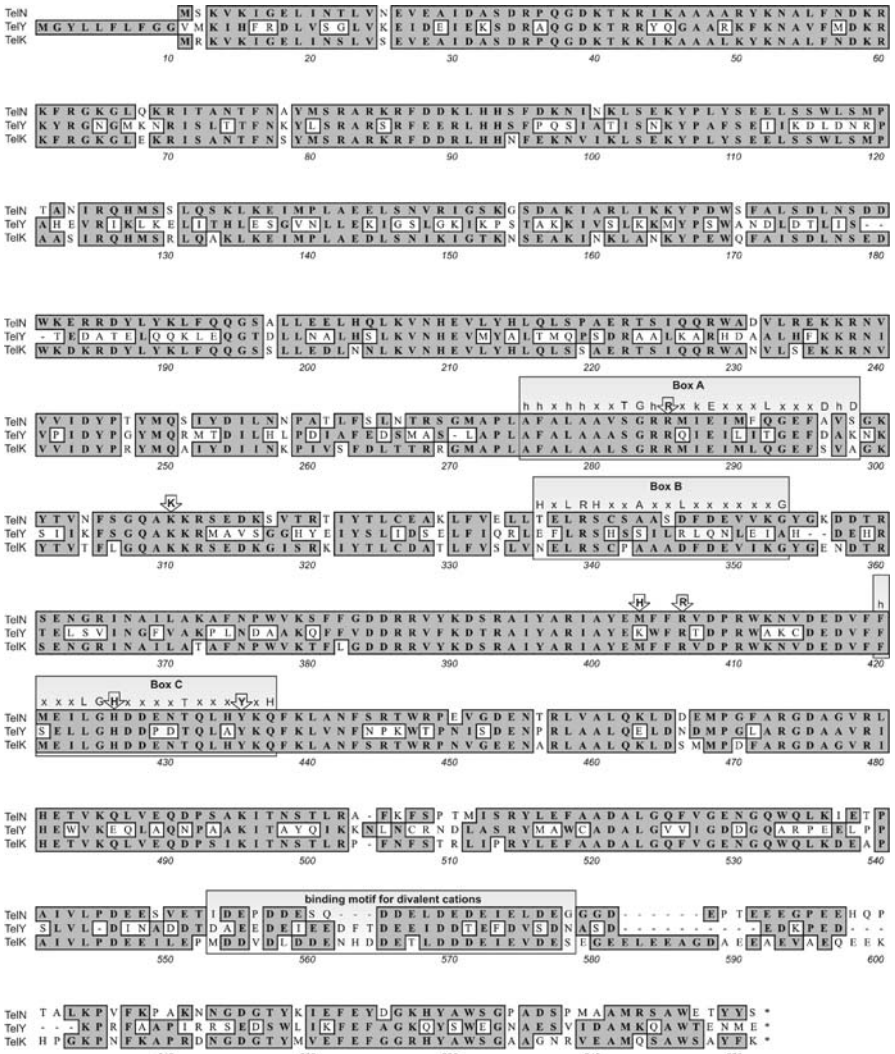


Fig. 3 Alignment of the TelN (N15), TelY (PY54) and TelK (ϕKO2) protelomerases. The positions of the boxes A, B and C are indicated. *Arrowheads* show the pentad "RKhRH" plus the catalytically active tyrosine residue, which are conserved in many tyrosine recombinases. The middle histidine in RKhRH is less common than the other residues of this motif, and is a lysine in the PY54 protelomerase and a methionine in the N15 and ϕKO2 protelomerases. A possible binding motif for divalent cations is also shown

(telomeres). However, comparable to phage integrases of other temperate phages, N15, PY54 and ϕKO2 possess an enzyme (protelomerase) essential for establishing the lysogenic life cycle. This occurs by conversion of the linear phage genome into the linear plasmid in which the protelomerase is responsible for the generation of the plasmid hairpin ends.

The phage-encoded protelomerases of more than 70 kDa are much larger than most phage integrases, which have sizes between 30 and 40 kDa. Alignments revealed 77% sequence identity between the N15 (TelN) and ϕKO2 (TelK) proteins, while the PY54 protelomerase TelY is approximately 41% identical to TelN and TelK (Fig. 3). The overall homologies to phage integrases as the λ integrase are low. Nevertheless, three major clusters, designated boxes A, B and C, and a pentad ("RKhRH") conserved in many integrases are also present in the protelomerases (Esposito and Scocca 1997; Groth and Calos 2004). Furthermore, the relationship between protelomerases and integrases is corroborated by the facts that in both enzymes a tyrosine residue is pivotal for the catalytic activity, and that during catalysis the protelomerases are transiently linked to a 3′-phosphoryl group of the target DNA (see Sect. 3.2.5).

3.1 Simple In Vitro and In Vivo Assays to Follow Protelomerase Activity

Deneke et al. (2000) designed an in vitro assay for protelomerases. DNA fragments containing possible substrates for the protelomerase are inserted into a cloning vector. The recombinant plasmids are linearized by a restriction endonuclease and then incubated with purified protelomerase. For TelN, an incubation at 30 °C for 30 min was used. Following this step, products are analysed on agarose gels under non-denaturing and alkaline conditions. Under alkaline (denaturing) conditions, DNA fragments containing a terminal hairpin exhibit twice the length as those observed under neutral conditions.

Hertwig et al. (2003a) developed an assay to investigate protelomerase activity in vivo. Plasmids containing possible substrates for the protelomerase are introduced into an indicator strain. Bacteria are then infected with a high-titre phage lysate. For PY54, a mutant was used that was unable to lyse the bacteria. After different times of infection (e.g. 10 min, 1 h, 6 h), plasmids are isolated and analysed on an agarose gel. Processing of the substrate is visible by linearization and subsequent loss of the plasmid, which can also be determined by plating the bacteria on agar containing the respective antibiotics.

3.2 The Telomere Resolution Site (*telRL*) Maps Adjacent to the Protelomerase Gene

The protelomerase gene of the phages N15, PY54 and ϕKO2 is located at a similar position in the middle of the phage genome (Fig. 2). Approximately

160 bp upstream, the genomes harbour a large inverted repeat (palindrome) of 56 bp (N15), 42 bp (PY54) and 50 bp (ϕKO2), termed the telomere resolution site (*telRL*, Fig. 4). *TelRL* can be divided into the left-hand arm *telL* and the right-hand arm *telR*. Using restriction and DNA sequence analyses, it was demonstrated that the nucleotide sequences of the arms correspond to the sequences of the plasmid hairpin ends (Casjens et al. 2004; Hertwig et al. 2003a; Malinin et al. 1992). Hence the phage genomes are approximately 50% circularly permuted to their respective plasmid prophages. By comparison it becomes evident that the *telRL* sites of the phages share a high degree of sequence homology. Ten nucleotides in the centre are even identical in all three repeats. The centres contain a 14-bp (N15 and ϕKO2) and 10-bp (PY54) alternating Pyr/Pur sequence that is likely to form a region of Z-DNA structure (Rybchin and Svarchevsky 1999).

Contrary to the perfect PY54 palindrome, those of N15 and ϕKO2 exhibit some mismatches. The N15 *telRL* site is composed of the central 22-bp *telO* palindrome and a 14-bp inverted repeat (L3/R3) separated from *telO* by three nucleotides (non-inverted sequence) on either side (Fig. 4). Following this nomenclature, the *telO* palindrome of the ϕKO2 *telRL* site is 28 bp in length and separated from the flanking 7-bp inverted repeat by 4-bp non-inverted repeat sequences. The *telRL* sites of N15 and PY54 are flanked by two and one further inverted repeats, respectively (Fig. 4). In N15, the complex comprising the 56-bp palindrome and its adjacent inverted repeats (L2/R2 and L1/R1) has been termed *tos* (telomerase occupancy site).

To study the catalytic properties of the protelomerases in detail, the phage-encoded *tel* genes were cloned and expressed in *E. coli*. The overproduced protelomerases displayed sizes of 75 to 78 kDa. The purified protelomerases were used for in vitro assays with plasmids containing the respective *telRL*

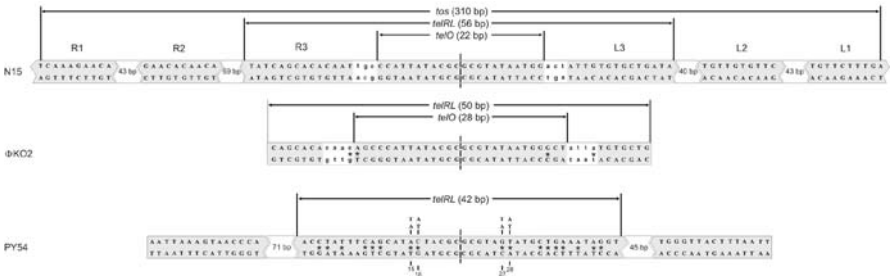


Fig. 4 The telomere resolution site regions of N15, ϕKO2 and PY54. *Horizontal arrows* indicate the sequences belonging to *tos*, *telRL* and *telO*. Nucleotides within the N15 and ϕKO2 *telRL* repeats diverging from perfect twofold symmetry are depicted by *lower-case letters*. Nucleotides in the ϕKO2 and PY54 *telRL* sites that differ from the N15 *telRL* sequence are marked by *asterisks*. Replacement of each of the nucleotides 15 and 16 (AC) and 27 and 28 (GT) with TA yields a PY54 *telRL* site that is a suitable substrate for TelN. The centres of dyad symmetry are shown by *broken vertical lines*

site. After incubation with the enzyme preparation, linearization was observed with each of the protelomerases in the presence of the corresponding *telRL* site (Deneke et al. 2000; Hertwig et al. 2003a; Huang et al. 2004). By contrast, the N15 *telO* repeat and a partial PY54 *telRL* site were not processed by the protelomerases. Processing of *telRL* occurred with both circular and linear substrates, indicating that negative supercoiling is not a prerequisite for cleavage. Kobryn and Chaconas (2002) performed similar experiments with the telomere resolvase ResT that generates the hairpin telomeres in *Borrelia* (Kobryn 2007, in this volume). Contrary to telomere resolution in prophages, negatively supercoiled target DNA was not processed by ResT. Instead, a 20-fold and greater stimulation of the reaction was observed by positive supercoiling (Bankhead et al. 2006).

Figure 5 shows the product analysis of TelN enzymatic activity. Preincubation of recombinant plasmids containing *telRL* with restriction endonucleases cutting the substrate once gave, as expected, two fragments. Under alkaline conditions, the fragments (single-stranded DNA) had twice the length of those observed under neutral conditions. DNA fragments generated by the protelomerase were isolated and sequenced. Each double-stranded frag-

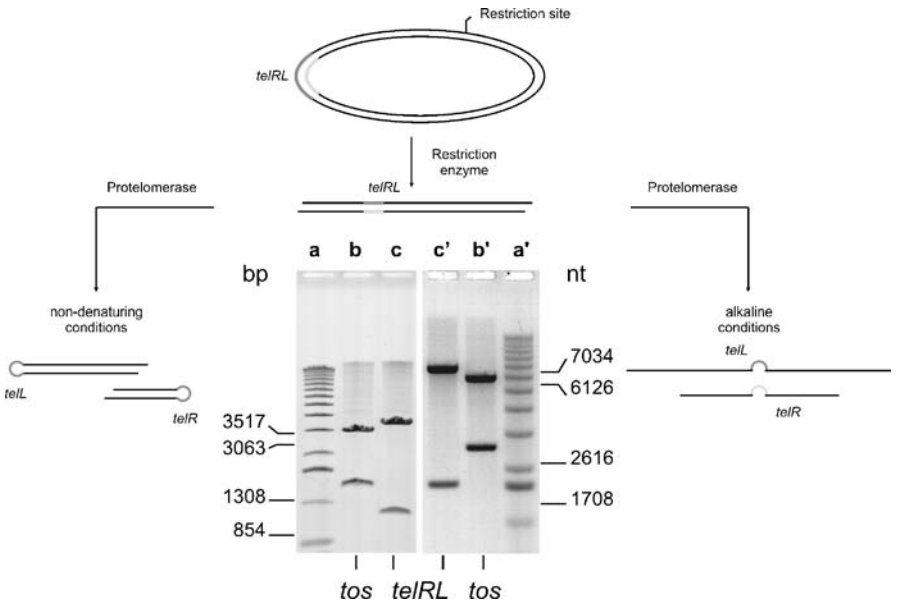


Fig. 5 The recognition sequence for TelN lies within *telRL*. Plasmids harbouring either *tos* or *telRL* were linearized with a restriction endonuclease and then incubated with TelN. Products of the reaction were analysed on 0.7 and 1% agarose gel under non-denaturing (*left panel*) and alkaline conditions (*right panel*), respectively. Lanes a and a′, size marker (1-kb DNA ladder); lanes b and b′, analysis of the *tos* product; lanes c and c′, analysis of the *telRL* product (Deneke et al. 2000)

ment could be sequenced as a continuous single strand. The nucleotide sequence contained either *telR* or *telL* in which the sequences of the upper and lower strands were joined together at the centre of *telRL* dyad symmetry. These results show that *telRL* is processed by the protelomerase without any other phage- or host-encoded factor or co-factor. The generated fragments contain one additional covalent linkage per fragment, resembling a hairpin-like structure (Fig. 1). Thus, the protelomerase has cleaving–joining activity.

Deneke et al. (2000) optimized the assay conditions for TelN and found that this enzyme is salt-sensitive and that the activity of TelN is optimal at approximately pH 7.5 and 25 °C. Divalent cations (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Mn^{2+}) stimulated the reaction significantly. A binding motif (amino acid positions 541 to 563) for divalent cations has already been predicted for the TelN sequence (Fig. 3). However, low concentrations (10–20 mM) of EDTA also had a stimulatory effect on the reaction and only at higher concentrations of EDTA was the TelN-mediated reaction inhibited (Deneke 2002). Therefore, a requirement for divalent cations is questionable.

3.2.1 The Most Efficient Binding Substrate for TelN is *tos*

Telomere resolution can be dissected into several steps. Prior to processing, the protelomerases have to recognize and bind their target DNAs. In order to study the recognition and binding characteristics of TelN, the binding and processing steps were uncoupled by use of a TelN mutant protein (TelNY424F), in which the proposed active tyrosyl residue Tyr-424 was replaced by a phenylalanine (Deneke et al. 2002). Preceding experiments had shown that TelNY424F lost its cleaving–joining activity but retained its specific DNA binding ability (Deneke et al. 2000). The modified TelN protein was incubated with either *tos*, *telRL* or *telO* followed by electron microscopic analysis of the complexes. Site-specific complex formation was observed with *tos* and *telRL* only while *telO* failed to yield specific complexes. Symmetric stepwise reduction of the 56-bp *telRL* sequence resulted in a gradual loss of DNA binding and at a substrate size of 36 bp, specific binding was no longer observed. Furthermore, the removal of one arm (R3) of the inverted repeat flanking *telO* abolished the sequence-specific binding (Fig. 4). Thus, the repetitive sequences L3 and R3 surrounding the *telO* palindrome in *telRL* are essential for TelN binding. This was confirmed by fragment retention experiments. On polyacrylamide gels, target fragments containing *tos* or *telRL* were bound by TelN resulting in decreased electrophoretic mobility. By contrast, no retention was observed with *telO*. The apparent equilibrium dissociation constant value ($k_{d(app)}$) for *telO* was the same as that for vector DNA. Besides the specific binding capacity, TelN also binds DNA in a non-sequence-specific way.

Interestingly, there was an almost fivefold increase in stability of the TelN–*tos* complexes compared with *telRL* complexes. This finding indicates that the additional repeats L1/R1 and L2/R2 present in *tos* are not vital for binding and for processing of the substrate in vitro but enhance the stability of the complexes (Fig. 4). The authors of this study speculate that the aforementioned repeats might serve as association points where TelN preferably binds and slides along the DNA until it hits *telRL*, where stable complex formation takes place. The sets of repeats within *tos* may also favour a structural alteration at *telRL* that improves DNA–TelN interactions to yield more stable complexes. Hertwig et al. (2003a) studied the processing of the PY54 palindrome in *Yersinia*. They ascertained that a plasmid harbouring the 42-bp palindrome was not linearized upon infection with PY54. However, a construct comprising the palindrome and flanking sequences including the 15-bp inverted repeat was clearly processed (Figs. 4 and 6). This observation suggests that the additional 15-bp repeat is essential for processing under in vivo conditions, and raises the question whether the repeats L1/R1 and L2/R2 present in *tos* are also required for telomere resolution in vivo.

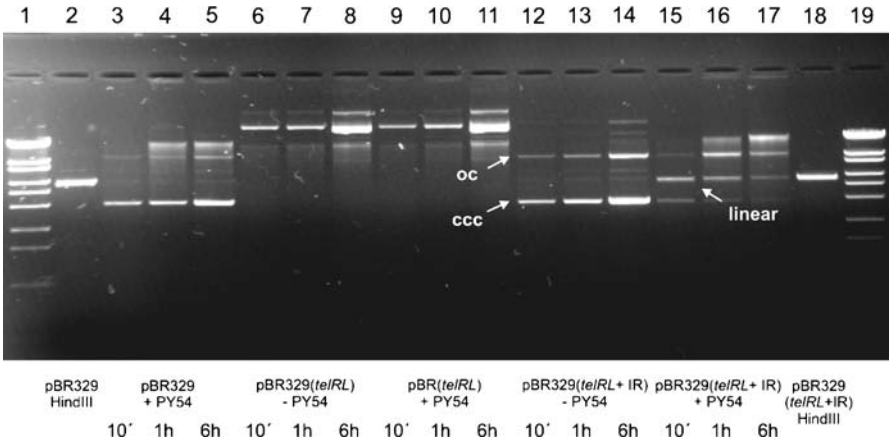


Fig. 6 In vivo linearization of the PY54 *telRL* site occurs only in the presence of its flanking inverted repeat (IR). *Y. enterocolitica* strains harbouring only pBR329 or pBR329 with either *telRL* or *telRL* and IR were infected by PY54. Plasmids were analysed on a 0.8% agarose gel after 10 min, 1 h or 6 h of infection. Lanes 1 and 19, size marker (λ Eco130I); lane 2, pBR329/HindIII; lanes 3–5, pBR329 isolated from an infected *Yersinia* strain; lanes 6–8, pBR329 with *telRL* isolated from a non-infected *Yersinia* strain; lanes 9–11, same as before, but from an infected *Yersinia* strain; lanes 12–14, pBR329 with *telRL* and IR isolated from a non-infected *Yersinia* strain; lanes 15–17, same as before, but from an infected *Yersinia* strain; lane 18, same construct as before, digested with HindIII. ccc, covalently closed circle; oc, open circle (Hertwig et al. 2003)

3.2.2
Two TelN Monomers Bind to a Segment of ~ 50 bp at *telRL*

Previous studies demonstrated that the N15 protelomerase TelN relies on the *telRL* repeats for efficient binding. To define the binding site on *telRL*, DNase I footprint analyses were carried out using the TelNY424F protein (Deneke et al. 2002). The footprints exhibited a large stretch of approximately 50 nucleotides on both DNA strands protected against nuclease attack. The stretches comprise *telO* and adjacent sequences including one arm of the L3/R3 repeat (Fig. 7a). On each strand, one hypersensitive band could be detected. The hypersensitive bands reveal a mirror symmetric arrangement and separate two protected regions of 52 and 6 bp which together cover *telRL* (Fig. 7a).

Surface plasmon resonance showed that TelN is monomeric in solution. In the presence of *telRL*, 1.64 TelN molecules per *telRL* target were determined, which allows the conclusion that two TelN molecules are bound to each *telRL* site. Huang and co-workers investigated the subunit structure of TelN and TelK by equilibrium ultracentrifugation (Huang et al. 2004). Both proteins are monomers in solution. The number of molecules needed to interact with the substrate was determined by the use of *telRL* duplex oligonucleotides as targets. To distinguish the left half of *telRL* from the right, the oligonucleotides had an extension of ten nucleotides at the L3 side (Fig. 8). When a nick was placed on the top strand at the centre of dyad symmetry between nucleotides 28 and 29, TelN could not process this suicide substrate and caused an accumulation of two DNA–protein complexes that could be separated on polyacrylamide gels. Upon ^{32}P labelling of the top left end of the suicide substrate and treatment with TelN, the autoradiogram showed the faster migrating protein–oligonucleotide complex, which contained the 28-bp (R3) side of *telRL*. Labelling of the bottom right end of the suicide substrate gave a slower migrating complex comprising TelN and the extended 38-bp L3 side of *telRL*. Each half of the dyad symmetrical target site thus provides a binding (and processing) site for each molecule of the protelomerase.

Which nucleotides of *telRL* in addition to the L3/R3 repeat are essential for TelN binding? To answer this question Deneke et al. (2002) introduced mutations within *telO* and examined the stability of the *telRL*/TelN complexes. A target mutation leading to an asymmetric sequence situation in *telO* resulted in less stable complex formation (Fig. 8). Restoring symmetry by introduction of a second mutation further destabilized the complexes. This already indicates that sequences within *telO* are also important for TelN binding. Substrates containing mutations in which part of the core sequence was replaced by a random symmetric sequence, or in which the central $(CG)_3$ was extended to a symmetric $(CG)_7$ by replacement of the flanking TATA region, behaved inertly in binding (Fig. 8). On the other hand, the replacement of the central $(CG)_3$ by TATATA yielding $(TA)_7$ retained full binding capacity.

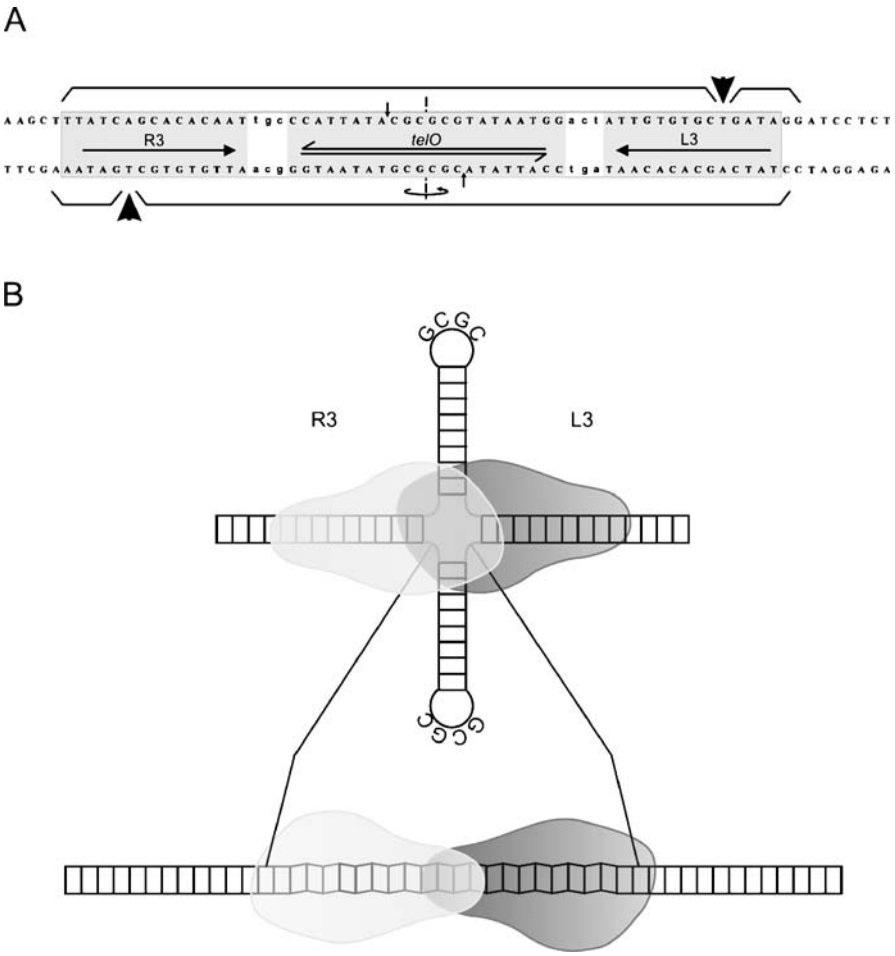


Fig. 7 Schematic representation of TelN-protected nucleotides of *telRL* determined by footprint analysis. **a** The nucleotide sequence of *telRL* and the flanking regions is shown. The inverted repeat R3/L3 of *telRL* is marked by *horizontal arrows* and the inner palindrome *telO* is indicated by a *double arrow*. *Brackets* denote nucleotides protected from DNase I attack. *Bold vertical arrowheads* above and below the sequence indicate nucleotide positions whose 5′-phosphodiester bonds are hypersensitive to DNase I cleavage. The symmetry centre and the cleavage sites are depicted as a *broken vertical line* and *small vertical arrows*, respectively. **b** Proposed models for TelN–*telRL* interaction. In the hairpin model (*upper panel*), the palindrome *telO* is extruded as hairpins. The lower panel shows the B–Z-form model, in which DNA in the Z conformation is shown as a *zigzag* (Deneke et al. 2000)

Consequently, it can be presumed that the sequences flanking the central 6-bp segment of *telO* but not the segment itself are important for complex formation.

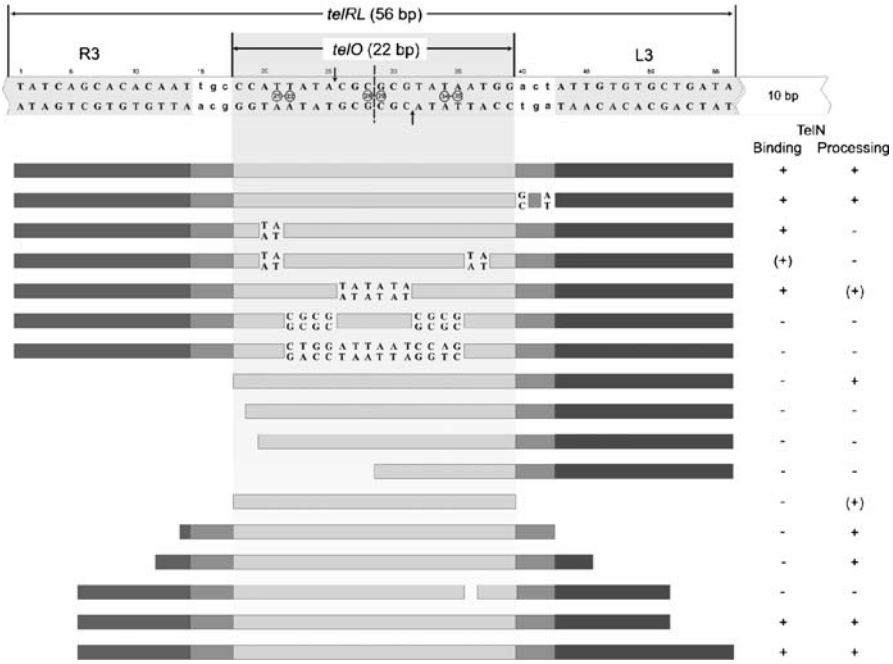


Fig. 8 Mutational analysis of the TelN (N15) target *telRL*. Substrates containing mutations were inserted into a vector and subjected to TelN processing. The nucleotide sequences of *telRL* and *telO*, the centre of dyad symmetry, and the cleavage sites for the protelomerase are shown. Nucleotide positions deviating from perfect twofold symmetry are given in *lower-case letters*. *Numbers* correspond to the length of *telRL* (56 bp). The extension of ten nucleotides at the L3 end, which were used to distinguish the products, is indicated. Mutants are represented by *bars*; base pairs exchanged are shown *within the bars*. Phosphodiester linkages critical for protein–substrate interactions are depicted by *connected circles*. (+), reduced amount of bound or processed substrate

3.2.3 *TelO* is Essential for the Processing of the Telomere Resolution Site

At very high TelN concentrations (~ 50-fold higher than for *telRL*), *telO* is weakly processed suggesting that *telO* contains all the sequence requirements needed for telomere resolution in vitro (Deneke et al. 2002). For that reason *telRL* substrates harbouring mutations within *telO* that had been used for the binding studies were also investigated in terms of recognition and processing by the N15 protelomerase. It turned out that the AT→TA mutation within *telO* resulted in a complete loss of processing (Fig. 8). Similarly, the substrate containing a second AT→TA mutation, which restored the sequence symmetry of *telO*, could not be processed. The data make clear that the cen-

tral palindrome of *telRL* is of great importance in telomere resolution, and indicate that apparently both the sequence itself and the symmetry of the sequence are crucial for processing. This was confirmed with other substrates in which the sequence of *telO* was changed (Fig. 8). By stepwise reduction of one end of *telO* it could be demonstrated that the core of this sequence lies between the base pairs 19 and 38 (Fig. 8).

The importance of the integrity of the central *telO* sequences for protelomerase activity was also reported by Huang et al. (2004) who used duplex oligonucleotides described in the previous section. The introduction of a single nick up to position 20 and beyond position 36 on the top strand of *telRL* did not affect protelomerase activity. By contrast, nicks introduced at locations between nucleotides 22 and 25 and between nucleotides 31 and 33 impeded the processing of the 28-bp (R3) half and 38-bp (L3) half of *telRL*, respectively, while the respective other half was processed. Interestingly, a nick placed between positions 21 and 22 and at the positions 34 and 35 completely abolished the protelomerase activity, indicating that these linkages are critical for protein–substrate interaction (Fig. 8). The *telO* sequence of N15 is also present in phage ϕKO2, whereas flanking sequences exhibit some mismatches (Fig. 4). Nevertheless, the protelomerases and target sites of these phages are interchangeable, supporting the notion that the 16 nucleotides in the middle of *telO* are essential for processing. Huang et al. (2004) isolated the protelomerases of both N15 and ϕKO2. The TelK enzyme cleaved its own target with the same efficiency as *telRL* of N15. The telomere resolution site of phage PY54 was not a suitable substrate for TelK. Figure 4 shows that ten nucleotides located around the centre of dyad symmetry within *telO* are identical in all three palindromes. However, the critical nucleotides mentioned above deviate in PY54. When the positions 15, 16, 27 and 28 of the PY54 palindrome were substituted by nucleotides matching the N15/ϕKO2 *telO* core sequence, *telRL* of PY54 was processed by TelK (Fig. 4). These data unequivocally demonstrate the impact of the central nucleotides of *telO* for telomere resolution.

3.2.4 The *telRL* Site is Cleaved by Staggered Cuts Six Base Pairs Apart Around the Centre of Dyad Symmetry

In order to determine the cleavage sites on *telRL*, Huang et al. (2004) used the duplex oligonucleotides already described. However, in this experiment the substrates were not nicked and the labelled phosphate was incorporated at different positions along the length of the *telRL* target on the top strand. After processing, those substrates whose phosphate is on the left of the cleavage site would retain the label on the 28-bp product. If the label is on the right side of the cleavage site, the 38-bp product would be labelled. When this set of substrates was treated with TelN, the labelled

phosphates up to nucleotide 25 were identified on the 28-bp product, while a phosphate at position 26 and beyond remained with the 38-bp product. Thus, cleavage of the top strand occurs between the nucleotides 25 and 26, three nucleotides away from the dyad symmetry centre (Fig. 8). The same experiments were performed with substrates in which the bottom strand was labelled. Similar to the top strand, cleavage occurred three nucleotides right of the dyad symmetry centre between positions 31 and 32. Hence the protelomerase makes during the generation of the hairpins a pair of dyad-symmetric staggered cuts 6 bp apart that leave 5′ protruding ends (Fig. 8).

3.2.5 TelN Binds Covalently to the 3′-Phosphoryl of the Cleaved Strands

Processing of the telomere resolution site can be divided into three steps: (1) recognition and binding of the substrate, (2) cleavage at the target site, and (3) joining (transesterification) yielding a linear molecule with covalently closed ends. Following cleavage, a transient covalent intermediate between the protelomerase and its target is generated through nucleophilic substitution by the hydroxyl group of an amino acid (probably Tyr-424). Then the new phosphodiester bond with the complementary strand is constituted in a transesterification reaction, catalysed by the TelN–DNA complex.

At which end of the target is TelN bound upon cleavage? To answer this question, Deneke et al. (2002) applied a *telRL* substrate ($(TA)_7$, see Fig. 8) that was cut but not joined by TelN. One of the products was isolated and labelled at its 3′ or 5′ ends. After cleavage with a restriction enzyme, the fragments were analysed by autoradiography. The 3′ labelling showed only one fragment carrying the label, whereas two fragments were visible by labelling of the 5′ ends. It can be concluded that one 3′ end was decorated with a TelN molecule which prevented labelling. Huang et al. (2004) used suicide substrates with a nick for labelling. They could demonstrate that one protelomerase molecule is linked to the 3′-phosphoryl of nucleotide 25 on the top strand and a second one to the corresponding residue of nucleotide 32 on the bottom strand. Hairpin formation then occurs by ligation of the 3′-phosphoryl with a 5′-OH created by cleavage by the second protelomerase molecule.

The formation of 3′ covalent protein–DNA intermediates is a hallmark of tyrosine recombinases and therefore protelomerases belong to this family of enzymes. The complex of protelomerase and its target site *telRL* has the potential to form a cruciform structure with two protruding stem-loops in which the cleavage and ligation sites are located at the base of the loops. Alternatively, a duplex structure with DNA in the Z conformation is conceivable (Deneke et al. 2000; Huang et al. 2004) (Fig. 7).

3.3 Protelomerases are also Involved in Linear Plasmid Replication and Maintenance

The phage-encoded protelomerase is essential for the conversion of the phage genome into the linear plasmid prophage. Regions on the N15 and PY54 genomes that are important for plasmid maintenance have been defined by the construction of replicative linear and circular miniplasmid derivatives (Hertwig et al. 2003a; Vostrov et al. 1992). It was demonstrated that for the replication of a circular miniplasmid, two phage-encoded loci are essential: (1) the reading frame of the replication initiation gene *repA* and (2) its origin located within the 3′ portion of *repA* (Mardanov and Ravin 2006; Ziegelin et al. 2005). The latter publication provides data on the 212-bp origin of PY54 and shows that some structural elements of the *ori* sequence are also present in the proposed origins of N15 and ϕKO2. Contrary to circular plasmids, linear plasmid replication requires the presence of the protelomerase gene and its target sequence. Ravin et al. (2001) showed that TelN and *telRL* constitute a functional unit acting on other replicons (F, P1) by conferring linearity on mini-F and mini-P plasmids. Moreover, the TelN/*telRL* system is also well suited to generate linear, closed mini-DNA with high stability in mammalian cells (Heinrich et al. 2002).

Transcription studies on N15 and PY54 documented that the protelomerase gene is transcribed not only early after infection but also during the lysogenic stage (Hertwig et al. 2003b; Ravin et al. 2000). The data imply that the protelomerase might additionally play a role in the replication of the linear prophage. In this context, Ravin et al. (2001) investigated TelN activity in vivo. They constructed an N15-based linear plasmid with a deletion in the *telN* gene, which could replicate only in the presence of TelN provided in trans by a second plasmid. On this circular plasmid, *telN* expression was under the control of an inducible promoter. Upon repression of the cloned *telN* gene, intermediates of the linear N15 plasmid replication accumulated. The structure of these molecules was investigated by restriction analyses and Southern hybridization exhibiting circular head-to-head dimer molecules that were cut into linear monomers after protelomerase synthesis was induced again. Using electron microscopy, the structures of partly replicated intermediates of an intact linear N15 miniplasmid were analysed (Ravin et al. 2003). Three different types of replicating molecules could be identified (Fig. 9a). The first type contained an internal bubble proximal to one end of the linear plasmid. It probably represents an early step of bidirectional replication. *RepA* is located close to the *telL* plasmid end. The authors had shown that N15 prophage replication starts at the *ori* site located within the *repA* gene and that it proceeds bidirectionally. Type 2 molecules were circles located at the end of the linear plasmid and might result from replication to one end (*telL*) of the molecule. Type 3 molecules were Y-shaped having

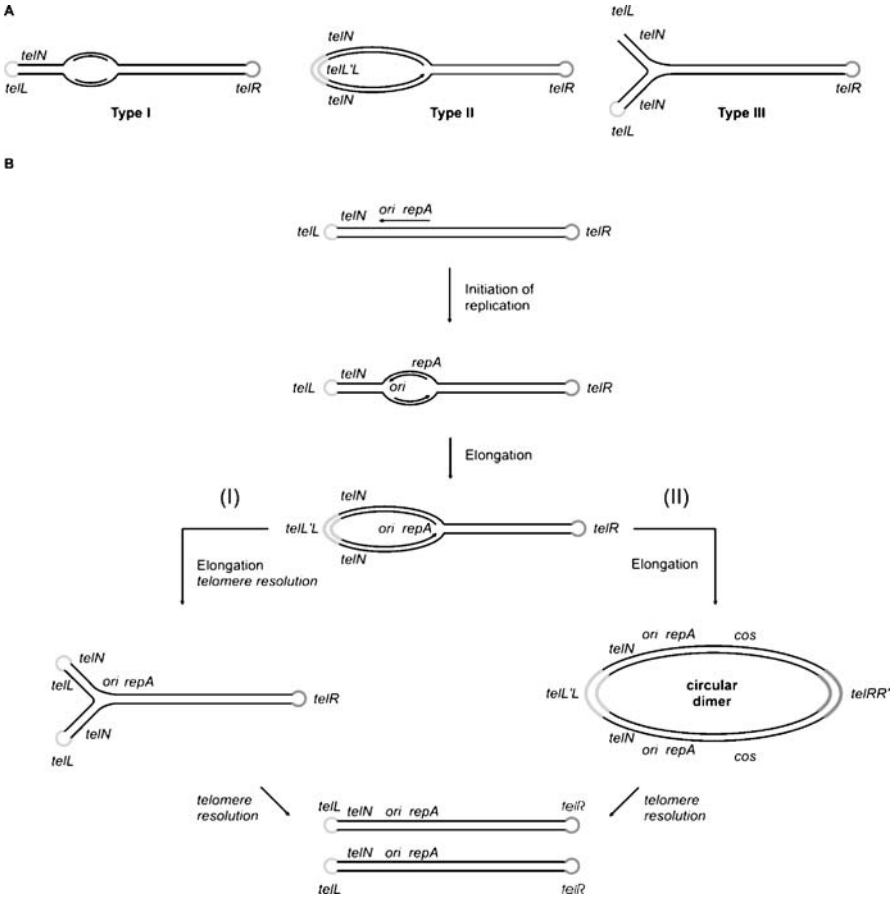


Fig. 9 Model of N15 plasmid replication, as proposed by Ravin et al. (2001). **a** Types of replicating molecules identified by electron microscopy. **b** Model of N15 plasmid prophage replication. (*I*) Protelomerase cuts before completion of replication, resulting in a Y-like structure. (*II*) Replication is finished before protelomerase cutting, giving a circular head-to-head dimer. *telL, telR*, left and right telomere ends of N15 prophage; *telN*, protelomerase gene; *ori*, origin of replication located within the gene *repA* (see text for details)

two equal-length arms consistent with the generation of a single fork. These molecules could result from TelN cleavage of the circle in type 2 molecules. Following the obtained data, Mardanov and Ravin (2006) suggested the following model of N15 plasmid replication (Fig. 9b). Replication is initiated at the origin located within *repA*, follows the θ mode of DNA replication and proceeds bidirectionally. After duplication of *telL*, TelN processes this site creating hairpin ends and a Y-shaped structure is formed. After replication of the right telomere and subsequent cleavage, two linear molecules

are produced. Under specific conditions, replication may precede end resolution and circular head-to-head dimers appear which then are processed by TelN.

The analysis of the N15 genome sequence revealed a plausible promoter upstream from the protelomerase gene (Ravin et al. 2000). As shown in Fig. 10, this putative promoter is flanked by the L2 and L3 repeats. Thus, it can be assumed that the expression of the *telN* gene is subjected to TelN autoregulation.

Dorokhov et al. (2003) studied the expression of the N15 *sop* operon that is located adjacent to the *tel* site (Fig. 2). The operon codes for two proteins (SopA and SopB) that are essential for the stable inheritance of the N15 plasmid (Ravin and Lane 1999). While SopB interacts with four centromere-like sites (*sopC*) scattered on the N15 genome, SopA is a transcriptional repressor of the partition operon (Grigoriev and Lobocka 2001). It binds to a promoter (P1) that was identified in short distance to the *sop* operon (Dorokhov et al. 2003). Further upstream of the *sopAB* genes, several additional putative promoters have been identified. Their activity was investigated by Northern blot hybridization of RNA isolated from cells carrying transcriptional fusions of upstream fragments to *lacZ* on high copy-number plasmids and subsequent primer extension mapping. One strong promoter (P2) has been detected that flanks the R2 site (Fig. 10). This P2 promoter was tightly repressed by basic levels of TelN but was not sensitive to the action of SopA and SopB. It can be reasoned that P2 repression is the consequence of binding of TelN to the repetitive sequences and particularly to R2. Hence TelN apparently exerts multiple functions: (1) it is responsible for the generation of the linear plasmid, (2) it cleaves intermediate products of linear plasmid replication at *telRL*, (3) it is a repressor for its own gene and (4) it is involved in the regulation of partitioning and therefore important for plasmid maintenance. Such different activities most likely require various protein domains.

To dissect the protelomerase functions, TelN proteins with deletions at the N or C terminus or with point mutations were analysed in terms of DNA binding and telomere resolution (unpublished results). Three major domains were identified. For unspecific binding of DNA, the presence of the N-terminal

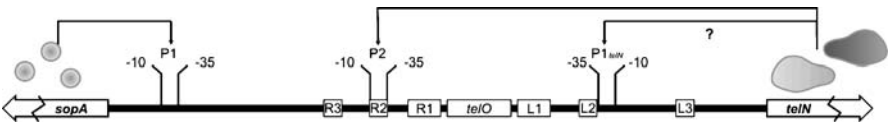


Fig. 10 DNA region between the N15 genes *sopA* and *telN*. The telomere resolution site *telRL* comprising *telO* and the repeats R1 and L1 is boxed. The boxes R3, R2, L2 and L3 show repeats belonging to the telomere occupancy site *tos*. P1*sop* is the promoter repressed by Sop. The promoters P2*sop* and P1*telN* are repressed by TelN and are located close to the *telRL* binding site

amino acid residues is mandatory. This region is also important for telomere resolution. The main catalytic domain is located around box C, which contains the essential tyrosine 424 (Fig. 3). Finally, the C-terminal region of TelN is responsible for the specific recognition of the *telRL* target. It can be expected that some additional domains exist, which are involved in the multiple functions of this interesting enzyme.

4 Perspectives

The mode of action of how protelomerases generate the covalently closed ends of the three known linear plasmid prophages has been elucidated to a large extent. Although most of the work has been done on the N15 protelomerase TelN, the homologies to TelK and TelY and the similar target sites of the protelomerases suggest the same principle by which the processing of the telomere resolution site occurs. To define the domains of the protelomerases involved in target binding, cleavage and joining, structural analyses of the proteins certainly would supply valuable information. Another aspect to be clarified pertains to the regulation of protelomerase synthesis. This enzyme is essential for the establishment and maintenance of the lysogenic cycle. However, it can be predicted that protelomerase activity would perturb lytic replication by cleaving the linear phage genomes. Hence the question arises of how protelomerase activity ceases upon induction. Transcription studies disclosed expression of the protelomerase genes of N15 and PY54 at the late stage of lytic growth (Hertwig et al. 2003b; Ravin et al. 2000). Huang et al. (2004) reported that TelK may act stoichiometrically and becomes inactive at the end of one transesterification. This scenario would provide a measure to ensure that no excess activity is present when it is no longer needed. It is also conceivable that protelomerase activity may be abolished by inactivation of the protein, by degradation or by an inhibitor protein that has not yet been identified. However, it cannot be excluded that the protelomerase also plays a role during lytic replication.

Acknowledgements The author thanks Erich Lanka for constructive discussions. The assistance of Jens A. Hammerl in preparing the figures is highly appreciated.

References

Bankhead T, Kobryn K, Chaconas G (2006) Unexpected twist: harnessing the energy in positive supercoils to control telomere resolution. Mol Microbiol 62:895–905

Casjens SR, Gilcrease EB, Huang WM, Bunny KL, Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW (2004) The pKO2 linear plasmid prophage of *Klebsiella oxytoca*. J Bacteriol 186:1818–1832

Danko AS, Luo M, Bagwell CE, Brigmon RL, Freedman DL (2004) Involvement of linear plasmids in aerobic biodegradation of vinyl chloride. Appl Environ Microbiol 70:6092–6097

Deneke J (2002) Das Tyrosinintegrase-Analog TelN katalysiert die telomere resolution im Bakteriophagen N15. PhD thesis, Freie Universität Berlin, Germany (in German)

Deneke J, Ziegelin G, Lurz R, Lanka E (2000) The protelomerase of temperate *Escherichia coli* phage N15 has cleaving–joining activity. Proc Natl Acad Sci USA 97:7721–7726

Deneke J, Ziegelin G, Lurz R, Lanka E (2002) Phage N15 telomere resolution. Target requirements for recognition and processing by the protelomerase. J Biol Chem 277:10410–10419

Dorokhov BD, Lane D, Ravin NV (2003) Partition operon expression in the linear plasmid prophage N15 is controlled by both Sop proteins and protelomerase. Mol Microbiol 50:713–721

Esposito D, Scocca JJ (1997) The integrase family of tyrosine recombinases: evolution of a conserved active site domain. Nucleic Acids Res 25:3605–3614

Fetzner S, Kolkenbrock S, Parschat K (2007) Catabolic Linear Plasmids. Microbiol Monogr 7

Grigoriev PS, Lobocka MB (2001) Determinants of segregational stability of the linear plasmid-prophage N15 of *Escherichia coli*. Mol Microbiol 42:355–368

Groth AC, Calos MP (2004) Phage integrases: biology and applications. J Mol Biol 335:667–678

Heinrich J, Schultz J, Bosse M, Ziegelin G, Lanka E, Moelling K (2002) Linear closed mini DNA generated by the prokaryotic cleaving–joining enzyme TelN is functional in mammalian cells. J Mol Med 80:648–654

Hendrix RW, Hatfull GF, Smith MC (2003) Bacteriophages with tails: chasing their origins and evolution. Res Microbiol 154:253–257

Hertwig S, Klein I, Lurz R, Lanka E, Appel B (2003a) PY54, a linear plasmid prophage of *Yersinia enterocolitica* with covalently closed ends. Mol Microbiol 48:989–1003

Hertwig S, Klein I, Schmidt V, Beck S, Hammerl JA, Appel B (2003b) Sequence analysis of the genome of the temperate *Yersinia enterocolitica* phage PY54. J Mol Biol 331:605–622

Huang WM, Joss L, Hsieh T, Casjens S (2004) Protelomerase uses a topoisomerase IB/Y-recombinase type mechanism to generate DNA hairpin ends. J Mol Biol 337:77–92

Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, Hendrix RW (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol 299:27–51

Jumas-Bilak E, Michaux-Charachon S, Bourg G, Ramuz M, Allardet-Servent A (1998) Unconventional genomic organization in the alpha subgroup of the Proteobacteria. J Bacteriol 180:2749–2755

Kobryn K (2007) The Linear Hairpin Replicons of *Borrelia burgdorferi*. Microbiol Monogr 7

Kobryn K, Chaconas G (2002) ResT, a telomere resolvase encoded by the Lyme disease spirochete. Mol Cell 9:195–201

Krum JG, Ensign SA (2001) Evidence that a linear megaplasmid encodes enzymes of aliphatic alkene and epoxide metabolism and coenzyme M (2-mercaptoethanesulfonate) biosynthesis in *Xanthobacter* strain Py2. J Bacteriol 183:2172–2177

Malinin A, Vostrov AA, Rybchin VN, Svarchevsky AN (1992) Structure of ends of linear plasmid N15. Mol Gen Mikrobiol Virusol 5–6:19–22 (in Russian)

Mardanov AV, Ravin NV (2006) Functional characterization of the repA replication gene of linear plasmid prophage N15. Res Microbiol 157:176–183

Plasterk RH, Simon MI, Barbour AG (1985) Transposition of structural genes to an expression sequence on a linear plasmid causes antigenic variation in the bacterium *Borrelia hermsii*. Nature 318:257–263

Popp A, Hertwig S, Lurz R, Appel B (2000) Comparative study of temperate bacteriophages isolated from *Yersinia*. Syst Appl Microbiol 23:469–478

Ravin N, Lane D (1999) Partition of the linear plasmid N15: interactions of N15 partition functions with the sop locus of the F plasmid. J Bacteriol 181:6898–6906

Ravin NV, Strakhova TS, Kuprianov VV (2001) The protelomerase of the phage-plasmid N15 is responsible for its maintenance in linear form. J Mol Biol 312:899–906

Ravin NV, Kuprianov VV, Gilcrease EB, Casjens SR (2003) Bidirectional replication from an internal ori site of the linear N15 plasmid prophage. Nucleic Acids Res 31:6552–6560

Ravin VK, Golub EI (1967) A study of phage conversion in *Escherichia coli* I. The acquisition of resistance to bacteriophage T1 as a result of lysogenization. Genetik 4:113–121 (in Russian)

Ravin VK, Ravin N, Casjens S, Ford ME, Hatfull GF, Hendrix RW (2000) Genomic sequence and analysis of the atypical temperate bacteriophage N15. J Mol Biol 299:53–73

Rybchin VN, Svarchevsky AN (1999) The plasmid prophage N15: a linear DNA with covalently closed ends. Mol Microbiol 33:895–903

Stoppel RD, Meyer M, Schlegel HG (1995) The nickel resistance determinant cloned from the enterobacterium *Klebsiella oxytoca*: conjugational transfer, expression, regulation and DNA homologies to various nickel-resistant bacteria. Biometals 8:70–79

Svarchevsky AN, Rybchin VN (1984) Physical mapping of plasmid N15 DNA. Mol Gen Microbiol Virusol 10:16–22 (in Russian)

Tardif G, Greer CW, Labbe D, Lau PC (1991) Involvement of a large plasmid in the degradation of 1,2-dichloroethane by *Xanthobacter autotrophicus*. Appl Environ Microbiol 57:1853–1857

Vostrov AA, Malinin A, Rybchin VN, Svarchevsky AN (1992) Construction of linear plasmid vectors for cloning in *Escherichia coli* cells. Genetika 28:186–188 (in Russian)

Ziegelin G, Tegtmeyer N, Lurz R, Hertwig S, Hammerl J, Appel B, Lanka E (2005) The repA gene of the linear *Yersinia enterocolitica* prophage PY54 functions as a circular minimal replicon in *Escherichia coli*. J Bacteriol 187:3445–3454

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_098/Published online: 29 June 2007

Retroplasmids: Linear and Circular Plasmids that Replicate via Reverse Transcription

Jeffrey T. Galligan · John C. Kennell (✉)

Department of Biology, Saint Louis University, St. Louis, MO 63103, USA
kennellj@slu.edu

1 Introduction 164

2 Circular Retroplasmids 167
2.1 Circular Retroplasmid Replication 169

3 Linear Retroplasmids 172
3.1 Linear Retroplasmid Replication 174
3.1.1 Characterization of In Vivo Replication Intermediates 174
3.1.2 Reverse Transcription 175
3.1.3 Linear Retroplasmid Replication Cycle 177
3.2 Double Hairpin Linear Retroplasmids 179

4 Evolutionary Significance 180

References 182

Abstract Retroplasmids are a unique group of genetic elements that replicate via reverse transcription. They are small, double-stranded DNAs that reside in mitochondria of certain filamentous fungal species. Retroplasmids are divided into two groups based on structure – circular and linear – and each group has distinct replication mechanisms. The reverse transcriptases (RTs) encoded by retroplasmids are deeply rooted in RT phylogeny and studies of retroplasmid replication have revealed unique mechanisms of reverse transcription that collectively suggest that retroplasmids are related to primitive genetic elements. Here, we review the types of retroplasmids that have been reported to date and focus on the studies of the replication cycle of the Mauriceville plasmid of *Neurospora crassa*, a prototype of circular retroplasmids, and studies of the pFOXC plasmids of *Fusarium oxysporum*, which are the first linear genetic elements shown to replicate via reverse transcription. The structural and mechanistic features of circular and linear retroplasmids suggest that at some point they could have been common ancestors of all known contemporary elements that replicate via reverse transcription and may hold clues to events associated with the transition of the RNA to DNA/protein world. While most mitochondrial plasmids appear to be benign, circular retroplasmids can integrate into mitochondrial DNAs and cause senescence. The significance of host–retroplasmid interactions are discussed in regard to the potential influence retroplasmids may have had on the evolution of mitochondrial DNAs and eukaryotic cells in general.

1 Introduction

Mitochondrial (mt) plasmids are defined as small (< 1–15 kb), autonomously replicating genetic elements that show no significant homology to mitochondrial DNA and have independent evolutionary origins. They are structurally diverse and are widely distributed in certain taxa (e.g., filamentous fungi), yet absent in others (e.g., animals; see Chap. 9, in this volume). They differ from bacterial plasmids in several important areas: most mt plasmids encode their own polymerase(s) that control key steps in replication, thus they are more like viruses or bacteriophages and, except in rare incidences, mt plasmids do not provide an obvious selective advantage to their host (see Chap. 8, in this volume). While most are considered to be benign, a few mt plasmids appear to provide a small benefit to their host (Leon et al. 1989; Bok et al. 2000), while others are clearly detrimental and are associated with senescence (reviewed in Griffiths 1992) or attenuation of virulence (reviewed in Bertrand and Baidyaroy 2002). Structural and phenotypic differences between mt plasmids and their presumed bacterial counterparts, including bacteriophages, likely reflect unique selective pressures experienced in residing in an organelle that has been transformed since its presumed former existence as a free-living α-proteobacterial-like bacterium (Andersson 1998). In general terms, mt plasmids can be considered to be selfish genetic elements that co-evolved with mitochondria and have adapted to residing within an organelle that has been extensively renovated. As expected of selfish genetic elements, mitochondrial plasmids exhibit drive and directly or indirectly increase their transmission frequencies by manipulating the inheritance of their mitochondrial host and can spread via lateral transfer mechanisms (reviewed in Griffiths 1995; Kennell, submitted for publication). In short, they may have had a significant impact on the development of eukaryotic cells and, in some lineages, continue to exert great influence over their hosts. Thus, they are of interest due to their unique evolutionary history and ways in which they influence mitochondrial function and evolution.

Mitochondrial plasmids can be categorized into discrete groups based on the type of nucleic acid associated with the replicative form: DNA, RNA, or those that have both RNA and DNA replication intermediates. DNA plasmids encode a DNA-dependent DNA polymerase and are subdivided into circular and linear types (see Chap. 8, in this volume). Mitochondrial double-stranded RNA elements that encode an RNA-dependent RNA polymerase are formally recognized as “mitoviruses” within the Narnaviridae family of naked viruses (Wickner et al. 2000); however, these elements lack an extracellular form and do not form virus particles, thus they have greater similarity to plasmids than to viruses (Kennell, submitted for publication). Retroplasmids are distinguished from DNA and RNA plasmids in that they encode an RNA-dependent DNA polymerase (i.e., reverse transcriptase, or RT) and replicate

through an RNA intermediate. Retroplasmids are divided into two groups based on structure, circular and linear, and each group has distinct replication mechanisms. Satellite plasmids represent an additional group of mt plasmids and are defined as genetic elements that depend on another plasmid for replication, yet are not derived from the helper element. The basic structural features of retroplasmids reported to date are illustrated in Fig. 1.

The origin of retroplasmids is thought to precede the endosymbiotic event that introduced the α-proteobacterial-like organism into the nascent eukaryotic cell. Phylogenetic analysis of retroplasmid RTs indicates that they form a monophyletic clade (Fig. 2) and are deeply rooted in RT phylogeny (Xiong and Eickbush 1990). This has led to the suggestion that retroplasmids are "molecular fossils", which are defined as contemporary genetic elements that are ancient in origin and provide insight into the evolutionary past (Maizels and Wiener 1993). If so, they are possible progenitors to a wide range of retroelements, including mobile group II introns, non-LTR retrotransposons, and perhaps telomerase (Wang and Lambowitz 1993; Eickbush 1997; Walther and Kennell 1999).

While the focus of this volume of *Microbial Monographs* is on linear genetic elements, the discovery of linear retroplasmids is relatively recent and many details of their replication cycle have yet to be characterized. In contrast, the Mauriceville retroplasmid was the first reported circular mitochondrial extrachromosomal element (Collins et al. 1981) and is currently the

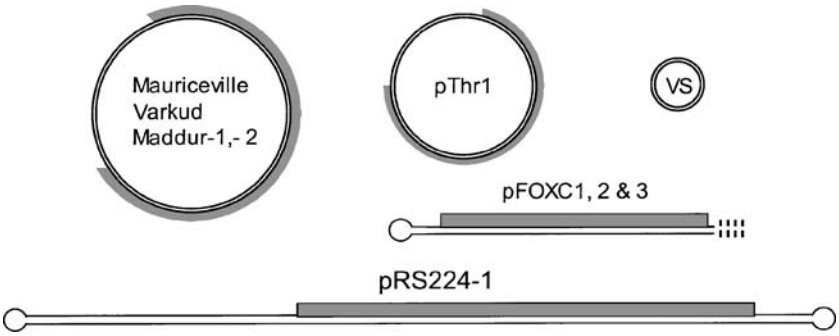


Fig. 1 Structural features of mitochondrial retroplasmids. The 3.6 kb Mauriceville plasmid of *Neurospora crassa*, 3.7 kb Varkud, Maddur-1, and Maddur-2 plasmids of *N. intermedia*, and 2.6 kb pThr1 plasmid of *Trichoderma harzianum* are circular dsDNAs that contain a single large open reading frame that encodes a reverse transcriptase (*gray box*). The VS plasmid of *N. intermedia* is an 881 bp circular dsDNA that lacks large ORFs and replicates as a satellite of the Varkud retroplasmid. The 1.9 kb pFOXC2 and pFOXC3 plasmids of *Fusarium oxysporum* are linear dsDNAs that have a clothespin structure with a hairpin at one end and three to five copies of a 5 bp repeat at the other terminus (*small vertical bars*). The 5.0 kb pRS224-1 plasmid of *Rhizoctonia solani* is a covalently-closed linear dsDNA having hairpins at both termini. Both clothespin and double-hairpin linear retroplasmids contain a single large ORF that encodes a reverse transcriptase (*gray box*)

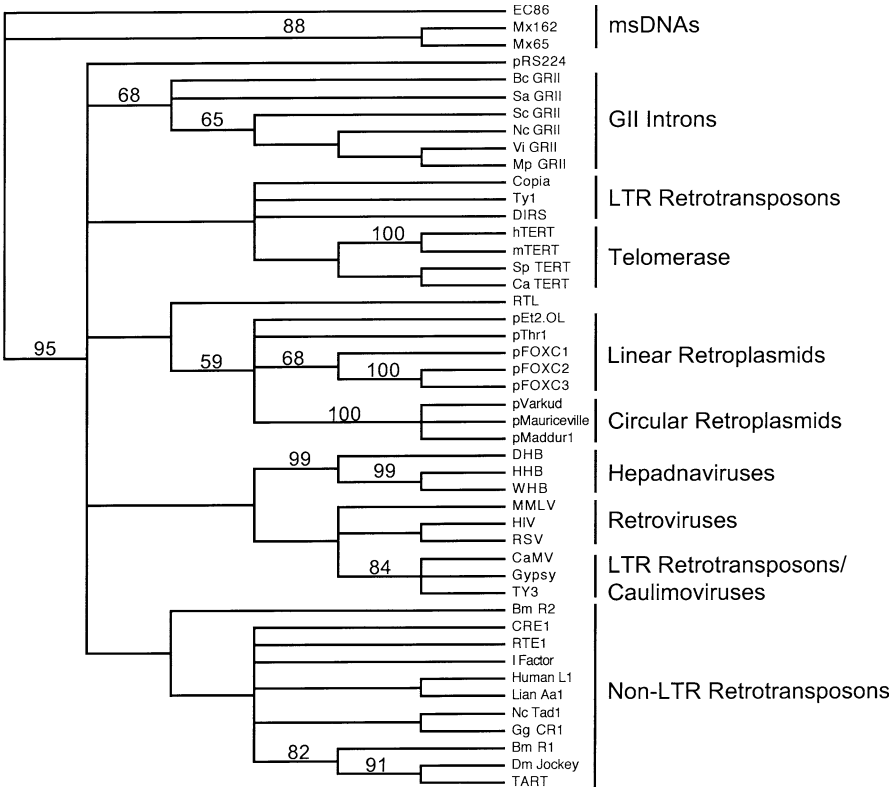


Fig. 2 Phylogenetic analysis of representative reverse transcriptase protein sequences. Amino acid sequences encompassing highly conserved domains I through VII (as described by Xiong and Eickbush 1990) were aligned in ClustalX and a maximum parsimony search using PAUP 4.0b10 generated 15 trees. The tree shown represents a strict consensus of recovered topologies and was rooted using the prokaryotic msDNAs. *Numbers* indicate the bootstrap support for the recovered clades. The accession numbers (in parentheses) of the elements used are as follows: EC86 (P23070), Mx162 (P23072), Mx65 (P23071), pRS224 (BAA94080), Bc GRII (ZP_00236429), Sa GRII (ZP_00784712), Sc GRII (AAA67532), Nc GRII (S07649), Vi GRII (AAB95256), Mp GRII (NP_054460), Copia (P04146), Ty1 (NP_058152), DIRS (AAA33195), hTERT (NP_937983), mTERT (NP_033380), Sp TERT (AAC49803), Ca TERT (AAF26733), RTL (CAA38646), pEt2.0L (CAA40486), pThr1 (NP_862333), pFOXC1 (AAD12231), pFOXC2 (NP_862680), pFOXC3 (NP_862679), pVarkud (AAA70286), pMauriceville (NP_041729), pMaddur1 (AAU25926), DHB (AAK85436), HHB (AAB59972), WHB (NP_671813), MMLV (P03355), HIV (AAL12183), RSV (CAA48535), CaMV (NP_056728), Gypsy (P10401), TY3 (AAA98435), Bm R2 (T18197), CRE1 (AAA75435), RTE1 (AAB71003), I Factor (AAA70222), Human L1 (AAC51279), Lian Aa1 (AAB65093), Nc Tad1 (AAA21781), Gg CR1 (AAA49027), Bm R1 (BAD82946), Dm Jockey (P21328), TART (T13173)

most thoroughly studied mitochondrial plasmid. Consequently, a brief review of its replication cycle is appropriate as features show direct similarity to certain steps in the replication of linear retroplasmids and can serve as a valuable model for retroplasmid replication in general.

2 Circular Retroplasmids

The Mauriceville plasmid of *Neurospora crassa* is a covalently closed circular DNA of approximately 3.6 kb. It has high sequence identity (97–99%) with the Varkud, Maddur-1, and Maddur-2 plasmids of a related species, *N. intermedia* (Nargang et al. 1984; Akins et al. 1988; D'Souza et al. 2005). The 2.6 kb pThr1 plasmid of *Trichoderma harzianum* is the only other circular retroplasmid reported outside of *Neurospora* species (Antal et al. 2002; Fig. 1). Recently, a possible related retroplasmid was identified among PCR-amplified products of mitochondrial DNAs of *Neurospora intermedia* isolates (Maas et al. 2007). An autonomously replicating form of the putative hybrid plasmid was not reported; thus it remains to be seen whether this represents a new linear or circular plasmid form. The Mauriceville retroplasmid has a single large open reading frame (ORF) that encodes a 710 amino acid polypeptide that has reverse transcriptase activity (Kuiper and Lambowitz 1988; Kuiper et al. 1990; Table 1). The plasmid replication cycle has been extensively characterized and in vitro studies demonstrate that the Mauriceville reverse transcriptase (pMaur-RT) has unique enzymatic properties (Wang and Lambowitz 1993a). Isolates containing Mauriceville or Varkud are phenotypically similar to wild-type strains, but unlike plasmid-free strains that show indefinite growth potential, repeated vegetative transfer of retroplasmid-containing strains often leads to growth defects and senescence. Presenescent cultures are found to contain variant forms of the plasmids that have cDNA copies of mitochondrial tRNAs or other sequences that derive from mitochondrial DNA (mtDNA; Chiang et al. 1994). These so-called "variant plasmids" may also suffer deletions in non-essential regions (Akins and Lambowitz 1990) or contain duplications of plasmid sequences (Stevenson et al. 2000).

Included among the circular retroplasmid group is the VS plasmid of *N. intermedia*. It is a small (881 bp), dsDNA that does not encode any large ORFs. The VS plasmid is only found in association with the Varkud retroplasmid and, in experiments in which plasmids were transferred between strains via transient hyphal fusion, the VS plasmid was always transferred along with the Varkud plasmid (Collins and Saville 1990). The VS plasmid appears to replicate via reverse transcription, using the RT encoded by the Varkud plasmid, thus it is also considered to be a satellite plasmid (Kennell et al. 1995). Notably, the VS RNA transcript is a ribozyme and shows highly efficient RNA cleavage and ligation activity (Saville and Collins 1990; Saville and Collins

Table 1 Mitochondrial retroplasmids

Retroplasmid	Type	Host	Size (bp)	RT size (aa)	Associated phenotype	Accession number	Refs.
Mauriceville	Circular	*Neurospora crassa*	3581	710	Senescence	NC_001570	Nargang et al. 1984
Varkud	Circular	*N. intermedia*	3675	710	Senescence	NC_001571	Akins et al. 1988
Maddur-1	Circular	*N. intermedia*	3614	710	Senescence	AY553872	D'Souza et al. 2005
Maddur-2	Circular	*N. intermedia*	3774	710	Senescence	AY553873	D'Souza et al. 2005
VS	Satellite	*N. intermedia*	881	–	–	M327994	Saville and Collins 1990
pThr1	Circular	*Trichoderma harzianum*	2619	606	–	AF163325	Antal et al. 2002
pFOXC1	Linear	*Fusarium oxysporum* f. sp. *conglutinans*	1900[a]	n.d.	–	AF005240	Kistler et al. 1996
pFOXC2	Linear	*F. oxysporum* f. sp. *raphani*	1905	527	–	AF124843	Walther and Kennell 1999
pFOXC3	Linear	*F. oxysporum* f. sp. *matthioli*	1836	527	–	AF124844	Walther and Kennell 1999
Et2.0L	Linear	*Epichloë typhina*	2000[a]	n.d.	–	X57200	Mogen et al. 1991
pRS224-1	Linear	*Rhizoctonia solani*	4986	887	–	AB035862	Katsura et al. 1997

n.d. not determined
[a] Estimated size

1991; Zamel et al. 2004). A variant form of the Varkud plasmid that contains a portion of the VS plasmid has also been identified (Arganoza and Akins 1995).

2.1 Circular Retroplasmid Replication

Studies of retroplasmid replication rely on analyzing polymerase activity in plasmid-containing ribonucleoprotein (RNP) particles isolated from mitochondria (Kuiper and Lambowitz 1988; Wang et al. 1992; Kennell et al. 1995; Walther and Kennell 1999) and characterization of in vivo replication intermediates (Stohl et al. 1982; Akins et al. 1986, 1989; Kennell et al. 1994; Nagasaka et al. 2003; Simpson et al. 2004). The limitations of efficiently transforming mitochondria outside of *S. cerevisiae* (Johnston et al. 1988) and *C. reinhardii* (Boynton and Gillham 1996) have been a major constraint in fully elucidating the mechanism of replication of mitochondrial plasmids. However, despite this limitation, studies of retroplasmid replication have revealed several surprising results, some of which have significant evolutionary implications.

The 3.6 kb Mauriceville plasmid is transcribed by the host mitochondrial RNA polymerase, and transcription initiation occurs approximately 260 nt upstream of the site at which cDNA synthesis begins (Kennell et al. 1994; Fig. 3). Transcription is thought to proceed by a rolling-circle mechanism. The resulting RNAs are cleaved at a specific site which generates unit-length transcripts (Akins et al. 1986; Kennell et al. 1994). The 3′ end of the plasmid transcript contains sequences that can potentially fold into a tRNA-like structure and terminate in tandem CCA sequences (Kuiper and Lambowitz 1988). These RNAs appear to function both as mRNAs for the synthesis of the pMaur-RT and as templates for (–) strand cDNA synthesis. Translation produces an 81 kDa polypeptide that has reverse transcriptase activity (Kuiper and Lambowitz 1988; Kuiper et al. 1990). In mitochondria, the pMaur-RT is tightly associated with the plasmid transcript in a ribonucleoprotein complex and in reactions having partially purified mt RNPs and labeled deoxynucleotides, pMaur-RT produces full-length (–) strand cDNA products (Kuiper and Lambowitz 1988; Wang et al. 1992; Kennell et al. 1994). Posttreatment of reaction products with RNase, protease, or alkali has no effect on the cDNA products, indicating they do not have large attached RNAs or proteins that could function as primers.

When liberated from endogenous RNAs by treatment with micrococcal nuclease (MN) or elution using polyethyleneimine (PEI), the pMaur-RT can be used in reactions containing exogenous RNAs. In these assays, pMaur-RT shows specificity for transcripts that contain the 3′ terminal tRNA-like structure and can initiate cDNA synthesis by two or more mechanisms (Wang et al. 1992; Wang and Lambowitz 1993a). When mt RNPs are treated with

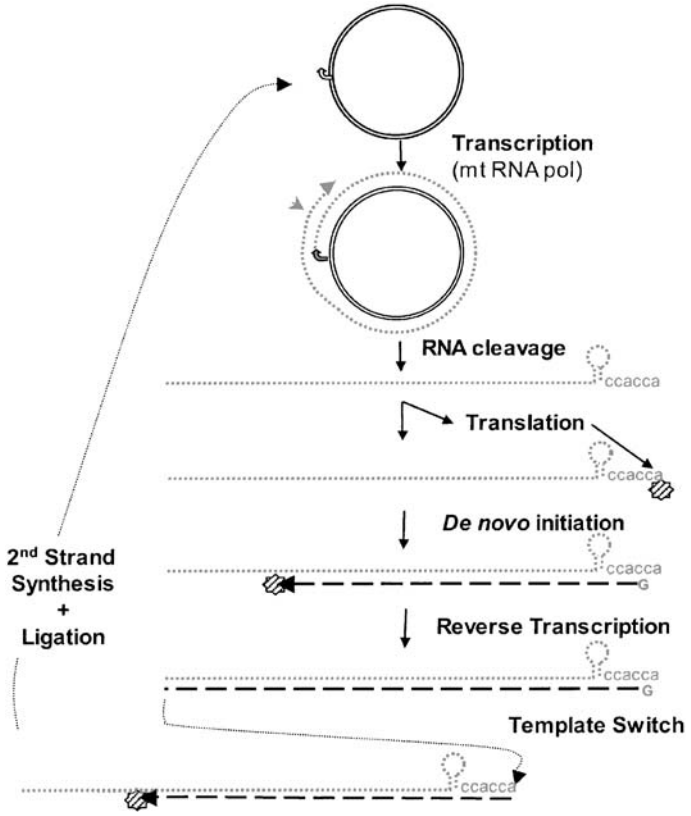


Fig. 3 Hypothetical replication cycle of circular retroplasmids. The double-stranded DNA plasmid is transcribed by the mt RNA polymerase and the transcript (*dotted line*) is cleaved at a site ~ 260 bp downstream from the promoter (*arrowhead*), yielding a monomer-length RNA having a tRNA-like structure and 3′ CCA. The transcript serves as both the mRNA for translation of the reverse transcriptase and as template for cDNA synthesis. The plasmid-encoded RT (*diagonally-striped star*) initiates (–) strand cDNA synthesis de novo, opposite the penultimate nucleotide of the transcript. Once the RT has reached the end of the RNA template, it can template switch to a second transcript, forming linear, concatenated cDNAs. Following (–) strand synthesis the RNA template is either degraded by the mt RNase H or displaced during (+) strand synthesis. Following the formation of the second strand, the ends of the plasmid DNAs may ligate to form the double-stranded circular DNA or homologous regions of concatenated dsDNAs may recombine to generate circular forms

micrococcal nuclease to degrade endogenous RNAs, the resulting RT preparations initiate cDNA synthesis using the 3′ hydroxyl of previously synthesized cDNAs that remain associated with the enzyme (Wang et al. 1992). This so-called "primer-dependent" initiation occurs at the extreme 3′ end of the RNA transcript. Subsequent studies showed that the pMaur-RT can also utilize synthetic DNA oligonucleotides, and shows little preference for specific se-

quences (Wang and Lambowitz 1993a; Chen and Lambowitz 1997). The ability to use DNAs (or cDNAs) to initiate new synthesis is reminiscent of template switching mechanisms associated with retroviral RTs; however, in the case of the pMaur-RT, DNA primers do not need to base pair with RNA templates, thus priming operates in a homology-independent manner (Kennell et al. 1994; Chiang and Lambowitz 1997; Chen and Lambowitz 1997). A similar mechanism appears to be used by non-LTR retrotransposons (Bibillo and Eickbush 2002).

When the pMaur-RT is isolated from mt RNP particles using PEI, it is free of associated nucleic acids and can initiate cDNA synthesis without a primer (de novo). Rather than starting at the extreme 3′ end of the plasmid transcript, primer-independent initiation occurs opposite the penultimate nucleotide of the RNA templates (C<u>C</u>A-3′), using guanine as the initiating nucleotide (Wang and Lambowitz 1993a). Initiation of DNA synthesis without the use of a primer is unprecedented and suggests that circular retroplasmid RTs have an evolutionary relationship with RNA-dependent RNA polymerases that initiate without the use of a primer at the 3′ end of RNA templates (RdRps; Wang and Lambowitz 1993a). Coupled with the basal position that retroplasmid RTs occupy in reverse transcriptase phylogenies, this finding supports the hypothesis that the *Neurospora* retroplasmids represent a contemporary example of a primitive genetic element, and suggests that they are related to early DNA-based life forms (Maizels and Weiner 1993).

Primer-independent initiation would result in the synthesis of (–) stand cDNA products that lack a terminal nucleotide. It is possible that a terminal T residue is added following synthesis or during second strand synthesis; however, this may not be necessary. Upon reaching the 5′ end of the RNA template, the pMaur-RT appears to template switch to the 3′ end of a second transcript, creating dimeric and multimeric cDNAs (Kennell et al. 1994). These concatenated molecules presumably serve as templates for second strand synthesis, and may obviate the need for a terminal nucleotide addition activity. At present, steps following (–) strand cDNA synthesis are poorly understood, and it is unclear whether the plasmid RT or another polymerase catalyzes the synthesis of the (+) strand. The RNA template may be displaced during second strand synthesis, or it could be degraded by a mitochondrial RNase H (Wang and Lambowitz 1993b). Following the formation of the second strand, the ends of the plasmid DNAs may ligate to create the double-stranded circular DNA form that, in turn, can recombine to generate the concatenated forms that predominate in mitochondria. Alternatively, homologous regions of concatenated dsDNAs could recombine to generate circular forms.

The finding that the Mauriceville plasmid transcript has a 3′ tRNA-like structure that ends in CCA suggests that the retroplasmids are related to certain plant RNA viruses (such as brome mosaic or turnip yellow mosaic virus; Miller et al. 1986; Kuiper and Lambowitz 1988). However, in vitro studies with the pMaur-RT show that the 3′ structure is less important than the ter-

minal CCA sequence (Chen and Lambowitz 1997). The pMaur-RT can also readily copy mitochondrial tRNAs and its promiscuous nature can lead to the formation of variant forms of dsDNA plasmids that contain cDNA copies of mitochondrial tRNAs (Akins et al. 1986, 1989; Kennell et al. 1994; Chiang et al. 1995; Chiang and Lambowitz 1997; Fox and Kennell 2001). The tRNA copies are inserted precisely downstream of the plasmid tRNA-like sequence and retain the posttranscriptionally added CC sequence, which indicates that they were copied via reverse transcription. Interestingly, the inserted sequences often include GC-rich sequences that derive from regions of the mt genome. Variant plasmids that contain cDNA copies of mt tRNAs can integrate into the mtDNA via homologous recombination or non-homologous mechanisms, and result in senescence of the host (Chiang et al. 1995; reviewed in Lambowitz and Chiang 1995). There is also a report of a variant plasmid that causes senescence without integrating into mtDNA (Stevenson et al. 2001). In this case, the variant plasmid replicates to a high copy number and interferes with the translation of mitochondrial genes.

The ability to copy non-plasmid-derived RNAs indicates that the RT can function in *trans*, a feature that is likely to have played an important role in maintenance of the VS satellite plasmid. The 881 bp circular VS plasmid is transcribed by the host mt RNA polymerase, and the VS RNA undergoes a self-catalyzed cleavage/ligation reaction to form a circular RNA that appears to function as a template for the Varkud RT. Steps following the synthesis of (–) strand cDNAs are presumably similar to those associated with the helper plasmid (Kennell et al. 1995).

3
Linear Retroplasmids

Linear retroplasmids are small, dsDNAs that have terminal hairpins (Katsura et al. 2001) or are in a "clothespin" structure that have a terminal hairpin at one end and telomere-like repeats at the other (Fig. 1; Walther and Kennell 1999). The pFOXC plasmids of *Fusarium oxysporum* are the most thoroughly characterized group of linear autonomously replicating retroelements described to date and serve as prototypical clothespin linear retroplasmids. Three related plasmids (termed pFOXC1, pFOXC2, and pFOXC3) have been found to reside in mitochondria of the fungal plant pathogen *Fusarium oxysporum*. *F. oxysporum* has a broad host range, and isolates are categorized based on the ability to cause disease in specific hosts [i.e., *F. oxysporum* forma specialis (f. sp.) *conglutinans* is pathogenic to cabbage while *F. oxysporum* f. sp. *raphani* infects radish]. Differences in host specialization are correlated with specific plasmid types, with pFOXC1, pFOXC2, and pFOXC3 being specific to *F. oxysporum* f. sp. *conglutinans*, f. sp. *raphani*, and f. sp. *matthioli*, respectively (Kistler and Leong 1986; Hirota et al. 1992; Kistler et al.

1997). Originally, this suggested that the plasmids play a role in host specificity or pathogenicity, and was tested by conducting pathogenicity assays with plasmid-free strains and strains in which plasmids had been transferred between different forma speciales (*conglutinans* and *matthioli*) that had been generated by protoplast fusion (Momol and Kistler 1992). Plasmid-free strains and non-parental combinations were found to incite disease based on the type of nuclear genome, rather than on mitochondrial or plasmid type, demonstrating that the plasmids do not dictate host specificity or pathogenicity (Momol and Kistler 1992).

The pFOXC plasmids are double-stranded DNA elements, 1.9 kb in length, and their linear structure was deduced by restriction digest analysis, electron microscopy, and $3' \rightarrow 5'$ exonuclease sensitivity (Kistler and Leong 1986). Subsequent studies showed that plasmids pFOXC2 and pFOXC3 possess a covalently closed hairpin at one terminus, while the other terminus consists of multiple tandem copies of a 5 bp sequence (Walther and Kennell 1999). The $5'$ end possesses a covalently bound terminal protein, a feature common to many linear genetic elements such as mitochondrial and cytoplasmic DNA plasmids, bacteriophage $\Phi 29$, and adenovirus (Salas 1991; Kempken 1995, Meinhardt et al. 1997; see Chap. 8, in this volume).

The pFOXC2 and pFOXC3 plasmids have been completely sequenced and each contains a single large ORF that is predicted to encode a 527 amino acid polypeptide. The polypeptide sequences have seven conserved amino acid motifs that are characteristic of reverse transcriptases and are most closely related to the RT encoded by the circular retroplasmids (30% identity within the conserved domains) of *Neurospora* spp. (Kistler et al. 1997; Walther and Kennell 1999). The pFOXC2 and pFOXC3 ORFs show 88% amino acid identity overall, and 93% identity within the seven conserved RT motifs (Walther and Kennell 1999). The polypeptides are predicted to be 62 kDa in size, which are amongst the smallest RTs described to date. They lack regions homologous to the RNAse H domain of other retroelements and the T motif of telomerase RTs (TERTs; McClure 1993; Nakamura et al. 1997; Walther and Kennell 1999). The pFOXC1 plasmid has only been partially sequenced, yet the incomplete open reading frame contains seven amino acid motifs conserved among RTs (Kistler et al. 1997; Walther and Kennell 1999). Interestingly, the predicted pFOXC1 polypeptide has much lower identity (39% within the highly conserved region) to the pFOXC2 and pFOXC3 RTs, indicating that pFOXC1 is distantly related (Walther and Kennell 1999; Fig. 2). Preliminary studies indicate that pFOXC1 has 3 bp terminal repeats rather than the 5 bp repeats of pFOXC2 and pFOXC3 (unpublished data).

The $3'$ terminal repeats of the pFOXC2 and pFOXC3 plasmids were identified by amplification of the downstream terminus via anchored PCR, cloning and sequencing of the amplified products (Walther and Kennell 1999). Both plasmids were found to have the same pentameric sequence ($5'$-ATCTA), and the number of repeats varied between two and five copies, suggesting that

the number of repeats is not fixed. To analyze the 5′ end, cycle sequencing of the isolated plasmid was employed. The results indicated that the 5′ terminus is complementary to the 3′ repeats (5′-TAGAT) and has the same relative number of repeats, suggesting the plasmid is blunt ended and lacks long single-stranded extensions. In addition, the mature plasmid contains a 5′-terminal protein, which is inferred by the inability to recover the plasmid from mitochondrial lysates following extraction with phenol, without pretreatment with a protease, as well as insensitivity to digestion with λ exonuclease that has 5′ → 3′ processivity (Walther and Kennell 1999).

To determine the nature of sequences upstream of the open reading frame, terminal plasmid restriction fragments were isolated and, following dephosphorylation and radiolabeling of the 5′ end using polynucleotide kinase, analyzed by electrophoresis in denaturing and non-denaturing polyacrylamide gels. Labeled restriction fragments migrated twice as fast through non-denaturing gels as through denaturing gels, confirming the presence of a hairpin (Walther and Kennell 1999). Terminal fragments were also amplified by anchored PCR, cloned, and sequenced, which revealed the presence of an inverted repeat.

3.1 Linear Retroplasmid Replication

Following steps that proved effective in elucidating the replication cycle of circular retroplasmids, in vivo replication intermediates of pFOXC2 and pFOXC3 were characterized and reverse transcriptase assays were performed using mt RNP particles from plasmid-containing strains.

3.1.1 Characterization of In Vivo Replication Intermediates

The 3′ ends of plasmid transcripts were analyzed to determine whether they terminate at a specific sequence. Full-length plasmid RNAs were isolated, tailed with A residues using polyA polymerase, and copied using reverse transcriptase with an oligo dT primer. The products were amplified, cloned, and sequenced. The length of the plasmid RNAs was heterogeneous and the number of repeats was, on average, slightly less than the number found at the 3′ end of plasmid DNAs. A similar approach was used to analyze the 5′ ends of in vivo (–) strand cDNA products. Like the plasmid DNA and RNA transcripts, the number of repeats at the 5′ end was variable; however, the cDNAs contained a greater number of repeats than the number found with the plasmid RNAs, and several cDNAs were longer than the average length of cloned sequences obtained from the mature plasmid DNA (Simpson et al. 2004). This suggested a mechanism exists for generating repeats during reverse transcription.

3.1.2 Reverse Transcription

The pFOXC2 and pFOXC3 ORF is predicted to encode a 62 kDa protein, and Western analysis using an antibody against a synthetic peptide derived from the pFOXC3-RT polypeptide sequence confirmed the presence of a ~60 kDa polypeptide in mt RNP particles (unpublished data). The pFOXC3-RT is concentrated in mt RNP particles and has reverse transcriptase activity. When mt RNPs from pFOXC-containing strains are incubated in appropriate buffers having [^{32}P]-dNTPs, high molecular weight labeled DNAs are produced. These reactions are insensitive to the DNA-dependent DNA polymerase inhibitor actinomycin D, and sensitive to pretreatment with RNase A (Walther and Kennell 1999). The products are cDNAs of approximately 1.9 kb in length and, when used as probes in Southern hybridizations, the products hybridize with plasmid (+) strand sequences indicating that the plasmid transcript serves as template.

The length of the (–) strand cDNA products matches the size of the plasmid transcript, suggesting the cDNA synthesis begins at the 3′ end of RNA templates. Posttreatment of (–) strand cDNA products with RNase, protease, or alkali did not affect the migration of the products in denaturing gels, indicating that the cDNAs were not bound by large RNA or protein primers (Simpson et al. 2004). The nature of the primer used in vivo has not been determined and is currently under investigation.

To gain a better understanding of the mechanistic properties of the pFOXC-RT, an in vitro reverse transcription system was developed that uses exogenously added RNA templates (i.e., exogenous reactions; Simpson et al. 2004). When liberated from endogenous RNAs via micrococcal nuclease digestion, the pFOXC-RT is able to copy in vitro synthesized RNAs added to reactions. To initially assess the template specificity of the pFOXC-RT, total mitochondrial RNA was added to reactions having MN-treated pFOXC-RT. The resulting cDNA products were found to hybridize to plasmid sequences rather than mtDNA fragments, indicating that the RT shows specificity for the plasmid transcript. Interestingly, unlike the circular retroplasmid RNAs, the 3′ ends of pFOXC RNAs are not predicted to be extensively base-paired and a recognition site has not been identified.

In exogenous reverse transcription assays having in vitro synthesized RNAs corresponding to the 3′ end of the plasmid transcript, the pFOXC-RT is capable of initiating cDNA synthesis using a variety of primers. In reactions containing the MN-treated pFOXC-RT and in vitro RNA templates, cDNA synthesis initiates via the extension of the 3′ hydroxyl of snapped-back RNAs. When DNA oligonucleotides complementary to regions of the RNA templates are included in the reactions, the pFOXC-RT can readily use the oligonucleotides as primers. In both instances, the pFOXC-RT shows little specificity for RNA templates. While the ability to use DNA oligonucleotides

and snapped-back RNAs as primers is not uncommon for retroviral RTs such as MMLV or AMV, comparative studies showed that the pFOXC-RT can use primers in an unconventional manner. The pFOXC-RT was found to require substantially less base-pairing between the DNA primers and RNA templates, and was able to extend snapped-back RNAs that had very minimal base-pairing interactions. DNA primers with as little as 5 bp of complementarity to sequences at the end of the RNA templates could be extended, and the pFOXC-RT showed a preference for primers bound at the 3′ terminus (Simpson et al. 2004).

In other reactions, the pFOXC-RT was also found to be able to copy DNA templates. It was observed that in reactions containing DNA oligonucleotides, small, highly abundant labeled DNAs were produced. These products derived from the extension of DNA primers that had self-annealed (Simpson et al. 2004). Experiments using a series of single oligonucleotides that had varying degrees of self-complementarity confirmed that the products derived from the copying of DNA templates and demonstrated that the pFOXC-RT could extend DNA primers that contained 3′ mismatches. Remarkably, primers having up to three terminal mismatches could be extended (Simpson et al. 2004). Taken together, these studies indicate that the pFOXC-RT has a remarkably loose specificity for primers and that it is capable of extending primers that have minimal base-pairing interactions, including those that have mismatched 3′ termini.

While there are some similarities to the mechanism of reverse transcription associated with circular retroplasmids of *Neurospora* spp., the activity of the pFOXC-RT differs in many regards. This is somewhat surprising considering that retroplasmid RTs are considered to be members of the same monophyletic family of RTs. The differences detected thus far include:

1. The pFOXC-RT readily uses the 3′ OH of RNA templates to prime cDNA synthesis, while the pMaur-RT rarely uses RNA primers and appears to depend on a specific RNA sequence, rather than base-pairing of the 3′ end of the RNA primer (Wang and Lambowitz 1993a)
2. The pFOXC-RT is able to use DNA primers that anneal to internal regions of the transcript (Simpson et al. 2004), whereas the pMaur-RT cannot (Wang et al. 1992; Chen and Lambowitz 1997)
3. The pFOXC-RT is able to copy DNA templates, while the pMaur-RT cannot
4. Treatment of pFOXC-containing mt RNPs with micrococcal nuclease results in RT preparations free of endogenous nucleic acids, whereas MN-treated pMaur-RT preparations contain endogenous cDNA products that are used as primers for reverse transcription (Wang et al. 1992)
5. The pFOXC-RT has low specificity for RNAs, whereas the pMaur-RT highly prefers RNAs having a 3′ terminal CCA sequence (Chen and Lambowitz 1997)

The enzymatic differences between RTs encoded by circular and linear retroplasmids likely reflect specific requirements involved in cDNA initiation and plasmid replication. For example, the flexibility the pFOXC-RT exhibits in using different types of primers and its ability to utilize minimally base-paired primer–template combinations could be properties that are essential for maintaining terminal repeat structures. In addition, its ability to extend mismatched 3′-terminal nucleotides would ensure that RNAs having heterogeneous terminal sequences could be copied. To our knowledge, the only other DNA polymerase known to extend primers having 3′ mismatches is the *Tetrahymena* TERT, and this property is associated with the "de novo" repeat addition that adds telomeres to non-telomeric DNAs (Wang and Blackburn 1997; Wang et al. 1998; Ware et al. 2000).

3.1.3 Linear Retroplasmid Replication Cycle

When the findings of the characterization studies of plasmid replication intermediates are combined with the properties of the pFOXC-RT revealed from in vitro reverse transcription assays, a model for linear clothespin retroplasmid replication can be developed (Fig. 4). Replication begins with transcription of the plasmid DNA, most likely by a host mitochondrial RNA polymerase. The length of the RNA transcripts suggests transcription likely begins on the top strand and proceeds around the hairpin to the terminal repeat region (Walther and Kennell 1999; Simpson et al. 2004). The full-length transcript potentially serves as both an mRNA for the reverse transcriptase and as a template for (–) strand cDNA synthesis. Following translation, the pFOXC-RT may bind directly to the RNA in *cis* and initiate cDNA synthesis at or near the extreme 3′ end. Although the nature of the primer has yet to be determined, in vitro studies indicate that the pFOXC-RT can use DNA or snapped-back RNA primers. Due to the phylogenetic relationship with the pMaur-RT, the pFOXC-RT may also be able to initiate cDNA synthesis de novo. In addition, since the mature plasmid contains a 5′-linked protein, it is also conceivable that a protein is used as a primer. In this scenario, the RT may function as a primer itself, as has been shown for hepadnaviruses, such as the hepatitis B virus (reviewed in Ganem and Schneider 2001). Figure 5a shows four potential initiation mechanisms.

Characterization of the in vivo replication intermediates suggests that the telomere-like repeats may be added during reverse transcription (Simpson et al. 2004). A potential mechanism by which repeats may be added is illustrated in Fig. 5b, and is based on mechanisms used by protein-primed linear DNA elements (Salas, 1991). In this model, cDNA synthesis initiates via the extension of a snapped-back RNA and the pFOXC-RT pauses after copying a few nucleotides. The nascent cDNA-RT complex may slide back and become repositioned on the plasmid RNA, prior to elongation. This may occur more

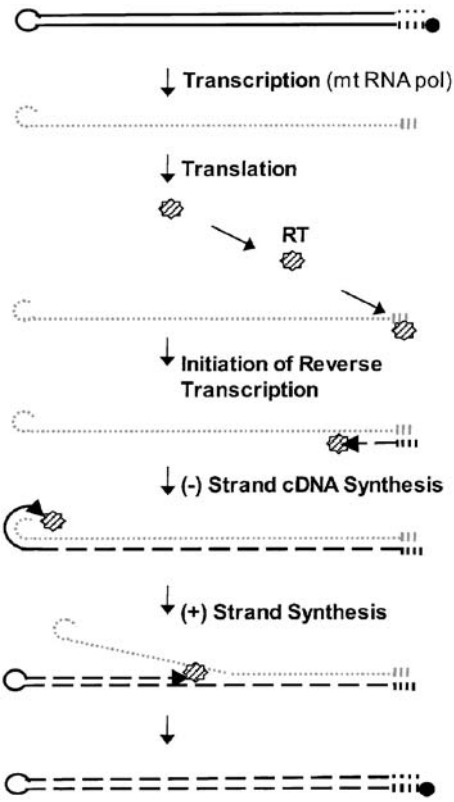


Fig. 4 Hypothetical replication cycle of linear clothespin retroplasmids. Replication begins with transcription of the plasmid DNA by the host mt RNA polymerase. Transcription begins on the top strand and proceeds around the hairpin to the 3′ repeat region (*vertical bars*). Transcripts (*dotted line*) appear to contain fewer 3′ repeats than are associated with plasmid DNAs (three repeats vs. four repeats). The transcript serves as both an mRNA and a template for (–) strand cDNA synthesis. Following translation, the RT (*diagonally-striped star*) associates with the 3′ end of the plasmid transcript where it initiates cDNA synthesis and may generate additional copies of the 5 bp repeat. Potential mechanisms used for initiation and repeat addition are shown in Fig. 5. Reverse transcription continues until the RT reaches the 5′ end of the RNA template, generating a full-length (–) strand cDNA (*dashed line*). The 3′ end of the (–) strand cDNA may serve as a primer for (+) strand synthesis, which is catalyzed by the plasmid RT or host DNA polymerase. The RNA template is degraded by a host RNAse H or displaced during the synthesis of the (+) strand. Following second strand synthesis, the RT (or mt DNA polymerase) may associate with the 5′ end to potentially function as a protective cap

than once and would generate (–) strand cDNAs that have a greater number of repeats than associated with the RNA template.

Following the synthesis of a full-length (–) strand cDNA, the RNA template is presumably degraded by a host RNAse H, or displaced during second

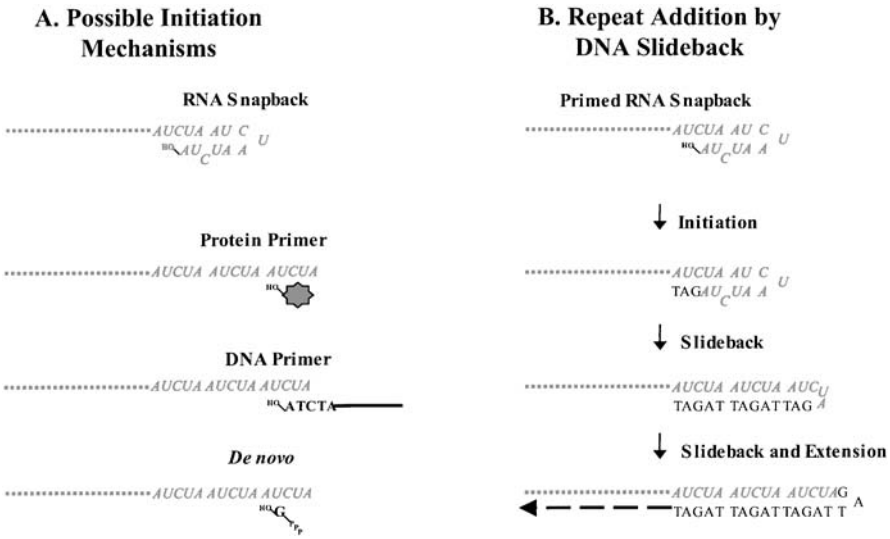


Fig. 5 Models of initiation of cDNA synthesis and repeat addition. **a** Potential mechanisms used to initiate (–) strand cDNA synthesis of pFOXC3. The pFOXC3 transcript may snap back upon itself and the 3′ hydroxyl of the terminal nucleotide is used as a primer (RNA snapback). A hydroxyl group may also be provided by a serine, threonine or tyrosine residue of the pFOXC-RT itself (protein priming) or from the 3′ end of a cDNA or DNA molecule (DNA primer). Due to the phylogenetic relationship of pFOXC-RT to the pMaur-RT, cDNA synthesis may potentially initiate without a primer (de novo). **b** Addition of the telomere-like repeats. Following models for 3′-terminal repeat addition of linear DNAs that initiate via a protein primer (Salas 1991), the nascent pFOXC3 cDNA may slideback and reposition itself on the RNA template prior to being extended. As shown here, the pFOXC3-RT extends a snapped-back RNA and synthesizes a nascent cDNA of three nucleotides. The cDNA is displaced from the transcript and repositioned on the template prior to a new round of synthesis. This slideback–extension mechanism may occur more than once, generating multiple copies of the pentameric repeat, prior to elongation of the (–) strand cDNA (*dashed line*)

strand synthesis. Due to the ability of the pFOXC-RT to copy DNA templates, the second strand may be synthesized by the pFOXC-RT; however, it is also possible that a host DNA polymerase is utilized. Once DNA synthesis is complete, the RT may dissociate from the plasmid DNA, or it may associate with the 5′ end to potentially function as a protective cap.

3.2 Double Hairpin Linear Retroplasmids

The pRS224-1 plasmid of *Rhizoctonia solani* isolate H-16 is a 5 kb linear double hairpin retroplasmid (Fig. 1; Katsura et al. 2001; see also Chap. 9, in this volume). *Rhizoctonia solani* is a soil-born basidiomycete that is capa-

ble of infecting a broad range of plants. The dual terminal hairpin structure of pRS224-1 was derived from sequence analysis, computer-assisted folding programs, and electron microscopy, which showed that the 5 kb linear plasmid became a single-stranded circular molecule under denaturing conditions (Katsura et al. 2001). The plasmid is classified as a retroplasmid, based on the presence of a single large ORF that potentially encodes a polypeptide having the seven conserved amino acid motifs found among all RTs. Complete sequence analysis of pRS224-1 indicated that the ORF codes for an 887 amino acid polypeptide with a predicted mass of 102 kDa (Katsura et al. 2001; Nagasaka et al. 2003). It shows 19% identity (30% similarity) with the pMaur-RT and 16% identity (27% similarity) with the pFOXC3-RT within the highly conserved domain region (Galligan and Kennell, unpublished observations). The coding sequence for this ORF was expressed in *E. coli* and a purified fusion protein was used to generate a polyclonal antibody. The antibody recognizes a $\sim$ 90 kDa protein that is localized to the mitochondrion (Nagasaka et al. 2003).

The putative pRS224-1 RT is unusual in that it has a region of more than 200 amino acids between highly conserved domains III and IV that fail to match other RTs currently in the database (Katsura et al. 2001). Interestingly, in the cladogram shown in Fig. 2, pRS224 fails to group with either circular or linear retroplasmids, which suggests that it is highly divergent. Reverse transcriptase activity has not been demonstrated for the endogenous plasmid protein and little is known about the replication of pRS224-1. Two major plasmid transcripts have been identified, 4.7 kb and 7.4 kb in length. The smaller transcript begins near the center of the element and includes the RT-ORF and sequence of one of the terminal hairpins. The larger transcript encompasses the entire sense strand and both hairpins, and potentially serves as a template in plasmid replication (Katsura et al. 2001).

Several other linear retroplasmids have been reported based on partial sequence information (Mogen et al. 1991) or hybridization studies (Hirato et al. 1992), but the structures of the plasmid termini have not been determined. Conversely, another double hairpin linear DNA has also been reported in *Rhizoctonia solani* (pRS64-2; Katsura et al. 1997; see also Chap. 9, in this volume), yet it does not appear to encode a polymerase (the largest ORF is 163 amino acids), thus it is difficult to classify at this time.

4 Evolutionary Significance

Both circular and linear retroplasmids show little direct relationship to other retroelements. As mentioned, the retroplasmid RTs are deeply rooted in RT phylogeny and are possible progenitors to a wide range of retroelements (Wang and Lambowitz 1993a; Walther and Kennell 1999; Eickbush and Malik

2002). Evidence in support of their putative ancient origin also derives from the distinct enzymatic properties of their RTs, such as the mechanistic similarity of the pMaur-RT to RNA-dependent RNA polymerases. The finding that the pFOXC retroplasmids have 3′ terminal repeats similar to those of chromosomal telomeres and replicate via reverse transcription makes them leading contenders to be precursors of telomerase (Walther and Kennell 1999). In addition, retroplasmid clothespin DNAs contain a 5′-linked protein, suggesting a potential relationship to hepadnaviruses (e.g., hepatitis B virus) that have covalently attached terminal proteins that serve as primers for cDNA synthesis (reviewed in Ganem and Schneider 2001). These properties help to justify the position of retroplasmids at the base of retroelement phylogeny and the collective features of circular and linear retroplasmids suggest that they are related to primordial genetic elements that could have been a common ancestor of all known contemporary elements that replicate via reverse transcription. Further studies of this interesting group of genetic elements will likely shed more light on the origin of retroelements as well as events involved in the transition from the RNA world to the DNA/protein world.

It is clear that mitochondrial plasmids, and circular retroplasmids in particular, are able to interact directly with mitochondrial DNAs. Variant forms of Mauriceville and Varkud plasmids readily integrate into mtDNA and cause senescence (Akins et al. 1986, 1989; Chiang et al. 1994; D'Souza et al. 2005). Consequently, the *Neurospora* retroplasmids serve as excellent tools to understand nuclear–mitochondrial interactions as well as host mechanisms that developed to control or cope with invasive genetic elements. Interestingly, the degree and manner in which plasmids affect their hosts appears to be host-specific and the relative rates of senescence vary widely between the nearly identical Mauriceville and Varkud plasmids of *N. crassa* and *N. intermedia*, respectively (Fox and Kennell 2001). Variation in the behavior of plasmids and ways in which hosts respond to plasmid integration may partly explain why plasmid-induced senescence is not commonly detected outside of *Neurospora* spp.

Integration of the linear DNA plasmids has also been reported to linearize portions of mtDNAs of *Podospora anserina* (Hermanns et al. 1995; see also Chap. 8, in this volume), *N. intermedia* (Bertrand et al. 1985, 1986), *Zea mays* (Schardl et al. 1985) and *Physarum polycephalum* (Nomura et al. 2005). While there is currently no evidence that linear retroplasmids have integrated into mtDNAs, it cannot be discounted. Recent studies suggest that mtDNAs of most eukaryotes exist as various types of linear molecules (Nosek and Tomaska 2003). At least six different telomeric structures have been identified, several of which share features with the termini of linear plasmids. This has led to the hypothesis that mt plasmids played a major role in the development of linear mitochondrial genomes (Nosek and Tomaska 2003). Mitochondrial retroplasmids may not only have had a significant impact on mitochondrial genomes, but could have played a role in the development of

eukaryotic cells as well. In a recent hypothesis paper, Martin and Koonin (2006) proposed that the invasion of mobile group II introns from the α-proteobacterial ancestor of mitochondria into the presumed archaebacterial genome of the primitive eukaryote created a selective pressure for the development of the nuclear membrane. They speculate that the nuclear membrane evolved as a mechanism to separate transcription from translation to ensure that only fully spliced mRNAs are translated. Building on this idea, it is possible that mitochondrial retroplasmids have interacted with nuclear DNAs and, over time, have had considerable influence on nuclear chromosome structure (Nosek et al. 2006) and the development of mechanisms used to exclude the transmission of invasive mitochondrial genetic elements (Kennell, submitted for publication).

Acknowledgements We thank Casey Dillman, Matthew Althage, and Rahul Malireddy for assistance in the phylogenetic analysis.

References

Akins RA, Grant DM, Stohl LL, Bottorff DA, Nargang FE, Lambowitz AM (1988) Nucleotide sequence of the Varkud mitochondrial plasmid of Neurospora and synthesis of a hybrid transcript with a 5′ leader derived from mitochondrial RNA. J Mol Biol 204:1–25

Akins RA, Kelley RL, Lambowitz AM (1986) Mitochondrial plasmids of Neurospora: integration into mitochondrial DNA and evidence for reverse transcription in mitochondria. Cell 47:505–516

Akins RA, Kelley RL, Lambowitz AM (1989) Characterization of mutant mitochondrial plasmids of Neurospora spp. that have incorporated tRNAs by reverse transcription. Mol Cell Biol 9:678–691

Akins RA, Lambowitz AM (1990) Analysis of large deletions in the Mauriceville and Varkud mitochondrial plasmids of Neurospora. Curr Genet 18:365–369

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG (1998) The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature 396:133–140

Antal Z, Manczinger L, Kredics L, Kevei F, Nagy E (2002) Complete DNA sequence and analysis of a mitochondrial plasmid in the mycoparasitic Trichoderma harzianum strain T95. Plasmid 47:148–152

Arganoza MT, Akins RA (1995) Recombinant mitochondrial plasmids in Neurospora composed of Varkud and a new multimeric mitochondrial plasmid. Curr Genet 29: 34–43

Bertrand H, Baidyaroy D (2002) Hypovirulence. In: Osiewacz HD (ed) Molecular biology of development. Marcel Dekker, New York, pp 457–476

Bertrand H, Chan BS, Griffiths AJ (1985) Insertion of a foreign nucleotide sequence into mitochondrial DNA causes senescence in Neurospora intermedia. Cell 41:877–884

Bertrand H, Griffiths AJ, Court DA, Cheng CK (1986) An extrachromosomal plasmid is the etiological precursor of kalDNA insertion sequences in the mitochondrial chromosome of senescent Neurospora. Cell 47:829–837

Bibillo A, Eickbush TH (2002) The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. J Mol Biol 316:459–473

Bok JW, Griffiths AJ (2000) Possible benefits of kalilo plasmids to their Neurospora hosts. Plasmid 43:176–180

Boynton JE, Gillham NW (1996) Genetics and transformation of mitochondria in the green alga Chlamydomonas. Methods Enzymol 264:279–296

Chen B, Lambowitz AM (1997) De novo and DNA primer-mediated initiation of cDNA synthesis by the Mauriceville retroplasmid reverse transcriptase involve recognition of a 3′ CCA sequence. J Mol Biol 271:311–332

Chiang CC, Kennell JC, Wanner LA, Lambowitz AM (1994) A mitochondrial retroplasmid integrates into mitochondrial DNA by a novel mechanism involving the synthesis of a hybrid cDNA and homologous recombination. Mol Cell Biol 14:6419–6432

Chiang CC, Lambowitz AM (1997) The Mauriceville retroplasmid reverse transcriptase initiates cDNA synthesis de novo at the 3′ end of tRNAs. Mol Cell Biol 17:4526–4535

Collins RA, Saville BJ (1990) Independent transfer of mitochondrial chromosomes and plasmids during unstable vegetative fusion in Neurospora. Nature 345:177–179

Collins RA, Stohl LL, Cole MD, Lambowitz AM (1981) Characterization of a novel plasmid DNA found in mitochondria of N. crassa. Cell 24:443–452

D'Souza AD, Sultana S, Maheshwari R (2005) Characterization and prevalence of a circular mitochondrial plasmid in senescence-prone isolates of Neurospora intermedia. Curr Genet 47:182–193

Eickbush TH (1997) Telomerase and retrotransposons: which came first? Science 277:911–912

Eickbush TH, Malik HS (2002) Origins and Evolution of Retrotransposons. In: Craig NL (ed) Mobile DNA II. American Society of Microbiology, Washington DC, pp 1111–1144

Fox AN, Kennell JC (2001) Association between variant plasmid formation and senescence in retroplasmid-containing strains of Neurospora spp. Curr Genet 39:92–100

Ganem D, Schneider RJ (2001) Hepadnaviridae: The viruses and their replication. In: Fields BN, Knipe DM, Howley PM (eds) Virology, vol 2. Lippincot Williams and Wilkins, Philadelphia PA, pp 2923–2969

Griffiths AJ (1992) Fungal senescence. Annu Rev Genet 26:351–372

Griffiths AJ (1995) Natural plasmids of filamentous fungi. Microbiol Rev 59:673–685

Hermanns J, Asseburg A, Osiewacz HD (1995) Evidence for giant linear plasmids in the ascomycete Podospora anserina. Curr Genet 27:379–86

Hirato N, Hashiba T, Yoshida H, Kikumoto T, Ehara Y (1992) Detection and properties of plasmid-like DNA in isolates from twenty-three formae speciales of Fusarium oxysoporum. Ann Phytopath Soc Japan 58:386–392

Johnston SA, Anziano PQ, Shark K, Sanford JC, Butow RA (1988) Mitochondrial transformation in yeast by bombardment with microprojectiles. Science 240:1538–1541

Katsura K, Sasaki A, Nagasaka A, Fuji M, Miyake Y, Hashiba T (2001) Complete nucleotide sequence of the linear DNA plasmid pRS224 with hairpin loops from Rhizoctonia solani and its unique transcriptional form. Curr Genet 40:195–202

Katsura K, Suzuki F, Miyashita SI, Nishi T, Hirochika H, Hashiba T (1997) The complete nucleotide sequence and characterization of the linear DNA plasmid pRS64–2 from the plant pathogenic fungus Rhizoctonia solani. Curr Genet 32:431–435

Kempken F (1995) Plasmid DNA in Mycelial Fungi. In: Kuck U (ed) The Mycota II: Genetics and biotechnology. Springer, Berlin, Heidelberg, New York, pp 169–187

Kennell JC, Saville BJ, Mohr S, Kuiper MT, Sabourin JR, Collins RA, Lambowitz AM (1995) The VS catalytic RNA replicates by reverse transcription as a satellite of a retroplasmid. Genes Dev 9:294–303

Kennell JC, Wang H, Lambowitz AM (1994) The Mauriceville plasmid of Neurospora spp. uses novel mechanisms for initiating reverse transcription in vivo. Mol Cell Biol 14:3094–3107

Kistler HC, Benny U, Powell WA (1997) Linear mitochondrial plasmids of Fusarium oxysporum contain genes with sequence similarity to genes encoding a reverse transcriptase from Neurospora spp. Appl Environ Microbiol 63:3311–3313

Kistler HC, Leong SA (1986) Linear plasmidlike DNA in the plant pathogenic fungus Fusarium oxysporum f. sp. conglutinans. J Bacteriol 167:587–593

Kuiper MT, Lambowitz AM (1988) A novel reverse transcriptase activity associated with mitochondrial plasmids of Neurospora. Cell 55:693–704

Kuiper MT, Sabourin JR, Lambowitz AM (1990) Identification of the reverse transcriptase encoded by the Mauriceville and Varkud mitochondrial plasmids of Neurospora. J Biol Chem 265:6936–6943

Lambowitz AM, Chiang CC (1995) The Mauriceville and Varkud plasmids - primitive retroelements found in Neurospora mitochondria. Can J Botany - Revue Canadienne De Botanique 73:S173–S179

Leon P, Walbot V, Bedinger P (1989) Molecular analysis of the linear 2.3 kb plasmid of maize mitochondria: apparent capture of tRNA genes. Nucleic Acids Res 17:4089–4099

Maas MFPM, Hoekstra RF, Debets AJM (2007) Hybrid mitochondrial plasmids from senescence suppressor isolates of Neurospora intermedia. Genetics 175:785–794

Maizels N, Weiner AM (1993) The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication. In: Gesteland RF, Atkins JF (eds) The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, pp 425–444

Martin W, Koonin EV (2006) Introns and the origin of nucleus-cytosol compartmentalization. Nature 440:41–45

McClure MA (1993) Evolutionary history of reverse transcriptase. In: Shalka AM, Goff SP (eds) Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, pp 425–444

Meinhardt F, Schaffrath R, Larsen M (1997) Microbial linear plasmids. Appl Microbiol Biotechnol 47:329–336

Miller WA, Bujarski JJ, Dreher TW, Hall TC (1986) Minus-strand initiation by brome mosaic virus replicase within the 3′ tRNA-like structure of native and modified RNA templates. J Mol Biol 187:537–546

Mogen KL, Siegel MR, Schardl CL (1991) Linear DNA plasmids of the perennial ryegrass choke pathogen, Epichloe typhina (Clavicipitaceae). Curr Genet 20:519–526

Momol EA, Kistler HC (1992) Mitochondrial plasmids do not determine host range in crucifer-infecting strains of Fusarium oxysporum. Plant Pathol 41:103–112

Nagasaka A, Sasaki A, Sasaki T, Yonezawa M, Katsura K, Hashiba T (2003) Expression and localization of the linear DNA plasmid-encoded protein (RS224) in Rhizoctonia solani AG2–2. FEMS Microbiol Lett 225:41–46

Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR (1997) Telomerase catalytic subunit homologs from fission yeast and human. Science 277:955–959

Nargang FE, Bell JB, Stohl LL, Lambowitz AM (1984) The DNA sequence and genetic organization of a Neurospora mitochondrial plasmid suggest a relationship to introns and mobile elements. Cell 38:441–453

Nomura H, Moriyama Y, Kawano S (2005) Rearrangements in the Physarum polycephalum mitochondrial genome associated with a transition from linear mF-mtDNA recombinants to circular molecules. Curr Genet 47:100–110

Nosek J, Kosa P, Tomaska L (2006) On the origin of telomeres: a glimpse at the pre-telomerase world. Bioessays 28:182–190

Nosek J, Tomaska L (2003) Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. Curr Genet 44:73–84

Salas M (1991) Protein-priming of DNA replication. Annu Rev Biochem 60:39–71

Saville BJ, Collins RA (1990) A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. Cell 61:685–696

Saville BJ, Collins RA (1991) RNA-mediated ligation of self-cleavage products of a Neurospora mitochondrial plasmid transcript. Proc Natl Acad Sci USA 88:8826–8830

Schardl CL, Pring DR, Lonsdale DM (1985) Mitochondrial DNA rearrangements associated with fertile revertants of S-type male-sterile maize. Cell 43:361–368

Simpson EB, Ross SL, Marchetti SE, Kennell JC (2004) Relaxed primer specificity associated with reverse transcriptases encoded by the pFOXC retroplasmids of Fusarium oxysporum. Eukaryot Cell 3:1589–1600

Stevenson CB, Fox AN, Kennell JC (2000) Senescence associated with the over-replication of a mitochondrial retroplasmid in Neurospora crassa. Mol Gen Genet 263:433–444

Stohl LL, Collins RA, Cole MD, Lambowitz AM (1982) Characterization of two new plasmid DNAs found in mitochondria of wild-type Neurospora intermedia strains. Nucleic Acids Res 10:1439–1458

Walther TC, Kennell JC (1999) Linear mitochondrial plasmids of F. oxysporum are novel, telomere-like retroelements. Mol Cell 4:229–238

Wang H, Blackburn EH (1997) De novo telomere addition by Tetrahymena telomerase in vitro. EMBO J 16:866–879

Wang H, Gilley D, Blackburn EH (1998) A novel specificity for the primer-template pairing requirement in Tetrahymena telomerase. EMBO J 17:1152–1160

Wang H, Kennell JC, Kuiper MT, Sabourin JR, Saldanha R, Lambowitz AM (1992) The Mauriceville plasmid of Neurospora crassa: characterization of a novel reverse transcriptase that begins cDNA synthesis at the 3′ end of template RNA. Mol Cell Biol 12:5131–5144

Wang H, Lambowitz AM (1993a) The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis de novo and may be related to reverse transcriptase and DNA polymerase progenitor. Cell 75:1071–1081

Wang H, Lambowitz AM (1993b) Reverse transcription of the Mauriceville plasmid of Neurospora. Lack of ribonuclease H activity associated with the reverse transcriptase and possible use of mitochondrial ribonuclease H. J Biol Chem 268:18951–18959

Ware TL, Wang H, Blackburn EH (2000) Three telomerases with completely non-telomeric template replacements are catalytically active. EMBO J 19:3119–3131

Wickner RB, Esteban R, Hillman BI (2000) Narnavirdae. In: Van Regenmortal MHV, Fauquet C, Bishop D, Carsten E, Estes M, Lemon SM, Maniloff J, Mayo M, McGeoch D, Pringle C, Wickner R (eds) Virus taxonomy. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, New York, pp 651–653

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362

Zamel R, Poon A, Jaikaran D, Andersen A, Olive J, De Abreu D, Collins RA (2004) Exceptionally fast self-cleavage by a Neurospora Varkud satellite ribozyme. Proc Natl Acad Sci USA 101:1467–1472

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_095/Published online: 12 May 2007

Linear Protein-Primed Replicating Plasmids in Eukaryotic Microbes

Roland Klassen · Friedhelm Meinhardt (✉)

Institut für Molekulare Mikrobiologie und Biotechnologie,
Westfälische Wilhelms-Universität Münster, Corrensstr. 3, 48149 Münster, Germany
meinhar@uni-muenster.de

1 Introduction . . . 188

2 Common Structural Features . . . 188

3 Linear Plasmids in Filamentous Fungi . . . 190
3.1 The DNA Polymerase . . . 192
3.2 The RNA Polymerase . . . 193
3.3 Additional ORFs . . . 194
3.4 Phenotypes Associated With Fungal Linear Plasmids . . . 194

4 Linear Plasmids in Yeasts . . . 196
4.1 Autonomous Elements . . . 196
4.1.1 DNA Polymerase, SSB, and Terminal Recognition Factor . . . 198
4.1.2 The Cytoplasmic Trancriptase Complex . . . 200
4.1.3 Cytoplasmic Transcription . . . 201
4.2 The Nonautonomous Elements . . . 202
4.2.1 Killer Plasmids . . . 203
4.2.2 Cryptic Elements . . . 213

5 Evolution of Linear Plasmids in Eukaryotes . . . 213

6 Concluding Remarks and Perspectives . . . 215

References . . . 216

Abstract Linear plasmids of eukaryotic microbes are contemporary manifestations of ancient viruses, which have adjusted to two cellular compartments during evolution, i.e., to the mitochondrium or to the cytoplasm. In either case, infectious viral functions do not (any longer) exist. Mitochondrial as well as cytoplasmic elements display a minimized gene equipment and an archetypical mode of replication. Plasmids of filamentous fungi are selfish DNA elements routinely residing in the mitochondria, in which they underwent coevolution with their hosts. In addition to the archetypical viral B-type DNA polymerase, they typically exclusively encode a viral RNA polymerase. Usually, there are neither positive nor negative impacts on their hosts. In a few instances, however, symptoms redolent of a molecular disease, such as the accumulation of defective mitochondria and early onset of senescence, manifest in plasmid-harboring strains. Cytoplasmic localization, which applies for almost all yeast linear plasmids known so far, evidently enforced a more complex enzyme repertoire (of viral origin) to accomplish autonomous extranuclear and extramitochondrial replication and transcription, such as a helicase,

ssDNA binding proteins, and a capping enzyme. Accompanying cytoplasmic plasmids relying functionally on an autonomous element are rather frequent; some encode protein toxins, which benefits the respective host while competing with other yeasts (killer phenotype). Such toxins assure autoselection of the plasmid system as well. Two distinct toxic principles are known up to the present: one was shown to be a tRNase, whereas the other clearly involves a DNA-damaging mode of action.

1 Introduction

Linear protein-primed replicating plasmids frequently occur in two kingdoms of eukarya, i.e., in planta and fungi. The S plasmids from *Zea mays* were the first discovered linear double-stranded DNA (dsDNA) elements (Pring et al. 1977); soon after, similar genetic traits were detected in the yeast *Kluyveromyces lactis* and in the filamentous ascomycete *Ascobolus immersus* (Gunge et al. 1981; Francou 1981). Though linear plasmids were initially considered to represent rare exceptions from the commonly accepted rule defining plasmids as extrachromosomal covalently closed circular replicons, a great and still growing number of linear elements have since been discovered. To date, they are recognized as rather frequently occurring, extrachromosomal genetic elements with growing evidence for a viral evolutionary origin. Here, we focus on protein-primed replicating linear plasmids in eukaryotic microorganisms, i.e., filamentous fungi and yeasts. Reviews addressing linear plasmids from filamentous fungi and yeasts separately have been published over the years (Meinhardt et al. 1990; Stark et al. 1990; Griffiths 1995; Gunge 1995; Meinhardt and Schaffrath 2001; Gunge and Tokunaga 2004; Jeske et al. 2006a). However, recently established techniques, such as allelic replacement, gene shuffling, site-specific mutagenesis, and—more importantly—the extended molecular characterization of a number of linear plasmids to the sequence level, have greatly increased our knowledge at the molecular genetic level.

2 Common Structural Features

Regardless of the cell type harboring them, linear replicons encounter the problem of avoiding shortening of the termini during each replication round. In eukaryotic chromosomes the telomerase, a ribonucleoprotein with reverse transcriptase activity, adds telomeric repeats copied from an internal RNA template onto the ends of the chromosomes (Greider and Blackburn 1985, 1989). Another strategy to maintain size is brought about by covalently closed DNA ends, as in some prokaryotic and eukaryotic linear DNA elements (see Chapters in this volume by Kobryn (2007), Hertwig (2007), Galligan and Kennell (2007), Hashiba and Nagasaka (2007)).

The majority of fungal and yeast linear plasmids employ a replication mode using protein priming, thereby facilitating full replication of termini. In addition to the protein primers, which remain covalently bound (terminal proteins, TPs) to the 5′ ends of the newly synthesized DNA molecule, all these elements have in common terminal inverted repeat (TIR) sequences resembling both linear genomes of bacteriophages and adenoviruses, suggesting a common mode of replication (Paillard et al. 1985; Kempken et al. 1989; Meinhardt et al. 1990). Indeed, the protein-primed replication mechanism is well studied in adenoviruses and bacteriophages, serving as a model for linear plasmid replication (reviewed by Challberg and Kelly 1982; Salas 1991; Meijer et al. 2001). For *Bacillus* phage phi29, a serine residue of the TP acts as the replication primer by facilitating desoxynucleotidylation of the free hydroxyl group. Along with the TP, the phage-encoded DNA polymerase forms an initiation complex at the second nucleotide of the TIR (Kamtekar et al. 2006). Subsequently, the TP-DNA polymerase complex slides back to the first nucleotide, starting elongation of replication and dissociation of the TP–polymerase complex (Mendez et al. 1992, 1997; Kamtekar et al. 2006). There are no internal origins of replication and the highly processive DNA polymerase is able to replicate the entire phi29 genome (Kamtekar et al. 2004). Though replication initiation may occur simultaneously at both ends, yielding two fully replicated copies, single-strand (ss) replication intermediates occur as well (Fig. 1). By hybridization of the inversely oriented TIR sequences such ssDNAs form so-called panhandle structures, providing a dsDNA end suitable for initiation of a novel replication cycle, eventually resulting in a fully replicated dsDNA copy of the genome (Fig. 1; Challberg and Kelly 1992).

Such a replication mechanism realized in adenoviruses, bacteriophages, and linear plasmids led to the suggestion that not only do the bacteriophages and adenoviruses share a common ancestor, but also linear plasmids may have evolved from an archetypical protein-primed replicating genome some billion years ago (Davison et al. 2003). Interestingly, in both adenoviruses and bacteriophages, the replication module consists of the DNA polymerase and a separate TP encoding gene, which are, however, directly neighbored; in all linear plasmids from fungi and yeasts, a single open reading frame (ORF) encodes both the TP and the DNA polymerase (Meinhardt and Rohe 1993).

Thus, in addition to linearity, and the presence of TPs and TIRs, such a (fusion) TP-DNA polymerase gene represents a common feature likely to be found in all eukaryotic linear plasmids employing protein-primed replication. However, linear plasmids often exist as pairs or triplets, and there are several instances in which only one of the respective plasmids encodes a TP-DNA polymerase, which probably manages replication of the other linear plasmids (lacking such a gene) as well (see below; Kempken et al. 1989; Rohe et al. 1991; Klassen et al. 2001, 2002). Eukaryotic linear plasmids generally also encode their own RNA polymerases for transcription of plasmid-based

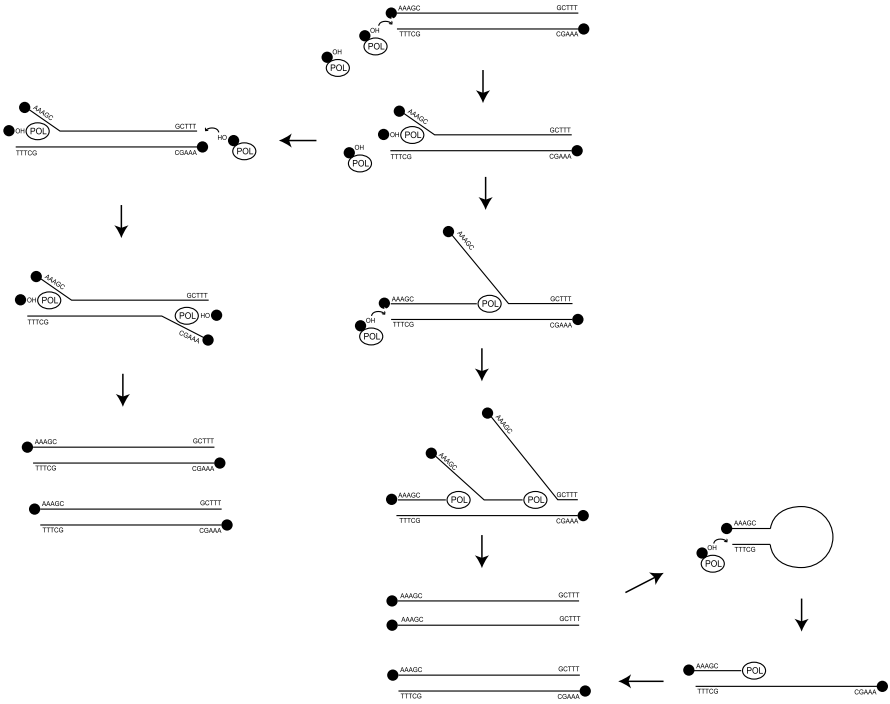


Fig. 1 Protein-primed strand displacement replication. The linear element is schematically represented by *two lines* with TIR sequences (AAAGC/GCTTT). *Filled circles* indicate plasmid-bound TPs; POL represents the TP-DNA polymerase fusion protein; the hydroxyl moiety (OH) of an amino acid acts as the primer

genes (see below), providing another clue for viral ancestry. As for the DNA polymerase, the RNA polymerase encoding gene is not necessarily found on each plasmid but is present at least once in each plasmid system (Kempken et al. 1992).

3 Linear Plasmids in Filamentous Fungi

Among the filamentous fungi, a great number of plasmids have been described for both ascomycetes and basidiomycetes; almost all of them reside in mitochondria (reviewed by Griffiths 1995). The only known exceptions are the linear plasmids from *Alternaria alternata*, which are apparently not localized in mitochondria (Shepherd 1992). Though there are circular plasmids, such as Mauriceville, Fiji, and LaBelle plasmids in *Neurospora* (Griffiths 1995), the great majority are linear molecules with TPs and TIRs. Frequencies of linear plasmids have been determined for a number of fungi belonging to

different genera, such as *Neurospora*, *Claviceps*, *Fusarium*, *Blumeria*, *Epichloë*, and *Podospora* (Arganoza et al. 1994; Samac and Leong 1988; Morgen et al. 1991; van der Gaag et al. 1998; Giese et al. 1990; Tudzynski et al. 1983; Tudzynski and Esser 1986; Düvell et al. 1988). Overall frequencies were estimated to match approximately 8–16% (van der Gaag et al. 1998); however, when the geographic distribution was taken into consideration, it became evident that linear plasmids may be endemic to specific geographic regions, in which they represent the norm rather than an exception. The kalilo plasmid may serve

Table 1 Mitochondrial linear plasmids of filamentous fungi for which complete nucleotide sequences are available (GenBank accession numbers in parentheses). Sizes of plasmids and TIRs are indicated. Coding capacity specifies genes identified; accession numbers for deduced polypeptides are indicated along with their sizes in amino acids (aa). DP, RP: DNA polymerase and RNA polymerase genes; unknown: gene of unknown function

Species	Plasmid (acc. no.)	Size/ TIRs (bp)	Coding capacity (size, acc. no.)	Refs.
Pleurotus ostreatus	pMLP1 (AF126285)	9879 381	unknown (239 aa; AAD39925) RP (903 aa; AAD39926) DP (1346 aa; AAD39927)	Kim et al. (2000)
Neurospora crassa	maranhar (X55361)	7052 349	RP (896 aa; CAA39045) DP (1021 aa; CAA39046)	Court and Bertrand (1992)
Neurospora intermedia	harbin-3 (NC_000843)	7050 350	RP (896 aa; AAD31445) DP (1035 aa; AAD31446)	Xu et al. (1999)
	kalilo (X52106)	8642 1365	RP (811 aa; CAA36326) DP (893 aa; CAA36327)	Chan et al. (1991b)
Gelasinospora sp.	Gel-kal (L40494)	8231 1137	RP (987 aa; AAB41447) DP (831 aa; AAB41448)	Yuewang et al. (1996)
Blumeria graminis	pBgh (AY189817)	7965 610	RP (973 aa; AAO37818) DP (1062 aa; AAO37819)	Giese et al. (2003)
Morchella conica	pMC3-2 (X63909)	6044 720	unknown (306 aa; CAB52198) DP (901 aa; CAA45364)	Rohe et al. (1991)
Podospora anserina	pAL2-1 (X60707)	8395 975	RP (948 aa; CAA43116) DP (1197 aa; CAA43117)	Hermanns and Osiewacz (1992)
Claviceps purpurea	pClKl (X15648)	6752 327	RP (970 aa; Nt. 112-3021) DP (1097 aa; Nt. 6641-3351) 4 minor ORFs	Oeser and Tudzynski (1989)
	pClT5 (X68490)	7113 574	RP (947 aa; Nt. 356-3199) DP (1050 aa; Nt. 6758-3606) 5 minor ORFs	Oeser et al. (1993)
Ascobolus immersus	pAI1 (X15982)	5142 535	DP (1202 aa; CAA34106) 4 minor ORFs	Kempken et al. (1989)

as an example, as it was found routinely in isolates from Hawaii (Yang and Griffiths 1993; Maas et al. 2005), but less frequently in Asia, Africa, Central America, and South Pacific islands (Arganoza et al. 1994). In *Claviceps purpurea* and *Blumeria graminis*, linear plasmids occur frequently irrespective of their geographic origins (Tudzynski and Esser 1986; Düvell et al. 1988; Giese et al. 1990, 2003), being indicative of a widespread distribution in these fungi.

Partial or entire nucleotide sequences are available for a considerable number of linear plasmids from both filamentous ascomycetes and basidiomycetes (Table 1). Sizes vary from 5142 bp (pAI1 of *Ascobolus immersus*) to 9879 bp (pMLP1 of *Pleurotus ostreatus*); all of them possess TIRs ranging in size from 327 to 1365 bp (Table 1).

Most known linear plasmids from filamentous fungi display a rather common genetic organization as for the maranhar plasmid from *Neurospora crassa* (Fig. 2), which contains two ORFs in an inverse orientation encoding the TP-DNA polymerase and the RNA polymerase, respectively. However, differently organized plasmids exist as well, such as pAI1 of *A. immersus* and pMC3-2 of the morel, which encode the typical TP-DNA polymerase fusion gene but lack an RNA polymerase gene (Kempken et al. 1989; Rohe et al. 1991; Fig. 2). In both instances the plasmids concomitantly exist along with other elements, suggesting the RNA polymerase to be encoded by an accompanying plasmid.

Such a scenario has indeed been proven for the linear plasmids discovered first, i.e., the S plasmids S1 and S2 from maize (Levings and Sederoff 1983; Paillard et al. 1985); S1 encodes the TP and DNA polymerase, while the RNA polymerase gene is located on S2.

3.1 The DNA Polymerase

DNA polymerases encoded by fungal linear plasmids are of the viral B type, like enzymes from adenoviruses and phages which display a replication mode employing protein primers (Ito and Braithwaite 1991; Kempken et al. 1992; Rohe et al. 1992). As already mentioned, however, linear plasmid encoded DNA polymerases carry the TPs as an N-terminal extension. Though suggested earlier (Oeser and Tudzynski 1989; Chan et al. 1991; Meinhardt and Rohe 1993), this structure was experimentally proven rather late for pGKL2 of *Kluyveromyces lactis* and later for pMLP1 of *Pleurotus ostreatus* (Takeda et al. 1996; Kim et al. 2000). Except for a weakly conserved motif (SYKN), there is only, if at all, a hardly detectable similarity among TP regions of distantly related plasmid DNA polymerases and the respective proteins from bacteriophages and adenoviruses, suggesting extended structural diversities. It remains totally obscure at present if, how, and where the TP is split off from the polymerase during replication. For the *Neurospora* plasmid kalilo, the molecular mass of the TP was determined to be approximately 120 kDa

(Vierula et al. 1990). It was suggested that it actually may represent the full-size TP-DNA polymerase fusion protein (Chan et al. 1991). On the contrary, however, for pMLP1 of *Pleurotus ostreatus*, as well as another linear plasmid, pMLP2, it was shown that TPs comprise only 70 and 72 kDa, respectively; consistently, DNA polymerase domains were immunologically not detectable, providing evidence for in vivo processing of the protein (Kim et al. 2000). Also, the molecular mass (approx. 36 kDa) of the TP of pGKL2 (Stam et al. 1986; Takeda et al. 1996) rules out a full-size TP-DNA protein. Replication initiation has been studied in detail for phage phi29; it was proven that the initiation complex consisting of TP and DNA polymerase starts DNA synthesis at the ends of the template. Dissociation of the above complex occurs shortly after elongation has begun, i.e., after the synthesis of approximately six nucleotides (Mendez et al. 1997; Kamtekar et al. 2006).

The highly processive phi29 DNA polymerase replicates the entire phage genome unassisted by a helicase or sliding clamp. Such unique features rely on the terminal protein region 2 (TPR2), which is similarly inserted in all DNA polymerases using protein primers. The TPR2 domain forms a doughnutlike structure, which enables high processivity and facilitates separation of template and nontemplate strands before entry into the polymerase active site (Kamtekar et al. 2004; Rodriguez et al. 2005). TPR2-like domains in linear plasmid encoded DNA polymerases (Dufour et al. 2000, 2003; Rodriguez et al. 2005) are assumed to compare functionally to the phi29 enzyme (Jeske et al. 2006a).

3.2 The RNA Polymerase

RNA polymerases encoded by mitochondrial linear plasmids from both plants and fungi are of the bacteriophage T7 type (Kuzmin et al. 1988; Oeser 1988; Oeser and Tudzynski 1989; Kempken et al. 1992; Cermakian et al. 1997). The latter does not employ a protein-primed replication mode; indeed, it is only distantly related to protein-primed replicating bacteriophages. The T7 RNA polymerase constitutes an archetypical single subunit enzyme which is not only encoded by T7-like phages and linear plasmids, but also is the common RNA polymerase in (plasmid free) mitochondria throughout the eukarya (Masters et al. 1987; Tiranti et al. 1997; Weihe et al. 1997; Cermakian et al. 1996). It has been suggested that the mitochondrial RNA polymerase, acquired from a T7 progenitor early in eukaryotic evolution, replaced the eubacterial multisubunit RNA polymerase that was originally present in the bacterial progenitor of the mitochondrium (Shutt and Gray 2006). Thus, while the DNA polymerase encoded by mitochondrial linear plasmids is clearly of viral origin (similar to adenoviral and phi29-like enzymes), it remains obscure whether the RNA polymerase encoding gene was acquired directly from a viral progenitor or from the eukaryotic host. Interestingly, phi29-like phages

do not possess their own RNA polymerases but recruit host encoded enzymes (reviewed by Meijer et al. 2001). It remains to be elucidated why T7-like RNA polymerases are required for maintenance of linear plasmids residing in mitochondria, although there is a T7-like polymerase anyway.

Though T7-like RNA polymerase genes are routinely found in known mitochondrial linear plasmids, it is not to be excluded that plasmid systems devoid of such genes may exist. Mitochondrial linear plasmids are known which encode a TP-DNA polymerase only (e.g., pAI1 and pMC2-3); however, sequencing of the entire set of plasmids present in such strains is necessary to resolve whether there is a T7-like RNA polymerase gene.

3.3
Additional ORFs

ORFs encoding predicted polypeptides differing articulately from DNA and/or RNA polymerases exist in a number of mitochondrial linear plasmids. Rather frequently, such reading frames overlap the DNA or RNA polymerase genes, as for pClK1, pClT5, and pAL-1 (Oeser and Tudzynski 1989; Oeser et al. 1993; Kempken et al. 1989).

In pMLP1 of *Pleurotus ostreatus* and pMC3-2 of *Morchella conica* large ORFs are located upstream of the DNA or RNA polymerase encoding genes (Rohe et al. 1991; Kim et al. 2000; Fig. 2). As the predicted protein of the respective pMLP1 ORF is remarkably basic, it was presumed to function as a TIR binding protein, possibly involved in replication initiation (Kim et al. 2000). However, such information is not conserved among mitochondrial linear plasmids which rules out its general requirement. A function is not known or attributable to any of the known additional reading frames; there is no obvious similarity either to functionally assigned genes or predicted polypeptides, or to each other. Hence, mitochondrial linear plasmids generally encode exclusively TPs, DNA polymerase, and RNA polymerase, and seldomly nonessential other proteins.

3.4
Phenotypes Associated With Fungal Linear Plasmids

Linear plasmids were detected in a number of plant pathogenic filamentous fungi; however, they apparently do not contribute to a pathogen's detrimental properties but rather persist as neutral passengers (Griffiths 1995).

In *Neurospora*, insertion of the linear plasmids kalilo and maranhar into the host's mitochondrial DNA routinely results in onset of senescence due to mitochondrial malfunction, ultimately leading to death (reviewed by Griffiths 1992; Bertrand 2000). Kalilo and maranhar insertion causes large inverted repetitions of mitochondrial DNA adjacent to the insertion site (Bertrand et al. 1985, 1986; Court et al. 1991). As a specific (defective) mitochondrial

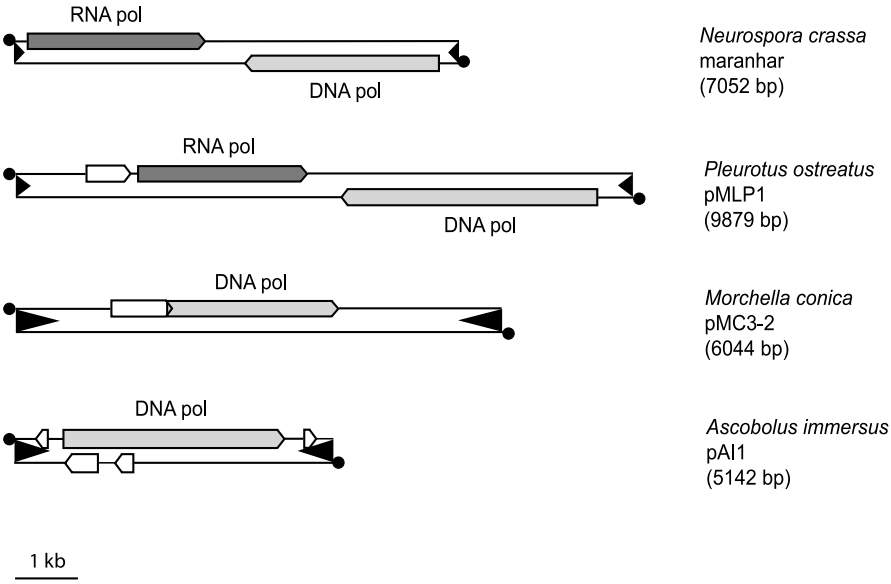


Fig. 2 Schematic representation of the structure and coding capacity of mitochondrial linear plasmids of filamentous fungi (a selection). *Arrows* indicate ORFs and their transcriptional direction; TPs are depicted as *filled circles*; TIRs correspond to *filled triangles*

DNA variant with inserted plasmids (IS-kalDNA or IS-marDNA) dominates in a particular senescent culture, it has apparently originated from a single insertion event (Myers et al. 1989). Supposedly, functionally impaired IS-kalDNA or IS-marDNA carrying mitochondria trigger mitochondrial replication to compensate for the defect and thus gradually replace wild-type mitochondria dividing at normal rates (Myers et al. 1989; Griffiths 1995).

Integration of a linear plasmid (pAL2-1) into the mitochondrial DNA has also been reported for *Podospora anserina* (AL2); however, in marked contrast to *Neurospora*, insertion of pAL2-1 resulted in a prolonged life span of the regularly aging fungus (Hermanns et al. 1994, 1995). Since other isolates of *P. anserina* carrying pAL2-1 homologous elements senesce normally (van der Gaag et al. 1998), such a phenomenon was apparently strain specific. Moreover, when *P. anserina* served as a model for studying the lifetime prolonging effect of calorie restriction (Maas et al. 2004), it was disclosed that life span (under calorie restriction) is indeed shortened by pAL2-1; hence, as for kalilo and maranhar, pAL2-1 was considered to induce an early onset of senescence (Maas et al. 2004, 2005).

In summary and by way of conclusion, most fungal linear plasmids appear to be cryptic elements, and occasionally they are associated with a molecular disease leading to massive accumulation of defective mitochondria which is—by the way—accompanied by a vast propagation of the inserted plasmid

copy. Thus, fungal linear plasmids represent selfish or parasitic DNAs with no discernible positive effects on their hosts.

4
Linear Plasmids in Yeasts

Yeast linear plasmids, originally detected in *K. lactis* (Gunge et al. 1981), are known to occur in a panoply of ascomycetous species belonging to different genera (such as *Pichia*, *Candida*, *Debaryomyces*, *Saccharomycopsis*, *Schwanniomyces*, *Botryoascus*) and also in the basidiomycetous representative *Trichosporon* (see Table 2; Kitada and Hishinuma 1987; Ligon et al. 1989; Worsham and Bolen 1990; Hayman and Bolen 1991; Bolen et al. 1992; Cong et al. 1994; Fukuhara 1995; Chen et al. 2000). A systematic screening among isolates deposited in the CBS-type culture collection revealed linear plasmids in 1.8% of the strains analyzed (Fukuhara 1995). At first sight—compared to filamentous fungi—few specimens seemed to harbor such genetic elements; however, plasmid occurrence in isolates of the same species is remarkably frequent, e.g., 16% for the pGKL system in *K. lactis* and 54% for pTP1 in *Trichosporon pullulans* (see also Table 2). The latter is so far the only basidiomycetous yeast species known to harbor linear plasmids and it remains to be elucidated whether this is an exception.

A striking difference between linear plasmids of filamentous fungi and yeasts concerns their localization (Table 2; Gunge et al. 1982; Ligon et al. 1989; Cong et al. 1994; Fukuhara 1995). There are only two mitochondrial yeast linear plasmids known (pPH1 and pPK1 from *Pichia heedi* and *P. kluyveri*; Blaissoneau et al. 1999), whereas almost all elements in filamentous fungi are mitochondrially associated (see above; Griffiths 1995). Based on the sequence data available for a number of cytoplasmic yeast linear plasmids and the mitochondrial pPH1, three types can be distinguished (Fig. 2): (1) mitochondrial elements, which resemble linear plasmids of filamentous fungi—indeed, genes encoding TP-DNA and RNA polymerase of pPH1 are arranged as for maranhar; (2) cytoplasmic autonomous elements, which are relatively large linear plasmids spanning at least 12 kb; in addition to TP-DNA and RNA polymerase other functions, apparently required for cytoplasmic inheritance, are encoded; and (3) cytoplasmic nonautonomous elements; smaller linear plasmids (i.e., < 12 kb), depending strictly on an aforementioned autonomous element; nonessential functions, such as a killer protein may be encoded.

4.1
Autonomous Elements

Extranuclear linear plasmids residing in the cytoplasm constitute the prevailing version of accessory genetic elements in yeast; routinely, they exist

Table 2 Yeast linear plasmids. Mitochondrial (pPH1, pPK1) and cytoplasmic linear plasmids of ascomycetes and basidiomycetes (pTP1) are given along with their sizes (in kb) and coding capacity. DP: DNA polymerase; RP: RNA polymerase; T: toxin; I: immunity; CB: chitin binding protein; CE: capping enzyme; SSB: single-strand binding protein; TRF: terminal recognition factor. Plasmid presence is derived from data accessible from the CBS (http://www.cbs.knaw.nl/) and/or original publications on the respective plasmid system (see also Cong et al. 1994; Fukuhara 1995)

Species	Plasmid system (kb)	Coding capacity	No. of strains analyzed (plasmids present)
Ascomycetes			
Pichia heedi	pPH1 (7.1)	DP/RP	3 (2)
Pichia kluyveri	pPK1 (7.2)	DP/RP	16 (3)
Kluyveromyces lactis	pGKL1 (8.9)	DP/T/I	31 (5)
	pGKL2 (13.5)	DP/RP/CE/SSB/TRF	
Debaryomyces hansenii	pDH1A (8.2)	DP/CB[a]	25 (1)
	pDH1B (14.4)	DP/RP[b]	
Debaryomyces hansenii TK	pDHL1 (8.4)	DP/CB[c]	1 (1)
	pDHL2 (9.2)		
	pDHL3 (15)		
Pichia acaciae	pPac1-2 (6.8)	T/I	2 (1)
	pPac1-1 (13.6)	DP/RP/CE/SSB/TRF	
Debaryomyces robertsiae	pWR1A (8.1)	T/I	2 (1)
	pWR1B (14.6)	DP/RP/CE/SSB/TRF	
Pichia pastoris	pPP1A (9.5)		4 (2)
	pPP1B (13)		
Candida tartarivorans	pCX1A (8)		1 (1)
	pCX1B (13)		
Pichia nakazawae	pPN1A (7)		2 (1)
	pPN1B (8)		
	pPN1C (15)		
Pichia inositovora	pPin1-3 (9.7)	DP/T	1 (1)
	pPin1-2 (13)		
	pPin1-1 (18)		
Debaryomyces etchellsii	pPE1A (6.7)	CB	4 (3)
	pPE1B (12.8)	DP/RP/CE/SSB/TRF	
Schwanniomyces occidentalis	pSoc1-1 (8.1)		2 (1)
	pSoc1-2 (13.4)		
Saccharomyces kluyveri	pSKL (14)	DP/RP/CE/SSB/TRF	5 (1)
Saccharomycopsis malanga	pSM2A (8.7)		2 (1)
	pSM2B (15)		
Saccharomycopsis crataegensis	pScr1-3 (5)		3 (2)
	pScr1-2 (7)		
	pScr1-1 (15)		
Botryoascus cladosporoides	pBC1A (7)		1 (1)
	pBC1B (14)		
Debaryomyces polymorphus	pDP1 (18)		1 (1)
Basidiomycetes			
Trichosporon pullulans	pTP1 (16)		13 (7)

[a] based on heterologous hybridization with pGKL1 derived probes
[b] based on heterologous hybridization with pGKL2 derived probes
[c] based on partial DNA sequences

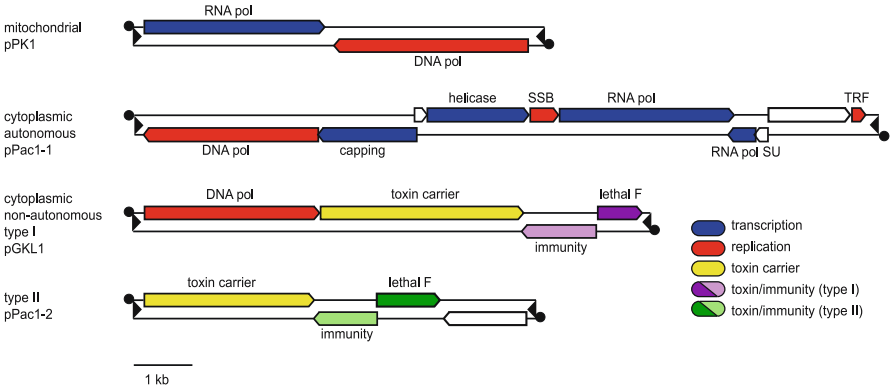


Fig. 3 Schematic representation of yeast linear plasmids: mitochondrial, cytoplasmic autonomous/nonautonomous, and toxin type I/II encoding. Symbols as in Fig. 2. Known or assumed functions of predicted proteins are given. Colors correspond to functional categories as indicated

as pairs or triplets, one of which is invariably self-sufficient with respect to replication and maintenance. Only in a few cases (e.g., *Saccharomyces kluyveri* pSKL, *T. pullulans* pTP1, and *Debaryomyces polymorphus* pDP1), was an autonomous element discovered without any attendance. Sequencing autonomous linear plasmids from *K. lactis*, *S. kluyveri*, *D. etchellsii*, and *P. acaciae* (Hishinuma et al. 1984; Tommasino et al. 1988; Hishinuma and Hirai 1991; Larsen and Meinhardt 2000; Klassen et al. 2001; Jeske and Meinhardt 2006) revealed a strictly conserved genetic organization, irrespective of the taxonomic or geographic origin. Exemplified for pPac1-1 in Fig. 3, such an illustration in principle applies for any of the autonomous cytoplasmic plasmids sequenced thus far. pGKL2 and pSKL from *K. lactis* and *S. kluyveri* negligibly deviate as they possess an additional nonessential gene of unknown function (ORF1). Since replication and transcription is confined genetically to the nucleus, autonomous cytoplasmic elements must necessarily provide the enzymatic repertoire ensuring both replication and transcription. With the exception of the above mentioned ORF1 in pGKL2 and pSKL, nonessential additional functions, such as the killer phenotype, are invariably encoded by an accompanying smaller linear plasmid, which, however, strictly depends on an autonomous element in terms of cytoplasmic gene expression and maintenance.

4.1.1 DNA Polymerase, SSB, and Terminal Recognition Factor

Each autonomous linear plasmid harbors a gene encoding a TP-DNA polymerase (Tommasino et al. 1988; Hishinuma and Hirai 1991; Hishinuma et al.

1984; Stark et al. 1984; Sor and Fukuhara 1985; Klassen et al. 2001; Klassen and Meinhardt 2003; Jeske and Meinhardt 2006). Hence, any cytoplasmic plasmid system possesses at least one viral B-type DNA polymerase; however, as for the pGKL system of *K. lactis*, another TP-DNA polymerase locus may be present. Since loss or knockout of either of the two polymerase encoding genes cannot be functionally complemented by the remaining locus, each plasmid evidently replicates via its own DNA polymerase (Kitada and Gunge 1988; Schaffrath et al. 1995). Such specificity is supported by the fact that TPs of pGKL1 and pGKL2 differ in size. Indeed, it was experimentally proven that TPs of pGKL2 are encoded as the N-terminal region of pGKL2 Orf2 (Takeda et al. 1986), agreeing with DNA polymerases specifically replicating the plasmid that encodes them.

In multiple plasmid systems possessing a single TP-DNA polymerase gene only (on the autonomous element), as for pPac1-1, pPac1-2 and pPE1A, pPE1B of *P. acaciae* and *D. etchellsii*, respectively, replication of nonautonomous elements is probably mediated by the same enzyme. Though it is to be expected that autonomous and nonautonomous elements should have identical TPs in these instances, there is no discernible sequence homology with respect to TIR sequences (Klassen et al. 2002, 2004; Jeske and Meinhardt 2006), which are considered to harbor replication origins. Thus, the TPs rather than the TIRs provide the basis for the specificity seen in systems with more than one DNA polymerase.

The viral B-type DNA polymerase–TP fusion protein is apparently sufficient for replication of linear plasmids in mitochondria. However, cytoplasmic replication needs additional plasmid encoded functions, such as the single-strand DNA binding protein (SSB), for stabilization of replication intermediates and the (rather basic) terminal recognition factor (TRF) which specifically binds to TIRs and is presumably involved in replication initiation (McNeel and Tamanoi 1991; Tommasino 1991; Schaffrath and Meacock 2001). Indeed, pGKL2 encoded SSB were shown to interact with ssDNA without sequence preference (Schaffrath and Meacock 2001), whereas binding sites for TRF comprise nucleotides 107–183 in the pGKL1 TIR and 126–179 in the pGKL2 TIR. Though there is no obvious sequence similarity, both TIR regions are rich in long dA–dT stretches presumably folding into similar secondary structures (McNeel and Tamanoi 1991).

Genes encoding TP-DNA polymerase, SSB, and TRF were proven to be essential for cytoplasmic inheritance in the pGKL1,2 system (Schaffrath and Meacock 1995; Schaffrath et al. 1995; Tiggemann and Meinhardt, unpublished results). The enzymatic machinery involved in replication of cytoplasmic linear plasmids clearly resembles the scenario seen in phi29-like bacteriophages, since virus encoded ss- and dsDNA binding proteins are involved in replication initiation in addition to the B-type DNA polymerase and TPs (Meijer et al. 2001).

4.1.2 The Cytoplasmic Trancriptase Complex

A rather unique RNA polymerase, a DExH/D box helicase and an mRNA capping enzyme, represent key elements of the cytoplasmic transcriptional apparatus. Consistent with their pivotal function, each of the above genes were proven essential for plasmid maintenance (Schaffrath et al. 1995, 1997; Larsen et al. 1998).

The RNA polymerase (Fig. 3) appears to consist of two different subunits, encoded by pGKL2 ORF6 and ORF7, and homologous genes in other plasmid systems. The architecture of the enzyme resembles β and β′-subunits of the multisubunit *Escherichia coli* RNA polymerase; all β and several domains of the β′ subunit exist in the predicted pGKL2/Orf6 protein (Wilson and Meacock 1988; Thuriaux and Sentenac 1992; Schaffrath et al. 1995). However, two β′ domains, located usually close to the C terminus, reside in a separately encoded protein (pGKL2 Orf7p or its homologues; Schaffrath et al. 1997; Jeske et al. 2006a), suggesting β′ functions can be assigned to separate subunits, a scenario known for the archeon *Halobacterium halobium* (Leffers et al. 1989). Thus, concerning its domain architecture the linear plasmid encoded RNA polymerase displays a rather unique and simple structure, most closely related to multisubunit RNA polymerases, in marked contrast to the single subunit, T7-like RNA polymerase encoded by mitochondrial linear plasmids.

The predicted protein encoded by pGKL2 ORF4 (and respective homologues in other autonomous linear plasmids) belongs to the DExH/D box family which comprises RNA helicases and NTPases (Tommasino et al. 1988; Stark et al. 1990; Jeske et al. 2006a). Though ORF4 was proven to be essential for plasmid maintenance (Schaffrath et al. 1997), the precise function of the predicted polypeptide remains to be elucidated. There are (at least) 34 DExD/H helicase homologues in *Saccharomyes cerevisiae*, most of which are essential (de la Cruz et al. 1999). Members of this family are instrumental in a number of diverse processes related to RNA metabolism (such as ribosome biogenesis, mRNA splicing, RNA degradation, nuclear export) as well as transcription and translation (de la Cruz et al. 1999; Tanner and Linder 2001). Not all representatives assigned to such a family must necessarily possess helicase activity but may rather supply energy for other processes (by NTP hydrolysis), such as the disruption of ribonucleoprotein complexes (Schwer 2001; Jankowsky and Bowers 2006). Interestingly, vaccinia virus NphI, a DExH/D protein, displays ATPase but lacks helicase activity. The enzyme, stimulated by ssDNA, acts as an energy-coupling factor for transcription elongation and mRNA release upon termination (Deng and Shuman 1998). Be that as it may, though the role of the DExH/D protein encoded by pGKL2 ORF4 systems remains obscure at present, it is likely to be involved in the rather unique cytoplasmic transcription process.

An unambiguously essential component of the cytoplasmic transcription machinery is the mRNA capping enzyme (encoded by pGKL2, ORF3, and homologous genes in other systems) (Larsen et al. 1998; Tiggemann et al. 2001; Klassen et al. 2001; Jeske and Meinhardt 2006). Orf3p displays RNA triphosphatase and guanylyltransferase activities (Tiggemann et al. 2001). During mRNA cap formation, RNA triphosphatase removes the γ-phosphate from the 5′ end of the nascent mRNA, followed by formation of an unusual 5′–5′ linkage between guanosine monophosphate and the processed 5′ transcript end by the guanylyltransferase, in a reaction that involves a covalent intermediate between a lysine residue of the enzyme and GMP. Finally, the cap methyltransferase modifies the guanosine residue at position N7 utilizing *S*-adenosylmethionine (SAM) as the methyl donor (Martin et al. 1975; Shuman and Hurwitz 1981; Cong and Shuman 1993; Shuman and Schwer 1995; Bisaillon and Lemay 1997). In contrast to RNA triphosphatase and guanylyltransferase activities, cap methyltransferase activity was not obtained for the purified Orf3p, though there is a potential SAM binding site (Larsen et al. 1998; Tiggemann et al. 2001). Overall, Orf3p resembles the capping enzyme known from the cytoplasmic vaccinia virus, in which triphosphatase, guanylyltransferase, and methyltransferase domains are located in the polypeptide (vD1) encoded by the vaccinia D1 gene. However, vD1 displays only faint cap methyltransferase activity, which is stimulated tremendously (more than 30-fold) upon heterodimerization with the vaccinia D12 gene product (Cong and Shuman 1992; Higman et al. 1992, 1994; Mao and Shuman 1994). It remains to be elucidated, however, whether the methyltransferase activity of the linear plasmid encoded capping enzyme (Orf3p) requires a stimulatory subunit, too. Concerning domain architecture and sequence similarities, the linear plasmid encoded Orf3p clearly corresponds to capping enzymes of irido- and poxviridae, which differ strikingly from fungal nuclearly encoded capping enzymes as—in the latter—RNA triphosphatase, guanylyltransferase, and cap methyltransferase activities reside in separate polypeptides. The persuasive similarity of the mRNA capping machinery of cytoplasmic linear plasmids to that of cytoplasmic pox viruses again strongly supports the conclusive presumption for viral ancestry.

4.1.3 Cytoplasmic Transcription

Since the cytoplasmic transcriptase complex in general, and the RNA polymerase in particular, convincingly differs from the nuclearly encoded host enzyme, cytoplasmic transcription of linear plasmid based genes was—not astonishingly—shown to be driven by unique promoters clearly differing from nuclear ones (Kämper et al. 1989a,b, 1991; Gunge et al. 2003; Romanos and Boyd 1988; Stark et al. 1990). Based on cytoplasmic in vivo recombination approaches using either yeast nuclear or reporter genes of bacterial

origin fused to pGKL1 and pGKL2 sequences, a 6-bp spanning conserved motif preceding the native ORFs was defined as being necessary and sufficient for cytoplasmic transcription (upstream conserved sequence, UCS: 5′-ATNTGA-3′; Schründer and Meinhardt 1995; Schaffrath et al. 1996; Schickel et al. 1996; Schründer et al. 1996). The UCS motif is typically located approximately 30 bp upstream of the start codon of an ORF; however, distances may vary. Sequences distal of the UCSs were found to be irrelevant, ruling out additional promoter motifs (Schickel et al. 1996).

Consistent with the high degree of conservation of transcriptase complexes, UCS motifs are located upstream of every ORF in any cytoplasmic plasmid system, irrespective of the host (Bolen et al. 1994; Hishinuma and Hirai 1991; Fukuda et al. 1997; Klassen et al. 2001, 2002, 2004; Klassen and Meinhardt 2003; Jeske and Meinhardt 2006). Accordingly, a UCS derived from the *D. etchellsii* plasmid pPE1B was shown to act as a functional promoter in the *K. lactis* pGKL system (Klassen et al. 2001), concomitantly providing further evidence for a functionally conserved cytoplasmic transcriptase.

Transcription initiation, analyzed for a number of ORFs in the pGKL system, occurs at multiple sites, routinely around 8–16 bp downstream of the UCS motif, with a preference for one site (Romanos and Boyd 1988; Jeske et al. 2006b). Since Northern analyses revealed monocistronic transcripts only, defined termination of transcripts was suggested to occur (Schaffrath et al. 1995; Jeske et al. 2006b). To date, however, the termination mechanism remains totally obscure, as for the potential 3′ modification such as polyadenylation.

Though there are similarities between a potential DExH/D box ATPase (Orf4p) encoded by linear plasmids and the viral NphI proteins being instrumental in transcription termination, it has not been proven that this largely uncharacterized enzyme functions similarly in linear plasmid systems.

4.2 The Nonautonomous Elements

Routinely, cytoplasmic linear plasmids occur as systems in which an autonomous element is accompanied by one or more nonautonomous attendants (Table 2), some of which encode their own DNA polymerase (pGKL1, pDHL1, pPin1-3, and pDH1A), while others are lacking such genes (pPE1A, pPac1-2, pWR1A) (Hishinuma et al. 1984; Stark et al. 1984; Sor and Fukuhara 1985; Cong et al. 1994; Fukuda et al. 1997; Klassen et al. 2002, 2004; Klassen and Meinhardt 2003). As already mentioned, the DNA polymerases of the pGKL system specifically replicate their own linear plasmid (Kitada and Gunge 1988; Schaffrath et al. 1995), consistent with different TPs of autonomous and nonautonomous elements. Such a scenario may also be realized in the pDHL, pPin, and pDH1 systems. Irrespective of the DNA polymerases, a nonautonomous plasmid not only depends on the corresponding autonomous element in terms of cytoplasmic replication (SSB, TRF) but also

in transcription (RNA polymerase, helicase, and capping enzyme; see above). An obvious function of nonautonomous elements is evident in the killer systems, as they typically encode a protein toxin as well as the respective immunity function, the latter safeguarding the killer strain against its own toxin.

4.2.1 Killer Plasmids

Linear plasmid encoded toxins were detected in *K. lactis*, *Pichia inositovora*, *P. acaciae*, and *Debaryomyces robertsiae* (synonym *Wingea robertsiae*) (Gunge et al. 1981; Worsham and Bolen 1990; Hayman and Bolen 1991; Klassen and Meinhardt 2002; see Table 2). Genetic and biochemical analyses provided evidence for large (>100 kDa) heteromeric proteins which are encoded by two separate genes (Table 2). The most thoroughly studied toxin is the *K. lactis* zymocin. A detailed picture is available with respect to biogenesis and multistep action on sensitive target cells. Though there is evidence for similar mechanisms being instrumental in target cell entrance, clearly different intracellular toxin targets must be conceded in other instances, eventually leading to classification of two functional subtypes (I and II).

4.2.1.1 The pGKL1 Encoded Zymocin

(a) Biogenesis

The *K. lactis* zymocin is a heterotrimeric (α, β, γ) glycoprotein, the subunits of which display molecular masses of 99, 30, and 28 kDa, respectively (Stark and Boyd 1986). The smallest subunit (γ) is encoded by ORF4 of pGKL1; the two larger ones originate from a single gene product (Orf2p) by posttranslational processing (Hishinuma et al. 1984; Stark et al. 1984, 1990; Sor and Fukuhara 1985; Stark and Boyd 1986; Tokunaga et al. 1987).

Both Orf2p and Orf4p carry typical signal peptides mediating their entry into the secretory pathway. Upon translocation to the endoplasmic reticulum (ER), signal peptides are probably cleaved off (Tokunaga et al. 1990); subsequently, in the Golgi, processing of Orf2p at positions KR_{29} and KR_{894} occurs by means of the Kex1 protease, ultimately giving rise to mature α- and β-subunits (Stark and Boyd 1986; Wésolowski-Louvel et al. 1988; Tanguy-Rougeau et al. 1988; Stark et al. 1990). Secreted holotoxin forms the α,β,γ-complex in which β and γ are covalently linked via a disulfide bond; the α-subunit at least contains one internal disulfide bridge (Stark and Boyd 1986; Stark et al. 1990). Protein glycosylation is restricted to the α-subunit, in which a single N-linked oligosaccharide chain was detected, whereas both β and γ are non-glycosylated (Stark and Boyd 1986). Inhibition of either α-subunit glycosylation by tunicamycin or precursor processing by *kex1* mutation pre-

vents zymocin secretion (Sugisaki et al. 1985; Wésolowski-Louvel et al. 1988; Tanguy-Rougeau et al. 1988). The γ-subunit, though carrying a functional signal peptide shown to efficiently mediate secretion of heterologous proteins, fails to be efficiently secreted in the absence of the αβ-precursor protein (Tokunaga et al. 1988, 1989, 1990, 1991). Thus, zymocin γ secretion apparently requires cosecretion and interaction with the glycosylated α-subunit or the αβ-precursor (Tokunaga et al. 1990). Interestingly, the γ-subunit is efficiently secreted without the need for the αβ-precursor when the heterologous pre–pro sequence of the *S. cerevisiae* α-mating-type factor is placed in between the γ-signal peptide and the mature polypeptide (Tokunaga et al. 1989, 1990, 1992). Since the pre–pro sequence is cleaved by the KEX protease in the Golgi and basal secretion levels of the native γ-subunit are not affected by a *kex* mutation, it has been assumed that both unprocessed pre–pro sequence and αβ-precursor govern γ passage from the ER to the Golgi (Tokunaga et al. 1990; Gunge and Tokunaga 2004). Hence, even though γ secretion requires cosecretion of toxin subunits, the assembly of holotoxin is not essential and interaction between separately encoded subunits (i.e., γ and αβ) most likely occurs prior to cleavage of the αβ-precursor into mature α- and β-subunits (Tokunaga et al. 1990).

The zymocin encoding killer plasmids pGKL1 and pGKL2 have been transferred to other yeast species, such as *Kluyveromyces marxianus* (synonyms *K. fragilis* and *Candida pseudotropicalis*) and *S. cerevisiae* (Gunge and Sakaguchi 1981; Gunge et al. 1982; Sugisaki et al. 1985). Irrespective of their host they confer the ability to secrete zymocin, suggesting toxin biogenesis to be similar. Indeed, *S. cerevisiae* cells harboring pGKL1 and pGKL2 were shown to secrete zymocin consisting of correctly processed subunits. Moreover, mutations in the *KEX2* gene, which is homologous to *K. lactis KEX1*, prevent secretion of both holotoxin and the αβ-precursor (Stark et al. 1990; Tokunaga et al. 1990).

(b) Mode of Action: Access to the Target Cell

Reminiscent of the long-known *S. cerevisiae* K28 toxin as well as of bacterial and plant A/B toxins, the lethal strategy of zymocin involves import of a toxic subunit into target cells. For zymocin, the cell cycle arresting activity resides exclusively in the γ-subunit (Tokunaga et al. 1989; Butler et al. 1991b). Conditional expression of the encoding gene (pGKL1 ORF4) mimics the effect of exogenously applied native zymocin, i.e., it provokes G1 cell cycle arrest, as does exozymocin (White et al. 1989; Tokunaga et al. 1989; Stark et al. 1990; Butler et al. 1991b,c).

Though γ is responsible for the G1 arrest, it requires both of the other subunits (α and β) to gain access to target cells, a process that involves cell wall contact and transmembrane passage (Tokunaga et al. 1989; Stark et al. 1990). For genetic analysis of factors involved in early (binding and uptake) and late events (intracellular action of γ) during zymocin action,

toxin-resistant mutants were isolated and classified according to resistance or sensitivity toward intracellular γ expression (Butler et al. 1991c, 1994; Kawamoto et al. 1993; Kishida et al. 1996; Frohloff et al. 2001). Mutants affected in early steps exclusively can survive exozymocin (class I), whereas loss of essential functions involved in intracellular action of γ renders a mutant resistant to both intracellularly expressed γ and exogenously applied holotoxin (class II).

Cell wall binding is mediated by the zymocin's α-subunit, which contains a cysteine-rich chitin binding domain (PROSITE 00026) typically found in plant chitin binding proteins and chitinases. There is also a characteristic chitinase domain (Pfam PF 00704) of the glycosyl hydrolase family 18 (Stark et al. 1990; Butler et al. 1991a; Jablonowski et al. 2001a). Only quite recently, we additionally identified a LysM signature (Pfam PF 01476) in front of the chitin binding domain. LysM signatures were originally identified in cell wall degrading proteins (lysins) of bacteria and bacteriophages, in which they function as peptidoglycan binding sites (Birkeland 1994; Pontig et al. 1999). Since carbohydrate moieties in both peptidoglycan and chitin are structurally similar, it is tempting to speculate that the LysM signature acts in concert with the chitin binding domain to facilitate the ascertained tight binding of the toxin complex to the target cell's chitin (Jeske et al. 2006a). Consistently, zymocin has both chitin binding and chitinase activity; moreover, several mutations affecting the major chitin synthase (Chs3) of *S. cerevisiae* cause class I zymocin resistance, concomitantly providing convincing evidence for chitin being the zymocin receptor (Takita and Castilho-Valavivius 1993; Butler et al. 1991a; Jablonowski et al. 2001a).

Chronologically following cell wall binding but still preceding target interference, there are two key players known to be involved in γ uptake, i.e., the plasma membrane sphingolipid mannosyl-diinositolphosphoceramide $M(IP)_2C$ (synthesized by the Kti6/Ipt1 protein) and the plasma membrane ATPase Pma1, both conjointly required for toxin action from outside as judged from class I resistance of *ipt1* and *pma1* mutants (Mehlgarten and Schaffrath 2004; Zink et al. 2005). The crucial role of $M(IP)_2C$ for zymocin uptake is further supported by reduced zymocin sensitivity of different sphingolipid mutants displaying shortened $M(IP)_2C$ levels (Zink et al. 2005). In contrast to *chs3* cells, which lack zymocin binding (due to the absence of the primary cell wall receptor, see above), the *ipt1* mutant, though proficient for cell wall binding of the toxin, denies γ import; thus, the Ipt1 synthesized sphingolipid might act as a secondary membrane receptor (Zink et al. 2005). Alternatively, however, a yet unidentified $M(IP)_2C$-dependent protein may serve as such a secondary plasma membrane receptor. Normal glutaredoxin levels are significant for toxin uptake as well, since high copy *GRX3* encoding glutaredoxin provokes class II zymocin resistance (Jablonowski et al. 2001b). Glutaredoxin exceeding normal levels may impact β/γ uncoupling or γ-toxin import (Schaffrath and Meinhardt 2004).

Following membrane passage, γ relies on intracellular activation mediated by the plasma membrane H^+ATPase Pma1 (Mehlgarten and Schaffrath 2004; Zink et al. 2005). Curiously, though they are not deficient in γ uptake, *pma1* mutants display class I zymocin resistance which can be overruled by an excess of H^+ (Mehlgarten and Schaffrath 2004). Hence, γ must exist during the uptake process in a transient dormant form, which is converted by H^+ flux into the active conformation (induced either experimentally by acidification of the cell exterior or by the H^+ATPase) (Mehlgarten and Schaffrath 2004). Interestingly, other protein toxins gaining access to target cells, such as the botulinum or diphtheria toxins, depend on acidification of endocytotic vesicles by plasma membrane H^+ATPases for induction of conformational changes as well, resulting in the insertion of a membrane spanning domain into the endosomal membrane normally facilitating toxin passage into the cytoplasm. It remains to be elucidated whether such a scenario also applies to zymocin action; if so, zymocin containing endocytotic vesicles must be formed independently of the well-established End3/End4-dependent pathway (Raths et al. 1993; Wesp et al. 1997) as zymocin action requires neither of them (Jablonowski et al. 2001a).

Though the exact mechanism facilitating transmembrane passage of the γ-subunit is not entirely understood, the remarkably hydrophobic β-subunit is believed to take part (Stark et al. 1990; Jablonowski et al. 2001a; Mehlgarten and Schaffrath 2004; Zink et al. 2005). Thus, it constitutes the prime candidate among the toxin subunits for mediating transmembrane passage of γ, either directly or with the aid of endocytotic vesicles. In support of this, a similarity (even though weak) of the β-peptide to *E. coli* TolQ has been noted (Stark et al. 1990), the latter being a factor required for transmembrane passage of colicins (Sun and Webster 1987). For understanding early toxin responses, it is of particular importance to know whether zymocin's toxicity involves breakdown of the disulfide bridge connecting β and γ in holotoxin, and to define detailed functions of the sphingolipid $M(IP)_2C$ and Pma1 in binding and mediating passage across the plasma membrane.

(c) Mode of Action: The Intracellular Target

The cell cycle arresting γ-subunit strictly requires the so-called Elongator complex, originally described to be involved in elongation during RNA polymerase II transcription (Otero et al. 1999; Wittschieben et al. 1999; Winkler et al. 2001; Frohloff et al. 2001). In addition, Elongator was considered to be instrumental in rather diverse processes, such as transfer RNA (tRNA) modification and exocytosis (Huang et al. 2005; Rahl et al. 2005). Only quite recently, however, it was shown that Elongator's key function concerns posttranscriptional modification of tRNAs, and concomitantly it was proposed that transcription and exocytosis defects of respective mutants are downstream effects due to defective tRNA modification (Esberg et al. 2006).

The Elongator complex consists of six subunits, each of which is required for the formation of 5-methoxy-carbonylmethyl (mcm^5) and 5-carbamoylmethyl groups (ncm^5) at the wobble nucleoside uridine in certain tRNAs (Huang et al. 2005). Target cells lacking any one of these subunits display class II toxin resistance and, consistent with tRNA modification representing the key function of Elongator, such modified tRNAs in fact constitute the molecular targets of zymocin. In vitro, the γ-subunit cleaves three mcm^5-modified tRNA species ($tRNA^{Glu}_{UUC}$, $tRNA^{Lys}_{UUU}$, and $tRNA^{Gln}_{UUG}$) specifically at the 3′ side of the wobble nucleoside with a strong preference for $tRNA^{Glu}$, generating a 5′-hydroxyl group and a 2′3′ cyclic phosphate end (Lu et al. 2005; Jablonowski et al. 2006). Supporting the conclusion that such modified tRNAs represent the target sites, overexpression of $tRNA^{Glu}_{UUC}$ alone or $tRNA^{Lys}_{UUU}$ and $tRNA^{Gln}_{UUG}$ in combination confers zymocin resistance (Butler et al. 1994; Jablonowski et al. 2006; Lu et al. 2005).

As for Elongator subunits Elp1-6, removal of several Elongator interacting proteins, such as Kti11-13, Sit4, Sap185, and Sap190, confers immunity toward zymocin (Frohloff et al. 2001, 2003; Jablonowski et al. 2001b,c, 2004; Fichtner et al. 2002a,b, 2003; Fichtner and Schaffrath 2002; Mehlgarten and Schaffrath 2003). Since *kti11-13* mutants not only display zymocin resistance but also lack mcm^5 modifications of tRNAs (Huang et al. 2005), the paramount importance of such modification for zymocin action again became evident.

Another zymocin relevant enzyme is Trm9/Kti1, a tRNA methyltransferase probably acting in concert with Elongator to carry out wobble nucleoside modification (Kalhor and Clarke 2003; Jablonowski et al. 2006). While tRNAs of cells mutated in either *ELP1-6* or *KTI11-13* do not possess the entire mcm^5 group, *trm9/kti1* mutants lack the methyl group of mcm^5 only; nevertheless, they display robust toxin resistance (Huang et al. 2005; Lu et al. 2005; Jablonowski et al. 2006). The significance of the methyl group for target recognition by γ is supported by the finding that the zymocin-protective effect of $tRNA^{Glu}$ overexpression is abolished when the methyltransferase Trm9 is present at elevated levels (Jablonowski et al. 2006).

Interestingly, tRNA targets are known from bacterial protein toxins as well, e.g., for colicins E5 and D (Ogawa et al. 1999; Tomita et al. 2000). As for the zymocin, both latter toxins are tRNases cleaving specifically either at the 5′ or the 3′ side of the wobble nucleoside of certain tRNAs and produce a 2′3′ cyclic phosphate end (Ogawa et al. 1999; Tomita et al. 2000). However, irrespective of a similar mode of action of both bacterial toxins, the lack of sequence homology and differences in substrate recognition as well as catalytic mechanisms have led to the conclusion that such toxins have acquired similar functions through convergent evolution (Masaki and Ogawa 2002). The γ-subunit of zymocin, representing a new member of toxic tRNases, does not display significant similarity to either of the others, which may support the opinion that

attacking tRNA constitutes an effective toxic principle that can be realized independently during evolution by making use of unrelated proteins.

(d) Toxin Immunity

Essentially, zymocin producing cells are resistant toward the toxin; such self-immunity strictly relies on the availability of the killer plasmid pair pGKL1 and pGKL2; however, partially cured strains carrying exclusively pGKL2 are susceptible to toxin (Gunge et al. 1981; Niwa et al. 1981). Accordingly, immunity is accomplished by pGKL1. As already outlined, three of the four pGKL1 genes are either involved in replication (ORF1) or encode toxin subunits (ORF2 and ORF4). Since deletion of a pGKL1 region spanning ORF2, ORF3, and ORF4 gave rise to nonimmune nonkillers, whereas deletion solely of ORF2 produced immune nonkillers, pGKL1 ORF3 was held responsible for zymocin immunity (Hishinuma et al. 1984; Sor and Fukuhara 1985; Stark and Boyd 1986; Stark et al. 1990). In fact, expression of ORF3 from an ARS vector in a strain devoid of pGKL1 conferred toxin resistance (Tokunaga et al. 1987). However, the acquired immunity was only partial; high toxin doses ($400\ ng\ ml^{-1}$) still caused noticeable growth inhibition, and were indeed tolerated by cells carrying pGKL1. Moreover, expression of ORF3 from an ARS vector conferred immunity exclusively in the presence of pGKL2 (Tokunaga et al. 1987). Thus, it was suggested that pGKL2 mediates expression of ORF3 from ARS vectors; however, this lags behind pGKL1 based expression (Tokunaga et al. 1987; Stark et al. 1990). Alternatively, pGKL2 might encode an additional factor required for full zymocin immunity. There is an additional nonessential ORF on pGKL2 that does not exist in other yeast autonomous elements (ORF1; see above). Though the predicted polypeptide displays significant similarities to pGKL1 Orf3p this ORF has evidently no function in zymocin immunity, as deletion mutants remained immune (Schaffrath et al. 1992).

The biochemical basis of immunity so far remains unknown; nevertheless, pGKL1 ORF3 not only protect cells from exogenously applied zymocin but also antagonizes intracellularly expressed γ (Tokunaga et al. 1989), indicating that zymocin protection is not realized by exclusion of γ but rather involves intracellular interference with γ function.

For colicin D, which exerts a zymocin-like toxic principle (see above), it was proven that the corresponding immunity protein (ImmD) structurally mimics the respective tRNA substrate and, forming a tight complex with the toxic domain, blocks the tRNase active site (Graille et al. 2004). Whether similarities between zymocin and colicins are restricted to (presumably convergently evolved) toxic principles or extend to like immunity mechanisms needs to be discovered.

4.2.1.2 Other Killer Plasmids and Encoded Toxins

Except for pGKL1, there are three nonautonomous linear plasmids from different yeast species known to encode killer toxins, i.e., pPac1-2 of *P. acaciae*, pPin1-3 of *P. inositovora*, and pWR1A of *D. robertsiae* (Gunge et al. 1981; Worsham and Bolen 1990; Hayman and Bolen 1991; Klassen and Meinhardt 2002). Each of the aforementioned elements encodes a chitin binding protein, structurally akin to the zymocin α-subunit (Klassen and Meinhardt 2002, 2003; Klassen et al. 2004). Additionally, hydrophobic C-terminal regions very similar to the zymocin β-subunit were identified in such deduced chitin binding polypeptides, suggesting chitin binding and subsequent import of a cell cycle arresting or lethal subunit as for zymocin. Consistently, mutations affecting the target cells major chitin synthase (Chs3) conferred toxin resistance to all of the toxins (Klassen and Meinhardt 2003; Klassen et al. 2004). Moreover, reminiscent of the γ-subunit, separate structural genes encoding potentially secreted proteins were identified in either case (Fig. 3).

For the *P. inositovora* plasmid pPin1-3, such protein does indeed compare to zymocin γ and—not astonishingly—as for zymocin, toxicity is impaired in an Elongator (*elp3*) mutant (Klassen and Meinhardt 2003). The functionally related *P. inositovora* toxin together with *K. lactis* zymocin were assigned to a corporate class of linear plasmid encoded toxins (type I), which are characterized by their dependency on cell wall chitin and, more importantly, on a functional Elongator complex (Schaffrath and Meinhardt 2004; Jeske et al. 2006b). Distinct from the *K. lactis* system, however, immunity is not encoded by the linear plasmids in *P. inositovora* (Hayman and Bolen 1991). Though there is a gene similar to the *K. lactis* pGKL1 ORF3 (see above) in the toxin encoding pPin1-3 (Klassen and Meinhardt 2003), its function has not been studied.

The killer plasmids pPac1-2 and pWR1A from *P. acaciae* and *D. robertsiae*, respectively, are structurally closely related, but their resemblance to type I toxin encoding elements (pGKL1 and pPin1-3) stretches only across the ORF encoding the chitin binding polypeptide, whereas the separately encoded and potentially secreted proteins (pPac1-2 Orf2p and pWR1A Orf3p) hardly bear any likeness to zymocin γ and/or the *P. inositovora* equivalent (Klassen et al. 2004). However, as for zymocin γ, intracellular expression of either protein phenocopies the impact of the extracellularly applied toxin on target cells; thus, their function as intracellularly acting toxin subunits immediately comes to mind (Klassen et al. 2004).

In accordance with functions different from zymocin γ, toxins of *P. acaciae* (PaT) and *D. robertsiae* do not require Elongator (Klassen et al. 2004). Hence, tRNAs with an Elongator-dependent mcm^5-modified wobble uridine are unlikely to be their targets (see above). Interestingly, a mutation in the *TRM9* gene (encoding a tRNA methyltransferase known to be involved in completion of the mcm^5 modification) (Kalhor and Clarke 2003; Lu et al.

2005; Jablonowski et al. 2006) confers resistance to the type II toxin of *P. acaciae* to a certain degree (McCracken et al. 1994; our unpublished results). Taking also into consideration that other than for zymocin, overexpression of $tRNA^{Glu}_{UUC}$ has no protective consequences in type II toxicity (our unpublished results), Trm9 may have an additional cellular role besides mcm^5 modification of tRNA.

Type II toxins arrest target cells in the S phase of the cell cycle and activate the intra-S-phase DNA damage checkpoint, a scenario that agrees with a replication inhibitory and/or DNA damaging function (Klassen et al. 2004). The *P. acaciae* toxin (PaT) induces death in a two-step fashion; during the first 3–4 h in toxin, target cells lose their viability to approximately 30% that of control levels. Such a phase is characterized by hyperphosphorylation of the DNA-damage checkpoint kinase Rad53 and mutation induction, indicating DNA damage to be involved in lethality at this stage (Klassen et al. 2004; Klassen and Meinhardt 2005). Our recent work has revealed that DNA damage induced by PaT occurs during the S phase of the cell cycle and probably involves formation of broken replication forks. Congruently, mutants impaired in stalled replication fork stabilization and recovery, as well as in double-strand break (DSB) repair, react extremely sensitively to PaT (our unpublished results).

A period comprising several hours of constant viability in toxin (∼30%) follows until the final decline to less than 1% occurs. Final cell death is characterized by the appearance of typical apoptotic markers, including abnormal nuclear morphology, reactive oxygen species, DNA fragmentation, and phosphatidylserine flipping (Klassen and Meinhardt 2005). Since Rheovirus encoded killer toxins with clearly distinct modes of action (such as K1 and K28 from *S. cerevisiae* and zygocin from *Zygosaccharomyces bailii*) also activate programmed cell death in target cells (Reiter et al. 2005), apoptosis is apparently an aftereffect and not confined to individual toxins.

Interestingly, type I and type II toxins are both influenced by the mating-type status of the target cell. Diploid cells are not only significantly more resistant to both of the toxins compared to either of the haploid parents, but also overexpression of the *MATa* locus in a *MAT*α strain, as well as mutations in genes required for silencing of the cryptic mating-type loci (*SIR* genes), render target cells resistant to toxins of both types (Butler et al. 1994; Klassen et al. 2006). Such an effect is likely due to the repression of haploid-specific genes required for the action of either toxin and, thus, may be due to the cellular uptake mechanism, which is akin in either case (see above; Table 3). Moreover, both zymocin and PaT interfere with the mating competence of target cells, as these are refractory (Klassen et al. 2006). Thus, both killers act efficiently on haploid cells only and prevent them from mating, thereby ensuring most effective target cell killing under unfavorable environmental conditions, and causing sporulation of target cells which compete for limited resources (Klassen et al. 2006).

Table 3 Genes essential for killer toxin action, arranged according to their function in target cells. Toxin resistance and susceptibility of individual mutants to killer toxins from *K. lactis* (Kl), *P. acaciae* (Pa), *P. inositovora* (Pi), and *D. robertsiae* (Wr) are indicated. Pa, Pi, and Wr resistance/sensitivity is not proven for each gene in a given category

Target cell modification (relevant genes)	Toxin relevant process	Toxin resistance	Toxin susceptibility	Refs.
Chitin synthesis defect *CHS3, CHS4, CHS5, CHS6, CHS7*	Cell wall receptor; uptake	Kl/Pa/Pi/Wr		Takita and Castilho-Valavivius 1993; Kawamoto et al. 1993; Butler et al. 1991a; Jablonowski et al. 2001a; Klassen et al. 2004; McCracken et al. 1994
Plasma membrane sphingolipid synthesis defect *IPT1, LAG1, LAC1*	Uptake; membrane receptor (?)	Kl	Pa	Zink et al. 2005; McCracken et al. 1994; unpublished results
Plasma membrane H^+ATPase defect *PMA1, PTK2*	Intracellular activation	Kl/Pa		Mehlgarten and Schaffrath 2004; McCracken et al. 1994; unpublished results
Simultaneous expression of *MAT*a and *MAT*α *SIR1-4; HMR, HML*	Uptake (?)	Kl/Pa		Butler et al. 1994; Klassen et al. 2006
High copy glutaredoxin *GRX3*	Uptake (?)	Kl		Jablonowski et al. 2001b
Elongator subunit defect *ELP1/IKI3, ELP2, ELP3, ELP4, ELP5/IKI1, ELP6*	tRNA modification, transcription, exocytosis	Kl/Pi	Pa/Wr	Yajima et al. 1997; Frohloff et al. 2001; Klassen et al. 2004; McCracken et al. 1994
Elongator relevant factor defect *KTI11, KTI12, ATS1/KTI13, HRR25/KTI14, URM1, UBA4, SIT4, SAP185, SAP190*; overexpression of *SAP155*; *KTI12*	Elongator modification/ interaction	Kl	Pa	Frohloff et al. 2001; Fichtner and Schaffrath 2002; Mehlgarten and Schaffrath 2003; Fichtner et al. 2003; Jablonowski et al. 2001b,c; Klassen et al. 2004; McCracken et al. 1994

Table 3 (continued)

Target cell modification (relevant genes)	Toxin relevant process	Toxin resistance	Toxin susceptibility	Refs.
tRNA overexpresion $tRNA^{Glu}_{UUC}$; $tRNA^{Lys}_{UUU}$ + $tRNA^{Gln}_{UUG}$	Target overexpression	Kl	Pa	Butler et al. 1994; Lu et al. 2005; Jablonowski et al. 2006; unpublished results
Methyltransferase *TRM9/KTI1*	tRNA methylation	Kl/Pa		Lu et al. 2005; Jablonowski et al. 2006; McCracken et al. 1994; unpublished results
Diphthamide synthesis *DPH2, YIL103w*	?	Kl		Fichtner et al. 2003

As for type I, immunity to type II toxins is procured by genes located on the nonautonomous, toxin encoding elements pPac1-2 and pWR1A (Worsham and Bolen 1991; our unpublished results). Predicted immunity factors encoded by pPac1-2 ORF4 and pWR1A ORF5 (our unpublished results) display striking similarities to each other but not to type I immunity encoded by pGKL1 ORF3 (see above). As for the latter, however, PaT immunity acts by intracellular interference with Orf2p function, rather than preventing its uptake (unpublished results).

In accordance with fundamental differences concerning both toxicity and immunity of type I and II toxins, pGKL1 Orf3p protects exclusively from the action of type I toxin, whereas at least partial cross-immunity was detected between *P. acaciae* and *D. robertsiae* toxin/immunity systems (our unpublished results).

Localization of an immunity conferring gene on the toxin encoding plasmid provides an autoselection mechanism for the entire system, as cells of a natural population having occasionally lost such an element are counterselected by toxin-producing neighbor cells. All known linear plasmid encoded killer toxins (except that of *P. inositovora*, see above) eliminate plasmid-free strains of the same species, suggesting that such toxins may indeed have autoselection purposes rather than providing an advantage to the host. Generally speaking, in killer/immunity combinations encoded by yeast linear plasmids there are (at least two) functionally distinct cargo proteins (types I and II), which exist conjointly with a respective compatible immunity factor. The cargo proteins, representing the effective principle of the toxin in either case, rely on a conserved carrier (the chitin binding/hydrophobic protein) which facilitates transport into target cells.

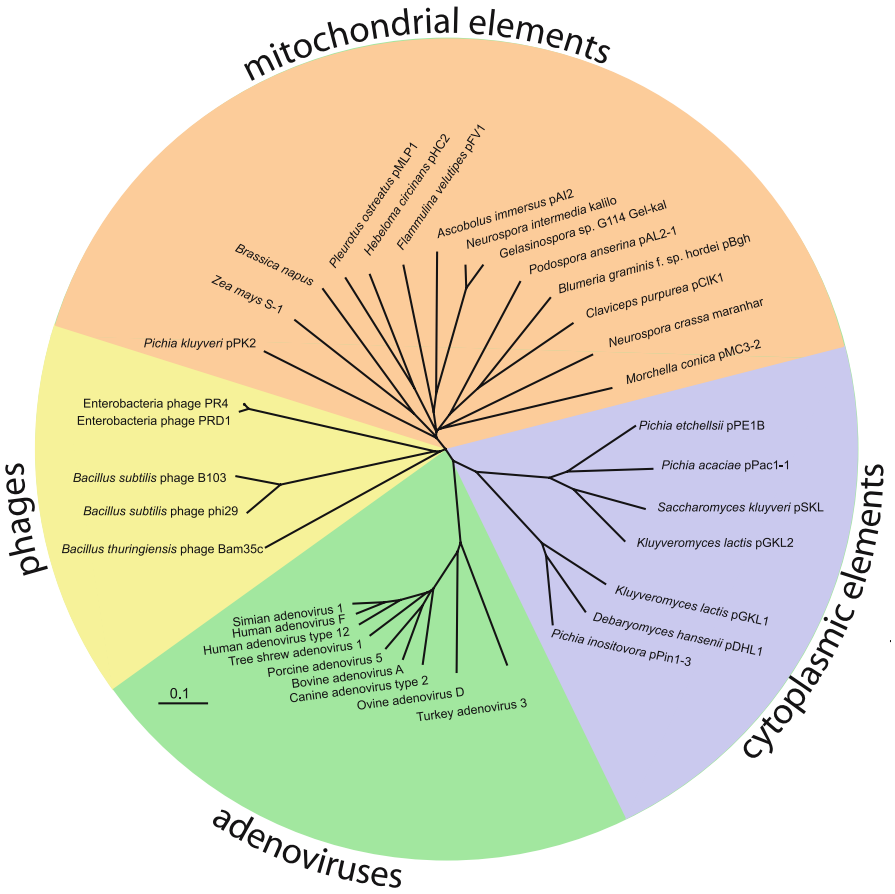
4.2.2 Cryptic Elements

Despite their frequent occurrence, a killer phenotype could not be attributed to most nonautonomous elements (Table 2). Since, to date, no other alternative traits (other than killer toxin production and immunity) could be disclosed, they were considered cryptic elements (Cong et al. 1994; Fukuhara 1995). For some of them, entire or partial nucleotide sequence data are available and potential coding capacities were addressed by heterologous hybridization to pGKL1 derived probes (Cong et al. 1994; Fukuda et al. 1997, 2004; Klassen et al. 2002). Accordingly, cryptic elements may either encode their own DNA polymerase (pDHL1, pDH1A) or are devoid of it (pPE1A); the latter holds true also for the nonautonomous killer elements (pPac1-2, pWR1A). Surprisingly, each of the nonautonomous plasmids contains an ORF encoding a polypeptide resembling the zymocin αβ-precursor, i.e., the conserved toxin carrier complex (see above). A detailed inspection, however, is only possible for pPE1A, since it is the only cryptic linear plasmid entirely sequenced. There is a mutation in the chitinase active site and no hydrophobic region similar to zymocin β (seen in all linear plasmid encoded killer toxins) could be identified. Moreover, an additional secreted protein (like zymocin γ) is not encoded. Thus, pPE1A unlikely encodes a killer toxin, which may have escaped detection. Nevertheless, the zymocin α-like protein encoded by pPE1A ORF2 is secreted and binds to chitin in vitro (Klassen et al. 2002).

In summary, there is convincing evidence for conserved zymocin α-like protein even in cryptic elements, though its function outside of a killer toxin complex remains obscure at present. Stable inheritance of such plasmids in many different yeast species could, however, hardly be understood in cases where there is no autoselection or an alternative positive effect for the host cell.

5 Evolution of Linear Plasmids in Eukaryotes

Eukaryotic linear plasmids, either mitochondrial or cytoplasmic, in principle replicate like genomes of adenoviruses and phi29-like bacteriophages (see above), which has led to the conclusion that they share a common ancestor (Meinhardt et al. 1986, 1990; Kuzmin et al. 1988; Oeser and Tudzynski 1989; Rohe et al. 1992; Kempken et al. 1992). For reconstruction of the phylogeny, both DNA and RNA polymerases have been employed, since the respective genes were found in almost every linear plasmid system. Figure 4 shows an updated phylogenetic tree based on the viral B-type DNA polymerase sequences with basic features agreeing with previous calculations (Rohe et al.



1992; Kempken et al. 1992). Mitochondrial and cytoplasmic linear plasmids are clearly separated, which is indicative of early phylogenetic divergence and adaptation to different cellular compartments. Moreover, cytoplasmic DNA polymerases are more closely related to enzymes from adenoviruses than to those of mitochondrial plasmids. Along with the plant linear plasmids, which also reside in mitochondria, the latter form a separate branch. Adenoviruses as well as the bacteriophage PRD1 (which infects Gram-negative bacteria and employs a phi29-like replication mechanism) were recently shown to display clear similarities concerning virion structure (Benap and Steven 2000; Benson et al. 1999, 2000; San Martin and Burnett 2003). Such findings agree with the assumption that adenoviruses and (PRD1/phi29-like) bacteriophages have a common ancestor which predates the occurrence of eukaryotes (Davison et al. 2003).

The calculated relatedness of cytoplasmic linear plasmids and adenoviruses points to separation from their ancestor prior to the emergence of

◀ **Fig. 4** Phylogenetic tree for protein-primed replicating genetic elements based on B-type DNA polymerases. Besides bacteriophages, present day manifestations in eukaryotes exist in mitochondria, in the cytoplasm, and in the nucleus (adenoviruses). Since mitochondrial elements are clearly separated from the cytoplasmic and nuclear plasmids, they must have diverged early during evolution; divergence of the nuclear and cytoplasmic elements dates back to a later point. Yet later, autonomous and nonautonomous elements split. The bar denotes 10% divergence. GenBank accession numbers of individual sequences: CAC08221.2: *Pichia etchellsii* pPE1A; CAA38621.1: *Saccharomyces kluyveri* pSKL; CAA25568.1: *Kluyveromyces lactis* pGKL1; CAA30603.1: *K. lactis* pGKL2; CAA09497.1: *Debaryomyces hansenii* pDHL1; CAD91889.1: *Pichia inositovora* pPin1-3; CAA72340.1: *Pichia kluyveri* pPK2; BAB13496.1: *Flammulina velutipes* pFV1; CAA36327.1: *Neurospora intermedia* kalilo; NP_053000.1: *Pleurotus ostreatus* pMLP1; CAA43117.2: *Podospora anserina* pAL2-1; NP_862206.1: *Blumeria graminis* f. sp. hordei pBgh; AAB41447.1: *Gelasinospora* sp. G114 Gel-kal; JQ0301: *Claviceps purpurea* pClK1; S26947: *Podospora anserina* pAL2-1; S05362: *Ascobolus immersus* pAI2; S26985: *Neurospora crassa* maranhar; CAA45364.2: *Morchella conica* pMC3-2; BAC16364.1: *Brassica napus* linear plasmid; P10582: *Zea mays* S-1; NP_659515.1: Ovine adenovirus D; AP_000236.1: Porcine adenovirus 5; YP_094032.1: Bovine adenovirus A; AAN84890.1: Tree shrew adenovirus 1; AP_000478.1: Turkey adenovirus 3; AP_000613.1: Canine adenovirus type 2; YP_213966.1: Simian adenovirus 1; AAA42478.1: Human adenovirus type 12; NP_040853.1: Human adenovirus F; P03680: phage phi29; AAP83475.1: phage Bam35c; NP_690635.1: phage B103; AAX45903.1: phage PRD1; AAX45594.1: phage PR4

eukaryotes. In contrast, the distinct separation of mitochondrial elements from both cytoplasmic linear plasmids and adenoviruses suggests that mitochondrial elements diverged later, possibly in the postendosymbiont era. Consistently, plasmid pPK2 of *Pichia kluyveri*, one of the few known mitochondrial linear yeast elements (Blaisonneau et al. 1999), is more closely related to mitochondrial linear plasmids from filamentous fungi and plants than to the cytoplasmic linear plasmids from other *Pichia* species. In general, the relatedness of mitochondrial linear elements reflects the phylogenetic relations of their hosts (Rohe et al. 1992), suggesting them to originate from a single mitochondrial ancestor which may have been acquired as an endosymbiotic (bacteriophage-like) genetic element. In contrast to bacteriophages and adenoviruses, linear plasmids lost almost all viral attributes (such as a capsid or an infectious cycle), presumably as an adaptation to the propagation mode of their hosts, since extranuclear inheritance, either as cytoplasmic or mitochondrial element, does not necessarily include an infectious virion for efficient propagation.

6 Concluding Remarks and Perspectives

A large number of eukaryotic microbes and plants contain linear plasmids which are descendants of ancient viruses. Hence they afford an opportu-

nity to trace evolutionary mapping beginning with preeukaryotic elements to the point of vertebrate adenoviruses. In filamentous fungi such virus-like elements are exclusively confined to mitochondria; here, they commonly exhibit extremely minimalized genomes encoding basically and essentially only a DNA and an RNA polymerase. Mitochondrial linear plasmids exist either as neutral passengers within the organelle or as harmful attendants which—by integrating into the mitochondrial genome—shorten the host's life span. Studying aging in such interactions may facilitate general insights into molecular mechanisms of senescence.

The phylogenetic root of the T7-like RNA polymerase, highly conserved among mitochondrial linear plasmids, still remains obscure even though the organelles possess enzymes of that type anyway. Concerning the enzymatic repertoire needed for cytoplasmic gene expression of yeast linear plasmids, much work is to be done until a comprehensive picture of the rather unique process of cytoplasmic transcription emerges, including termination and posttranscriptional modification.

For killer toxins of either type (I and II), research will have to address early events, such as binding to the cell wall receptor and the probable transport to a yet unknown membrane receptor. Elucidating the definite mode of DNA damage occurring in type II toxin-treated cells constitutes another challenge. Investigating transmembrane passage and intracellular trafficking of the toxic subunits will not only provide basic knowledge for this rather peculiar phenomenon, but will also help to establish similarities and differences to known protein toxins of clearly diverse evolutionary origins.

Acknowledgements We thank J. Paluszynski for helpful comments on the manuscript. Financial support by the Deutsche Forschungs Gemeinschaft (DFG) grant no. ME 1142/5-1 is gratefully acknowledged.

References

Arganoza MT, Min J, Hu Z, Akins RA (1994) Distribution of seven homology groups of mitochondrial plasmids in Neurospora: evidence for widespread mobility between species in nature. Curr Genet 26:62–73

Belnap DM, Steven AC (2000) "Déjà vu all over again": the similar structures of bacteriophage PRD1 and adenovirus. Trends Microbiol 8:91–93

Benson SD, Bamford JKH, Bamford DH, Burnett RM (1999) Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. Cell 98:825–833

Benson SD, Bamford JKH, Bamford DH, Burnett RM (2002) The X-ray crystal structure of P3, the major coat protein of the lipid-containing bacteriophage PRD1, at 1.65 Å resolution. Acta Crystallogr D Biol Crystallogr 58:39–59

Bertrand H (2000) Role of mitochondrial DNA in the senescence and hypovirulence of fungi and potential for plant disease control. Annu Rev Phytopathol 38:397–422

Birkeland NK (1994) Cloning, molecular characterization, and expression of the genes encoding the lytic functions of lactococcal bacteriophage phi LC3: a dual lysis system of modular design. Can J Microbiol 40:658–665

Bisaillon M, Lemay G (1997) Viral and cellular enzymes involved in synthesis of mRNA cap structure. Virology 236:1–7

Blaisonneau J, Nosek J, Fukuhara H (1999) Linear DNA plasmid pPK2 of *Pichia kluyveri*: distinction between cytoplasmic and mitochondrial linear plasmids in yeasts. Yeast 15:781–791

Bolen PL, Kurtzman CP, Ligon JM, Mannarelli BM, Bothast RJ (1992) Physical and genetic characterization of linear DNA plasmids from the heterothallic yeast *Saccharomycopsis crataegensis*. Antonie Van Leuwenhoek 61:195–295

Bolen PL, Eastman EM, Cihak PL, Hayman GT (1994) Isolation and sequence analysis of a gene from the linear DNA plasmid pPacl-2 of *Pichia acaciae* that shows similarity to a killer toxin gene of *Kluyveromyces lactis*. Yeast 10:403–414

Butler AR, O'Donnell RW, Martin VJ, Gooday GW, Stark MJ (1991a) *Kluyveromyces lactis* toxin has an essential chitinase activity. Eur J Biochem 199:483–488

Butler AR, Porter M, Stark MJR (1991c) Intracellular expression of Kluyveromyces lactis toxin γ subunit mimics treatment with exogenous toxin and distinguishes two classes of toxin-resistant mutant. Yeast 7:617–625

Butler AR, White JH, Stark MJR (1991b) Analysis of the response of *Saccharomyces cerevisiae* cells to *Kluyveromyces lactis* toxin. J Gen Microbiol 137:1749–1757

Butler AR, White JH, Folawiyo Y, Edlin A, Gardiner D, Stark MJR (1994) Two *Saccharomyces cerevisiae* genes which control sensitivity to G1 arrest induced by *Kluyveromyces lactis* toxin. Mol Cell Biol 14:6306–6316

Cermakian N, Ikeda TM, Cedergren R, Gray MW (1996) Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. Nucleic Acids Res 24:648–654

Cermakian N, Ikeda TM, Miramontes P, Lang BF, Gray MW, Cedergren R (1997) On the evolution of the single-subunit RNA polymerases. J Mol Evol 45:671–681

Challberg MD, Kelly TJ (1982) Eukaryotic DNA replication: viral and plasmid model systems. Annu Rev Biochem 51:901–934

Chan BS, Court DA, Vierula PJ, Bertrand H (1991) The kalilo linear senescence-inducing plasmid of Neurospora is an invertron and encodes DNA and RNA polymerases. Curr Genet 20:225–237

Chen WB, Han JF, Jong SC, Chang SC (2000) Isolation, purification, and characterization of a killer protein from *Schwanniomyces occidentalis*. Appl Environ Microbiol 66:5348–5352

Cong P, Shuman S (1992) Methyltransferase and subunit association domains of Vaccinia virus mRNA capping enzyme. J Biol Chem 267:16424–16429

Cong P, Shuman S (1993) Covalent catalysis in nucleotidyl transfer: a KTDG motif essential for enzyme–GMP complex formation by mRNA capping enzyme is conserved at the active sites of RNA and DNA ligases. J Biol Chem 268:7256–7260

Cong YS, Yarrow D, Li YY, Fukuhara H (1994) Linear DNA plasmids from *Pichia etchellsii*, *Debaryomyces hansenii* and *Wingea robertsiae*. Microbiology 140:1327–1335

Court DA, Bertrand H (1992) Genetic organization and structural features of maranhar, a senescence-inducing linear mitochondrial plasmid of *Neurospora crassa*. Curr Genet 22:385–397

Davison AJ, Benko M, Harrach B (2003) Genetic content and evolution of adenoviruses. J Gen Virol 84:2895–2908

de la Cruz J, Kressler D, Linder P (1999) Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. Trends Biochem Sci 24(5):192–198

Deng L, Shuman S (1998) Vaccinia NPH-I, a DExH-box ATPase, is the energy coupling factor for mRNA transcription termination. Genes Dev 12:538–546

Dufour E, Mendez J, Lazaro JM, de Vega M, Blanco L, Salas M (2000) An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. J Mol Biol 304:289–300

Dufour E, Rodriguez I, Lazaro JM, de Vega M, Salas M (2003) A conserved insertion in protein-primed DNA polymerases is involved in primer terminus stabilisation. J Mol Biol 331:781–794

Düvell A, Hessberg-Stutzke H, Oeser B, Rogmann-Backwinkel P, Tudzynski P (1988) Structural and functional analysis of mitochondrial plasmids in *Claviceps purpurea*. Mol Gen Genet 214:128–134

Esberg A, Huang B, Johansson MJ, Bystrom AS (2006) Elevated levels of two tRNA species bypass the requirement for elongator complex in transcription and exocytosis. Mol Cell 24:139–148

Fichtner L, Schaffrath R (2002) KTI11 and KTI13, *Saccharomyces cerevisiae* genes controlling sensitivity to G1 arrest induced by *Kluyveromyces lactis* zymocin. Mol Microbiol 44:865–875

Fichtner L, Frohloff F, Burkner K, Larsen M, Breunig KD, Schaffrath R (2002a) Molecular analysis of KTI12/TOT4, a *Saccharomyces cerevisiae* gene required for *Kluyveromyces lactis* zymocin action. Mol Microbiol 43:783–791

Fichtner L, Frohloff F, Jablonowski D, Stark MJ, Schaffrath R (2002b) Protein interactions within *Saccharomyces cerevisiae* Elongator, a complex essential for *Kluyveromyces lactis* zymocicity. Mol Microbiol 45:817–826

Fichtner L, Jablonowski D, Schierhorn A, Kitamoto HK, Stark MJ, Schaffrath R (2003) Elongator's toxin-target (TOT) function is nuclear localization sequence dependent and suppressed by post-translational modification. Mol Microbiol 49:1297–1307

Francou F (1981) Isolation and characterization of a linear DNA molecule in the fungus *Ascobolus immersus*. Mol Gen Genet 184:440–444

Frohloff F, Fichtner L, Jablonowski D, Breuning KD, Schaffrath R (2001) *Saccharomyces cerevisiae* elongator mutations confer resistance to the *Kluyveromyces lactis* zymocin. EMBO J 20:1993–2003

Frohloff F, Jablonowski D, Fichtner L, Schaffrath R (2003) Subunit communications crucial for the functional integrity of the yeast RNA polymerase II elongator (γ toxin target (TOT)) complex. J Biol Chem 278:956–961

Fukuda K, Maebuchi M, Takata H, Gunge N (1997) The linear plasmid pDHL1 from *Debaryomyces hansenii* encodes a protein highly homologous to the pGKL1-plasmid DNA polymerase. Yeast 13:613–620

Fukuda K, Jin-Shan C, Kawano M, Sudo K, Gunge N (2004) Stress responses of linear plasmids from *Debaryomyces hansenii*. FEMS Microbiol Lett 237:243–248

Fukuhara H (1995) Linear DNA plasmids of yeasts. FEMS Microbiol Lett 131:1–9

Galligan JT, Kennell JC (2007) Retroplasmids: Linear and circular plasmids that replicate via reverse transcription. In: Meinhardt F, Klassen R (eds) Microbiology Monographs, vol 7: Microbial linear plasmids. Springer, Berlin Heidelberg (in press)

Giese H, Christiansen SK, Jensen HP (1990) Extrachromosomal plasmid-like DNA in the obligate parasitic fungus *Erysiphe graminis* f. sp. *hordei*. Theor Appl Genet 79:56–64

Giese H, Lyngkjaer MF, Stummann BM, Grell MN, Christiansen SK (2003) Analysis of the structure and inheritance of a linear plasmid from the obligate biotrophic fungus *Blumeria graminis* f. sp. *hordei* Mol Genet Genomics 269:699–705

Graille M, Mora L, Buckingham RH, van Tilbeurgh H, de Zamaroczy M (2004) Structural inhibition of the colicin D tRNase by the tRNA-mimicking immunity protein. EMBO J 23:1474–1482

Greider CW, Blackburn EH (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell 43:405–413

Greider CW, Blackburn EH (1989) A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature 337:331–337

Griffiths AJF (1992) Fungal senescence. Annu Rev Genet 26:351–372

Griffiths AJF (1995) Natural plasmids of filamentous fungi. Microbiol Rev 59:673–685

Gunge N (1995) Plasmid DNA and the killer phenomenon in *Kluyveromyces*. In: Kück U (ed) Genetics and biotechnology: the mycota, vol 2. Springer, Berlin, pp 189–209

Gunge N, Sakaguchi K (1981) Intergeneric transfer of deoxyribonucleic acid killer plasmids, pGKl1 and pGKl2, from *Kluyveromyces lactis* into *Saccharomyces cerevisiae* by cell fusion. J Bacteriol 147:155–160

Gunge N, Tokunaga M (2004) Linear DNA plasmids and killer system in *Kluyveromyces lactis*. In: Kück U (ed) Genetics and biotechnology: the mycota II, 2nd edn. Springer, Berlin, pp 199–217

Gunge N, Tamaru A, Ozawa F, Sakaguchi K (1981) Isolation and characterization of linear deoxyribonucleic acid plasmids from *Kluyveromyces lactis* and the plasmid-associated killer character. J Bacteriol 145:382–390

Gunge N, Murata K, Sakaguchi K (1982) Transformation of *Saccharomyces cerevisiae* with linear DNA killer plasmids from *Kluyveromyces lactis*. J Bacteriol 151:462–464

Gunge N, Takata H, Matsuura A, Fukuda K (2003) Progressive rearrangement of telomeric sequences added to both the ITR ends of the yeast linear pGKL plasmid. Biol Proced Online 5:29–42

Hashiba T, Nagasaka A (2007) Hairpin plasmids from the plant pathogenic fungus *Rhizoctonia solani* and *Fusarium oxysporum*. In: Meinhardt F, Klassen R (eds) Microbiology Monographs, vol 7: Microbial linear plasmids. Springer, Berlin Heidelberg (in press)

Hayman GT, Bolen BL (1991) Linear DNA plasmids of *Pichia inositovora* are associated with a novel killer toxin activity. Curr Genet 19:389–393

Hermanns J, Osiewacz HD (1992) The linear mitochondrial plasmid pAL2–1 of a long-lived *Podospora anserina* mutant is an invertron encoding a DNA and RNA polymerase. Curr Genet 22:491–500

Hermanns J, Asseburg A, Osiewacz HD (1994) Evidence for a life span-prolonging effect of a linear plasmid in a longevity mutant of *Podospora anserina*. Mol Gen Genet 243:297–307

Hermanns J, Asseburg A, Osiewacz HD (1995) Evidence for giant linear plasmids in the ascomycete *Podospora anserina*. Curr Genet 27:379–386

Hertwig S (2007) Linear Plasmids and Prophages in Gram Negative Bacteria. In: Meinhardt F, Klassen R (eds) Microbiology Monographs, vol 7: Microbial linear plasmids. Springer, Berlin Heidelberg (in press)

Higman HA, Bourgeois N, Niles EG (1992) The Vaccinia virus mRNA (guanine-N7-)-methyltransferase requires both subunits of the mRNA capping enzyme for activity. J Biol Chem 267:16430–16437

Higman MA, Christen LA, Niles EG (1994) The mRNA (guanine-7)methyltransferase domain of the Vaccinia virus mRNA capping enzyme: expression in *Escherichia coli* and structural and kinetic comparison to the intact capping enzyme. J Biol Chem 269:14974–14981

Hishinuma F, Hirai K (1991) Genome organization of the linear plasmid, pSKL, isolated from *Saccharomyces kluyveri*. Mol Gen Genet 226:97–106

Hishinuma F, Nakamura K, Hirai K, Nishizawa R, Gunge N, Maeda T (1984) Cloning and nucleotide sequence of the DNA killer plasmids from yeast. Nucleic Acids Res 12:7581–7597

Huang B, Johansson MJ, Bystrom AS (2005) An early step in wobble uridine tRNA modification requires the Elongator complex. RNA 11:424–436

Ito J, Braithwaite DK (1991) Compilation and alignment of DNA polymerase sequences. Nucleic Acids Res 19:4045–4057

Jablonowski D, Fichtner L, Martin VJ, Klassen R, Meinhardt F, Stark MJR, Schaffrath R (2001a) *Saccharomyces cerevisiae* cell wall chitin, the potential *Kluyveromyces lactis* zymocin receptor. Yeast 18:1285–1299

Jablonowski D, Butler AR, Fichtner L, Gardiner D, Schaffrath R, Stark MJR (2001b) Sit4p protein phosphatase is required for sensitivity of *Saccharomyces cerevisiae* to *Kluyveromyces lactis* zymocin. Genetics 159:1479–1489

Jablonowski D, Frohloff F, Fichtner L, Stark MJR, Schaffrath R (2001c) *Kluyveromyces lactis* zymocin mode of actin is linked to RNA polymerase II function via Elongator. Mol Microbiol 42:1095–1105

Jablonowski D, Fichtner L, Stark MJ, Schaffrath R (2004) The yeast elongator histone acetylase requires Sit4-dependent dephosphorylation for toxin-target capacity. Mol Biol Cell 15:1459–1469

Jablonowski D, Zink S, Mehlgarten C, Daum G, Schaffrath R (2006) $tRNA^{Glu}$ wobble uridine methylation by Trm9 identifies Elongator's key role for zymocin-induced cell death in yeast. Mol Microbiol 59:677–688

Jankowsky E, Bowers H (2006) Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. Nucleic Acids Res 34:4181–4188

Jeske S, Meinhardt F (2006) Autonomous cytoplasmic linear plasmid pPac1-1 of *Pichia acaciae*: molecular structure and expression studies. Yeast 23:476–486

Jeske S, Meinhardt F, Klassen R (2006a) Extranuclear inheritance: virus-like DNA elements in yeast. In: Esser K, Lüttge U, Beyschlag W, Murata J (eds) Progress in Botany, vol 68. Springer, Berlin Heidelberg New York, pp 98–129

Jeske S, Tiggemann M, Meinhardt F (2006b) Yeast autonomous linear plasmid pGKL2 ORF9 is an actively transcribed essential gene with multiple transcription start points. FEMS Microbiol Lett 255:321–327

Kalhor HR, Clarke S (2003) Novel methyltransferase for modified uridine residues at the wobble position of tRNA. Mol Cell Biol 23:9283–9292

Kämper J, Meinhardt F, Gunge N, Esser K (1989a) New recombinant linear DNA elements derived from *Kluyveromyces lactis* killer plasmids. Nucleic Acids Res 17(4):1781

Kämper J, Meinhardt F, Gunge N, Esser K (1989b) In vivo construction of linear vectors based on killer plasmids from *Kluyveromyces lactis*: selection of a nuclear gene results in attachment of telomeres. Mol Cell Biol 9(9):3931–3937

Kämper J, Esser K, Gunge N, Meinhardt F (1991) Heterologous gene expression on the linear DNA killer plasmid from *Kluyveromyces lactis*. Curr Genet 19:109–118

Kamtekar S, Berman AJ, Wang J, Lázaro JM, de Vega M, Blanco L, Salas M, Steitz TA (2004) Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29. Mol Cell 16:609–618

Kamtekar S, Ho RS, Cocco MJ, Li W, Wenwieser SV, Boocock MR, Grindley ND, Steitz TA (2006) Implications of structures of synaptic tetramers of gamma delta resolvase for the mechanism of recombination. Proc Natl Acad Sci USA 103:10642–10647

Kawamoto S, Sasaki T, Itahashi S, Hatsuyama Y, Ohno T (1993) A mutant allele skt5 affecting protoplast regeneration and killer toxin resistance has double mutations in

its wild-type structural gene in *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem 57:1391–1393

Kempken F, Meinhardt F, Esser K (1989) In organello replication and viral affinity of linear, extrachromosomal DNA of the ascomycete *Ascobolus immersus*. Mol Gen Genet 218:523–530

Kempken F, Hermanns J, Osiewacz HD (1992) Evolution of linear plasmids. J Mol Evol 35:502–513

Kim EK, Jeong JH, Youn HS, Koo YB, Roe JH (2000) The terminal protein of a linear mitochondrial plasmid is encoded in the N-terminus of the DNA polymerase gene in white-rot fungus *Pleurotus ostreatus*. Curr Genet 38:283–290

Kishida M, Tokunaga M, Katayose Y, Yajima H, Kawamura-Watabe A, Hishinuma F (1996) Isolation and genetic characterization of pGKL killer-insensitive mutants (iki) from *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem 60:798–801

Kitada K, Gunge N (1988) Palindrome-hairpin linear plasmids possessing only a part of the ORF1 gene of the yeast killer plasmid pGKL1. Mol Gen Genet 215:46–52

Kitada K, Hishinuma H (1987) A new linear plasmid isolated from the yeast *Saccharomyces kluyveri*. Mol Gen Genet 206:377–381

Klassen R, Meinhardt F (2002) Linear plasmids pWR1A and pWR1B of the yeast *Wingea robertsiae* are associated with a killer phenotype. Plasmid 48:142–148

Klassen R, Meinhardt F (2003) Structural and functional analysis of the killer element pPin1-3 from *Pichia inositovora*. Mol Genet Genomics 270:190–199

Klassen R, Meinhardt F (2005) Induction of DNA damage and apoptosis in *Saccharomyces cerevisiae* by a yeast killer toxin. Cell Microbiol 7:393–401

Klassen R, Tontsidou L, Larsen M, Meinhardt F (2001) Genome organization of the linear cytoplasmic element pPE1B from *Pichia etchellsii*. Yeast 18:953–961

Klassen R, Jablonowski D, Schaffrath R, Meinhardt F (2002) Genome organization of the linear *Pichia etchellsii* plasmid pPE1A: evidence for expression of an extracellular chitin binding protein homologous to the α-subunit of the *Kluyveromyces lactis* killer toxin. Plasmid 47:224–233

Klassen R, Teichert S, Meinhardt F (2004) Novel yeast killer toxins provoke S-phase arrest and DNA damage checkpoint activation. Mol Microbiol 53:263–273

Klassen R, Jablonowski D, Stark MJR, Schaffrath R, Meinhardt F (2006) Mating type locus control of killer toxins from *Kluyveromyces lactis* and *Pichia acaciae*. FEMS Yeast Res 6:404–413

Kobryn K (2007) The linear hairpin replicons of *Borrelia burgdorferi*. In: Meinhardt F, Klassen R (eds) Microbiology Monographs, vol 7: Microbial linear plasmids. Springer, Berlin Heidelberg (in press)

Kuzmin EV, Levchenko IV, Zaitseva GN (1988) S2 plasmid from cms-S-maize mitochondria potentially encodes a specific RNA polymerase. Nucleic Acids Res 16:4177

Larsen M, Meinhardt F (2000) *Kluyveromyces lactis* killer system: identification of a new gene encoded by pGKL2. Curr Genet 38:271–275

Larsen M, Gunge N, Meinhardt F (1998) *Kluyveromyces lactis* killer plasmid pGKL2: evidence for a viral-like capping enzyme encoded by OFR3. Plasmid 40:243–246

Leffers H, Gropp F, Lottspeich F, Zillig W, Garrett RA (1989) Sequence, organisation, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. J Mol Biol 206:1–17

Levings CS, Sederoff RR (1983) Nucleotide sequence of the S-2 mitochondrial DNA from the S cytoplasm of maize. Proc Natl Acad Sci USA 80:4055–4059

Ligon JM, Bolen PL, Hill DS, Bothast RJ, Kurtzman CP (1989) Physical and biological characterization of linear DNA plasmids of the yeast *Pichia inositovora*. Plasmid 2:185–194

Lu J, Huang B, Esberg A, Johansson MJ, Bystrom AS (2005) The *Kluyveromyces lactis* γ toxin targets tRNA anticodons. RNA 11:1648–1654

Maas MF, de Boer HJ, Debets AJ, Hoekstra RF (2004) The mitochondrial plasmid pAL2-1 reduces calorie restriction mediated life span extension in the filamentous fungus *Podospora anserina*. Fungal Genet Biol 41:865–871

Maas MF, van Mourik A, Hoekstra RF, Debets AJ (2005) Polymorphism for pKALILO based senescence in Hawaiian populations of *Neurospora intermedia* and *Neurospora tetrasperma*. Fungal Genet Biol 42:224–232

Mao X, Shuman S (1994) Intrinsic RNA (guanine-7)methyltransferase activity of the Vaccinia capping enzyme D1 subunit is stimulated by the D12 subunit: identification of amino acid residues in the D1 protein required for subunit association and methyl group transfer. J Biol Chem 269:24472–24479

Martin SA, Paoletti E, Moss B (1975) Purification of mRNA guanylyltransferase and mRNA (guanine-7)methyltransferase from Vaccinia virions. J Biol Chem 250:9322–9329

Masaki H, Ogawa T (2002) The modes of action of colicins E5 and D, and related cytotoxic tRNases. Biochimie 84:433–438

Masters BS, Stohl LL, Clayton DA (1987) Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. Cell 51:89–99

McCracken DA, Martin VJ, Stark MJ, Bolen PL (1994) The linear-plasmid-encoded toxin produced by the yeast *Pichia acaciae*: characterization and comparison with the toxin of *Kluyveromyces lactis*. Microbiology 140:425–431

McNeel DG, Tamanoi F (1991) Terminal region recognition factor 1, a DNA-binding protein recognizing the inverted terminal repeats of the pGKl linear DNA plasmids. Proc Natl Acad Sci USA 88:11398–11402

Mehlgarten C, Schaffrath R (2003) Mutant casein kinase I (Hrr25p/Kti14p) abrogates the G1 cell cycle arrest induced by *Kluyveromyces lactis* zymocin in budding yeast. Mol Genet Genomics 269:188–196

Mehlgarten C, Schaffrath R (2004) After chitin docking, toxicity of *Kluyveromyces lactis* zymocin requires *Saccharomyces cerevisiae* plasma membrane H^+-ATPase. Cell Microbiol 6:569–580

Meijer WJ, Horcajadas JA, Salas M (2001) Phi29 family of phages. Microbiol Mol Biol Rev 65:261–287

Meinhardt F, Rohe R (1993) Extranuclear inheritance: linear protein-primed replicating genomes in plants and microorganisms. In: Esser K, Lüttge U, Kadereit JW, Beyschlag W (eds) Progress in botany, vol 54. Springer, Heidelberg, pp 51–70

Meinhardt F, Schaffrath R (2001) Extranuclear inheritance: cytoplasmic linear double-stranded DNA killer elements of the dairy yeast *Kluyveromyces lactis*. In: Esser K, Lüttge U, Kadereit JW, Beyschlag W (eds) Progress in botany, vol 62. Springer, Heidelberg, pp 51–70

Meinhardt F, Kempken F, Esser K (1986) Proteins are attached to the ends of a linear plasmid in the filamentous fungus *Ascobolus immersus*. Curr Genet 11:243–246

Meinhardt F, Kempken F, Kamper J, Esser K (1990) Linear plasmids among eukaryotes: fundamentals and application. Curr Genet 17:89–95

Mendez J, Blanco L, Esteban JA, Bernad A, Salas M (1992) Initiation of phi29 DNA replication occurs at the second 3′ nucleotide of the linear template: a sliding-back mechanism for protein-primed DNA replication. Proc Natl Acad Sci USA 89(20):9579–9583

Mendez J, Blanco L, Salas M (1997) Protein-primed DNA replication: a transition between two modes of priming by a unique DNA polymerase. EMBO J 16:2519–27

Mogen KL, Siegel MR, Schardl CL (1991) Linear DNA plasmids of the perennial ryegrass choke pathogen, *Epichloe typhina* (Clavicipitaceae). Curr Genet 20:519–526

Myers CJ, Griffiths AJ, Bertrand H (1989) Linear kalilo DNA is a *Neurospora* mitochondrial plasmid that integrates into the mitochondrial DNA. Mol Gen Genet 220:113–120

Niwa O, Sakaguchi K, Gunge N (1981) Curing of the killer deoxyribonucleic acid plasmids of *Kluyveromyces lactis*. J Bacteriol 148:988–990

Oeser B (1988) S2 plasmid from *Zea mays* encodes a specific RNA polymerase: an alternative alignment. Nucleic Acids Res 16:8729

Oeser B, Tudzynski P (1989) The linear mitochondrial plasmid pClK1 of the phytopathogenic fungus *Claviceps purpurea* may code for a DNA polymerase and an RNA polymerase. Mol Gen Genet 217:132–140

Oeser B, Rogmann-Backwinkel P, Tudzynski P (1993) Interaction between mitochondrial DNA and mitochondrial plasmids in *Claviceps purpure*: analysis of plasmid-homologous sequences upstream of the lrRNA gene. Curr Genet 23:315–322

Ogawa T, Tomita K, Ueda T, Watanabe K, Uozumi T, Masaki H (1999) A cytotoxic ribonuclease targeting specific transfer RNA anticodons. Science 283:2097–2100

Otero G, Fellows J, Li Y, de Bizemont T, Dirac AM, Gustafsson CM, Erdjument-Bromage H, Tempst P, Svejstrup JQ (1999) Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. Mol Cell 3:109–118

Paillard M, Sederoff RR, Levings CS (1985) Nucleotide sequence of the S-1 mitochondrial DNA from the S cytoplasm of maize. EMBO J 4:1125–1128

Pontig CP, Aravind L, Schultz J, Bork P, Koonin EV (1999) Eukaryotic signalling domain homologues in archaea and bacteria: ancient ancestry and horizontal gene transfer. J Mol Biol 289:729–745

Pring DR, Levings CS III, Hu WWL, Timothy DH (1977) Unique DNA associated with mitochondria in the "S"-type cytoplasm of male-sterile maize. Proc Natl Acad Sci USA 74:2904–2908

Rahl PB, Chen CZ, Collins RN (2005) Elp1p, the yeast homolog of the FD disease syndrome protein, negatively regulates exocytosis independently of transcriptional elongation. Mol Cell 17:841–853

Raths S, Rohrer J, Crausaz F, Riezman H (1993) *end3* and *end4*: two mutants defective in receptor-mediated and fluid-phase endocytosis in *Saccharomyces cerevisiae*. J Cell Biol 120:55–65

Reiter J, Herker E, Madeo F, Schmitt MJ (2005) Viral killer toxins induce caspase-mediated apoptosis in yeast. J Cell Biol 168:353–358

Rodriguez I, Lazaro JM, Blanco L, Kamtekar S, Berman AJ, Wang J, Steitz TA, Salas M, de Vega M (2005) A specific subdomain in phi29 DNA polymerase confers both processivity and strand-displacement capacity. Proc Natl Acad Sci USA 102:6407–6412

Rohe M, Schrage K, Meinhardt F (1991) The linear plasmid pMC3-2 from *Morchella conica* is structurally related to adenoviruses. Curr Genet 20:527–533

Rohe M, Schründer J, Tudzynski P, Meinhardt F (1992) Phylogenetic relationships of linear protein-primed replicating genomes. Curr Genet 21:173–176

Romanos M, Boyd A (1988) A transcriptional barrier to expression of cloned toxin genes of the linear plasmid k1 of *Kluyveromyces lactis*: evidence that native k1 has novel promoters. Nucleic Acids Res 16:7333–7350

Salas M (1991) Protein priming of DNA replication. Annu Rev Biochem 60:39–71

Samac DA, Leong SA (1988) Two linear plasmids in mitochondria of *Fusarium solani* f. sp. cucurbitae. Plasmid 19:57–67

San Martin C, Burnett RM (2003) Structural studies on adenoviruses. Curr Top Microbiol Immunol 272:57–94

Schaffrath R, Meacock PA (2001) An SSB encoded by and operating on linear killer plasmids from *Kluyveromyces lactis*. Yeast 18:1239–1247

Schaffrath R, Meinhardt F (2004) *Kluyveromyces lactis* zymocin and other plasmid-encoded yeast killer toxins. In: Schmitt M, Schaffrath R (eds) Topics in current genetics, vol 11: microbial protein toxins. Springer, Heidelberg, pp 133–155

Schaffrath R, Stark MJR, Gunge N, Meinhardt F (1992) *Kluyveromyces lactis* killer system: ORF1 of pGKL2 has no function in immunity expression and is dispensable for killer plasmid replication and maintenance. Curr Genet 21:357–363

Schaffrath R, Soond SM, Meacock PA (1995) The DNA and RNA polymerase genes of yeast plasmid pGKL2 are essential loci for plasmid integrity and maintenance. Microbiology 141:2591–2599

Schaffrath R, Meinhardt F, Meacock PA (1996) Yeast killer plasmid pGKL2: molecular analysis of UCS5, a cytoplasmic promoter element essential for ORF5 gene function. Mol Gen Genet 250:286–294

Schaffrath R, Meinhardt F, Meacock PA (1997) ORF7 of yeast plasmid pGKL2: analysis of gene expression in vivo. Curr Genet 31:190–192

Schickel J, Helmig C, Meinhardt F (1996) *Kluyveromyces lactis* killer system: analysis of cytoplasmic promoters of linear plasmids. Nucleic Acids Res 24:1879–1886

Schmitt MJ, Breinig F (2002) The viral killer system in yeast: from molecular biology to application. FEMS Microbiol Rev 26:257–276

Schründer J, Meinhardt F (1995) An extranuclear expression system for analysis of cytoplasmic promoters of yeast linear killer plasmids. Plasmid 33:139–151

Schründer J, Gunge N, Meinhardt F (1996) Extranuclear expression of the bacterial xylose isomerase (*xylA*) and the UDP-glucose dehydrogenase (*hasB*) genes in yeast with *Kluyveromyces lactis* linear killer plasmids as vectors. Curr Microbiol 33:323–330

Schwer B (2001) A new twist on RNA helicases: DExH/D box proteins as RNPases. Nat Struct Biol 8:113–116

Shepherd HS (1992) Linear, non-mitochondrial plasmids of *Alternaria alternata*. Curr Genet 21:169–172

Shuman S, Hurwitz J (1981) Mechanism of mRNA capping by Vaccinia virus guanylyltransferase: characterization of an enzyme-guanylate intermediate. Proc Natl Acad Sci USA 78:187–191

Shuman S, Schwer B (1995) RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyltransferases. Mol Microbiol 17:405–410

Shutt TE, Gray MW (2006) Bacteriophage origins of mitochondrial replication and transcription proteins. Trends Genet 22:90–95

Sor F, Fukuhara H (1985) Structure of a linear plasmid of the yeast *Kluyveromyces lactis*: compact organization of the killer genome. Curr Genet 9:147–155

Stam JC, Kwakman J, Meijer M, Stuitje AR (1986) Efficient isolation of the linear DNA killer plasmid of *Kluyveromyces lactis*: evidence for location and expression in the cytoplasm and characterization of their terminally bound proteins. Nucleic Acids Res 14:6871–6884

Stark MJR, Boyd A (1986) The killer toxin of *Kluyveromyces lactis:* characterization of the toxin subunits and identification of the genes which encode them. EMBO J 5:1995–2002

Stark MJ, Mileham AJ, Romanos MA, Boyd A (1984) Nucleotide sequence and transcription analysis of a linear DNA plasmid associated with the killer character of the yeast *Kluyveromyces lactis*. Nucleic Acids Res 12:6011–6030

Stark MJR, Boyd A, Mileham AJ, Romanos MA (1990) The plasmid encoded killer system of *Kluyveromyces lactis:* a review. Yeast 6:1–29

Sugisaki Y, Gunge N, Sakaguchi K, Yamasaki M, Tamura G (1985) Transfer of DNA killer plasmids from *Kluyveromyces lactis* to *Kluyveromyces fragilis* and *Candida pseudotropicalis*. J Bacteriol 164:1373–1375

Sun TP, Webster RE (1987) Nucleotide sequence of a gene cluster involved in entry of E colicins and single-stranded DNA of infecting filamentous bacteriophages into *Escherichia coli*. J Bacteriol 169:2667–2674

Takeda M, Hiraishi H, Takesako T, Tanase S, Gunge N (1996) The terminal protein of the linear DNA plasmid pGKL2 shares an N-terminal domain with the plasmid encoded DNA polymerase. Yeast 12:241–246

Takita MA, Castilho-Valavicius B (1993) Absence of cell wall chitin in *Saccharomyces cerevisiae* leads to resistance to *Kluyveromyces lactis* killer toxin. Yeast 9:589–598

Tanguy-Rougeau C, Wesolowski-Louvel M, Fukuhara H (1988) The *Kluyveromyces lactis KEX1* gene encodes a subtilisin-type serine proteinase. FEBS Lett 234:464–470

Tanner NK, Linder P (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell 8:251–262

Thuriaux P, Sentenac A (1992) Yeast nuclear RNA polymerases. In: Jones EW, Pringle JR, Broach JR (eds) The molecular and cellular biology of the yeast *Saccharomyces cerevisiae*. Cold Spring Harbor, New York, pp 1–48

Tiggemann M, Jeske S, Larsen M, Meinhardt F (2001) *Kluyveromyces lactis* cytoplasmic plasmid pGKL2: heterologous expression of Orf3p and proof of guanylyltransferase and mRNA triphosphatase activities. Yeast 18:815–825

Tiranti V, Savoia A, Forti F, D'Apolito MF, Centra M, Rocchi M, Zeviani M (1997) Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database. Hum Mol Genet 6:615–625

Tokunaga M, Wada N, Hishinuma F (1987) Expression and identification of immunity determinants on linear DNA killer plasmids pGKL1 and pGKL2 in *Kluyveromyces lactis*. Nucleic Acids Res 15:1031–1046

Tokunaga M, Wada N, Hishinuma F (1988) A novel yeast secretion signal isolated from 28K killer precursor protein encoded on the linear DNA plasmid pGKL1. Nucleic Acids Res 16:7499–7511

Tokunaga M, Kawamura A, Hishinuma F (1989) Expression of pGKL killer 28K subunit in *Saccharomyces cerevisiae*: identification of 28K subunit as a killer protein. Nucleic Acids Res 17:3435–3446

Tokunaga M, Kawamura A, Kitada K, Hishinuma F (1990) Secretion of killer toxin encoded on the linear DNA plasmid pGKL1 from *Saccharomyces cerevisiae*. J Biol Chem 265:17274–17280

Tokunaga M, Kawamura A, Omori A, Hishinuma F (1991) Purification and determination of the NH2-terminal amino acid sequence of mouse alpha-amylase secreted from *Saccharomyces cerevisiae*: correct processing of the secretion signal from pGKL killer 28-kDa precursor protein. Biochim Biophys Acta 1080:135–137

Tokunaga M, Kawamura A, Omori A, Hishinuma F (1992) Structure of yeast pGKL 128-kDa killer-toxin secretion signal sequence. Processing of the 128-kDa killer-toxin-secretion-signal-alpha-amylase fusion protein. Eur J Biochem 203:415–423

Tomita K, Ogawa T, Uozumi T, Watanabe K, Masaki H (2000) A cytotoxic ribonuclease which specifically cleaves four isoaccepting arginine tRNAs at their anticodon loops. Proc Natl Acad Sci USA 97:8278–8283

Tommasino M (1991) Killer system of *Kluyveromyces lactis*: the open reading frame 10 of the pGK12 plasmid encodes a putative DNA binding protein. Yeast 7:245–252

Tommasino M, Ricci S, Galeotti C (1988) Genome organization of the killer plasmid pGKL2 from *Kluyveromyces lactis*. Nucleic Acids Res 16:5863–5978

Tudzynski P, Esser K (1986) Extrachromosomal genetics of *Claviceps purpurea*. II. Plasmids in various wild strains and integrated plasmid sequences in mitochondrial genomic DNA. Curr Genet 10:463–467

Tudzynski P, Düvell A, Esser K (1983) Extrachromosomal genetics of *Claviceps purpurea*. I. Mitochondrial DNA and mitochondrial plasmids. Curr Genet 7:145–150

van der Gaag M, Debets AJ, Osiewacz HD, Hoekstra RF (1998) The dynamics of pAL2-1 homologous linear plasmids in *Podospora anserina*. Mol Gen Genet 258:521–529

Vierula PJ, Cheng CK, Court DA, Humphrey RW, Thomas DY, Bertrand H (1990) The kalilo senescence plasmid of *Neurospora intermedia* has covalently linked 5′ terminal proteins. Curr Genet 17:195–201

Weihe A, Hedtke B, Borner T (1997) Cloning and characterization of a cDNA encoding a bacteriophage-type RNA polymerase from the higher plant *Chenopodium album*. Nucleic Acids Res 25:2319–2325

Wésolowski-Louvel M, Tanguy-Rougeau C, Fukuhara H (1988) A nuclear gene required for the expression of the linear DNA-associated killer system in the yeast *Kluyveromyces lactis*. Yeast 4:71–81

Wesp A, Hicke L, Palecek J, Lombardi R, Aust T, Munn AL, Riezman H (1997) End4p/Sla2p interacts with actin-associated proteins for endocytosis in *Saccharomyces cerevisiae*. Mol Biol Cell 8:2291–2306

White JH, Butler AR, Stark MJR (1989) *Kluyveromyces lactis* toxin does not inhibit yeast adenylyl cylcase. Nature 341:666–668

Wilson DW, Meacock PA (1988) Extranuclear gene expression in yeast: evidence for a plasmid encoded RNA polymerase of unique structure. Nucleic Acids Res 16:8097–8112

Winkler GS, Petrakis TG, Ethelberg S, Tokunaga M, Erdjument-Bromage H, Tempst P, Svejstrup JQ (2001) RNA polymerase II elongator holoenzyme is composed of two discrete subcomplexes. J Biol Chem 276:32743–32749

Wittschieben BO, Otero G, de Bizemont T, Fellows J, Erdjument-Bromage H, Ohba R, Li Y, Allis CD, Tempst P, Svejstrup JQ (1999) A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. Mol Cell 4:123–128

Worsham PL, Bolen PL (1990) Killer toxin production in *Pichia acaciae* is associated with linear DNA plasmids. Curr Genet 18:77–80

Xu Y, Yang S, Turitsa I, Griffiths A (1999) Divergence of a linear and a circular plasmid in disjunct natural isolates of the fungus *Neurospora*. Plasmid 42:115–125

Yang X, Griffiths AJ (1993) Plasmid diversity in senescent and nonsenescent strains of *Neurospora*. Mol Gen Genet 237:177–186

Yuewang W, Yang X, Griffiths AJ (1996) Structure of a *Gelasinospora* linear plasmid closely related to the kalilo plasmid of *Neurospora intermedia*. Curr Genet 29:150–158

Zink S, Mehlgarten C, Kitamoto HK, Nagase J, Jablonowski D, Dickson RC, Stark MJR, Schaffrath R (2005) Mannosyl-diinositolphosphoceramide, the major yeast plasma membrane sphingolipid, governs toxicity of *Kluyveromyces lactis* zymocin. Eukaryot Cell 4:879–889

Microbiol Monogr (7)
F. Meinhardt · R. Klassen: Microbial Linear Plasmids
DOI 10.1007/7171_2007_096/Published online: 26 May 2007

Hairpin Plasmids from the Plant Pathogenic Fungi *Rhizoctonia solani* and *Fusarium oxysporum*

Teruyoshi Hashiba[1] (✉) · Atsushi Nagasaka[2]

[1]Department of Environmental Biotechnology, Graduate School of Agriculture, Tohoku University, 1-1 Tsutsumidori-Amamiyamachi, Aoba-ku, 981-8555 Sendai, Japan
hashiba@rouge.plala.or.jp

[2]National Agricultural Research Center for Tohoku Region, Arai, 960-2156 Fukushima, Japan

1 Introduction . . . 228

2 Detection of the Fungal Hairpin Plasmid . . . 228

3 Incidence and Distribution . . . 229
3.1 Linear Plasmids of *Rhizoctonia solani* . . . 229
3.2 Linear Plasmids of *Fusarium oxysporum* . . . 232

4 Structure . . . 233
4.1 Covalently Closed Hairpin Ends Structure . . . 233
4.2 Clothespin Structure . . . 234

5 Mitochondrial Location of DNA Plasmids . . . 235

6 Chromosome-Associated Plasmid . . . 235

7 Function . . . 236
7.1 Replication . . . 236
7.1.1 Replication of pRS64 Plasmids . . . 236
7.1.2 Replication of pFOXC2 and pFOXC3 Plasmids . . . 237
7.2 Transcription . . . 237
7.3 Translation . . . 238

8 Effects of Plasmids . . . 241

References . . . 242

Abstract This review focuses on the biology of the hairpin plasmids from the plant pathogenic fungi *Rhizoctonia solani* and *Fusarium oxysporum*. *R. solani* linear plasmid is single-stranded with covalently closed hairpin loops at each end. *F. oxysporum* plasmid, however, has a clothespin structure, which includes a terminal hairpin and noncovalently linked ends at the other terminus. We present the nucleotide sequence of the hairpin plasmids, the terminal hairpin structure, an analysis of an open reading frame in the sequence, the unique transcriptional form of this plasmid, and a proposed model for their replication.

Keywords Fungal hairpin plasmid · Hairpin-loop structure · Clothespin structure · Rhizoctonia solani · Fusarium oxysporum · Plant pathogenic fungus

1 Introduction

Linear plasmids are common genetic elements in both prokaryotes and eukaryotes. Most linear plasmids found in eukaryotes are double-stranded DNA molecules, with proteins covalently linked to their 5′ terminus (reviewed by Meinhardt and Rohe 1993; Griffiths 1995; Gunge et al. 1995; Kempken 1995; Meinhardt et al. 1997). Almost all linear plasmids sequenced to date have long, inverted terminal repeats and include a putative DNA polymerase gene. A different type of linear plasmid, the so-called hairpin plasmid with covalently closed ends, has been described in a wide variety of organisms, including the vaccinia virus genome (Baroudy et al. 1982), poxvirus (DeLange et al. 1986), African swine fever virus (Gonzalez et al. 1986), *Chlorella* viruses (Rohozinski et al. 1989), *Escherichia coli* N15 prophage (Rybchin and Svarchevsky 1984), and the linear plasmids of bacteria in the genus *Borrelia burgdorferi* (Barbour and Garon 1987; Hinnebusch and Barbour 1991; Casjens 1999). The Borrelia linear plasmid consists of a polynucleotide chain that is fully base-paired except for short, single-strand hairpin loops at each end (Hinnenbusch and Barbour 1991). The vaccinia virus genome is incompletely base-paired, with hairpin-loop structures that exist in isomeric forms that are inverted and complementary in sequence ("flip-flopped" structures) (Baroudy et al. 1982).

Hairpin plasmids have been found in a limited number of eukaryotes. The fungal plant pathogen *Rhizoctonia solani* linear plasmid is single-stranded with hairpin loops at each end (Hashiba et al. 1984; Miyashita et al. 1990). They do not undergo flip-flop inversion, unlike the vaccinia virus DNA. The pFOXC2 and pFOXC3 of the fungal plant pathogen *Fusarium oxysporum*, however, have a "clothespin" genomic structure, which includes a terminal hairpin and noncovalently linked ends at the other terminus (Walter and Kennel 1999; Simpson et al. 2004). This review focuses on the biology of linear DNA plasmids of the pathogenic fungi *R. solani* and *F. oxysporum*, which have unique terminal structures.

2 Detection of the Fungal Hairpin Plasmid

Cultures of *R. solani* were grown without shaking at 26 °C in a liquid potato/dextrose broth (Difco) containing 1% polypeptone (Hashiba et al. 1984). After 5 to 7 days, mycelia were harvested on filter paper and lyophilized. Plasmid DNA was detected using the partially modified procedure of Hirt (1967) (Chen et al. 1992).

3
Incidence and Distribution

3.1
Linear Plasmids of *Rhizoctonia solani*

The plant pathogenic fungus *R. solani* has a wide range of hosts and is one of the most serious fungal pathogens in the world. Fungal isolates were assigned to incompatibility groups based on their affinities for hyphal fusion with members of designated anastomosis groups (AGs). Japanese isolates were divided into ten AGs (Anderson 1982; Ogoshi 1987; Carling et al. 2002). Of these, AG1, AG2-1, AG2-2, AG3, and AG4 correspond to sasakii type (IA) and web-blight type (IB), winter crop type (II), rush type (IIIB), root rot type (IV) and potato type (IV), and praticola type (IIIA), respectively, based on their pathogenicity to field crops and growth characteristics in culture (Watanabe and Matsuda 1966).

A survey of 114 isolates of *R. solani* from a wide range of plant hosts revealed that 48 isolates contained one to three plasmids, ranging in size from 2.2 to 7.0 kb (Hashiba et al. 1984, 1987a, 1987b; Shimma et al. 1988; Miyasaka et al. 1990; Chen et al. 1992). We examined the sequence homology among DNA plasmids in representative isolates using Southern blot analysis with nick-translated DNA plasmids as probes (Miyasaka et al. 1990; Hashiba et al. 2001; Chen et al. 1992). Considerable sequence homology was detected among DNA plasmids obtained from isolates within the same AG, but no relationship was detected among DNA plasmids obtained from isolates belonging to different AGs, emphasizing the fact that AGs are genetically distinct (Table 1).

Jabaji-Hare et al. (1994) examined 54 isolates from five AGs in Canada, and found a 2.4-kb plasmid (pRS104) in AG4 isolates and a 3.6-kb plasmid (pRS188) in AG5 isolates. DNA plasmids in *R. solani* were identified by Rubio et al. (1996). They surveyed 15 AG4 isolates, 13 from Spain and two from Israel; 13 of these isolates contained a 2.6-kb DNA plasmid (pMe8-2). As expected, there was no sequence homology between the pMe8-2 plasmid (AG4, Spain) and the pRS188 plasmid (AG5, Canada). These findings are consistent with the results obtained with Japanese isolates by Miyasaka et al. (1990).

In addition, Northern blot analysis of total RNA obtained from six AGs of *R. solani* using cloned fragments revealed the presence of RNAs hybridizing to the DNA plasmids in various *R. solani* isolates. The RNAs hybridizing to DNA plasmids represent the transcription products of the plasmid DNA and indicate that the genetic information in the DNA plasmid is being expressed. There is considerable sequence homology among plasmid-encoded transcripts obtained from isolates of the same AG, but no sequence homology among transcripts obtained from isolates of different AGs (Table 2). These results suggested that the distribution of DNA plasmids correlates with the

Table 1 Southern hybridization analysis of DNA plasmid from *R. solani* AG1 to AG6 isolates probed with DNA plasmid from a representative isolate of each AG or ISG

AG	Pathological type	Isolate	Size (kb)	Source of probe: C-527 (AG1)	C-94 (AG2-2 IIIB)	H-16 (AG2-2 IV)	P-7-1 (AG3)	RI-64 (AG4)	ST-8 (AG5)	HI-5-16 (AG6)
1	Web-blight type (IB)	C-527	5.5	**P**[a]						
2-2	Rush type (IIIB)	C-6	3.2	–[b]	+[c]	–	–	–	–	–
		C-94	4.0	–	–	–	–	–	–	–
			3.0	–	**P**	–	–	–	–	–
		C-125	4.3	–	–	–	–	–	–	–
			2.9	–	+	–	–	–	–	–
		C-328	4.2	–	–	–	–	–	–	–
			2.9	–	+	–	–	–	–	–
		C-329	4.7	–	–	–	–	–	–	–
			2.9	–	+	–	–	–	–	–
2-2	Root rot type (IV)	A-14-1	4.7	–	–	+	–	–	–	–
		C-616	3.0	–	–	+	–	–	–	–
		C-623	5.5	–	–	–	–	–	–	–
			2.9	–	–	+	–	–	–	–
		H-16	5.4	–	–	+	–	–	–	–
			4.7	–	–	**P**	–	–	–	–
		SH-17	5.0	–	–	+	–	–	–	–
			4.2	–	–	+	–	–	–	–
3	Potato type (IV)	P-5	6.2	–	–	–	+	–	–	–
		P-7-1	6.2	–	–	–	**P**	–	–	–
		ST-3	6.2	–	–	–	+	–	–	–
4	Praticola type (IIIA)	GM-7	2.7	–	–	–	–	+	–	–
		GM-11	2.7	–	–	–	–	+	–	–
		RI-64	2.7	–	–	–	–	**P**	–	–
5		ST-2-1	3.2	–	–	–	–	–	+	–
		ST-8	3.2	–	–	–	–	–	**P**	–
		W-3	3.2	–	–	–	–	–	+	–
		501	3.2	–	–	–	–	–	+	–
6		HI-5-16	5.2	–	–	–	–	–	–	**P**

[a] The entire length of a DNA plasmid obtained from the representative isolates of one AG or ISG was used as the hybridization probe
[b] No sequence homology was observed
[c] Considerable sequence homology was observed

AGs and intraspecific groups (ISGs) of *R. solani* and might be used to classify this fungus. The relationship between the DNA plasmids and AGs or ISGs is consistent with the results of DNA base sequence homology (Kuninaga and Yokosawa 1985), serologic comparison (Adams and Butler 1979), DNA base comparison (Kuninag and Yokosawa 1980), ecologic and morphologic characteristics (Ogoshi 1976), and electrophoresis of isoenzymes (Matsuyama et al. 1987) and soluble proteins (Reynolds et al. 1983). Thus, these DNA plasmids are host-specific.

Table 2 Northern hybridization analysis of total RNAs from *R. solani* AG1 to AG6 isolates probed with the cloned fragment of plasmid DNA from a representative isolate of each AG or ISG

AG	Isolate	Source of probe[a] C-527[b] (AG1)	C-330 (AG2-2 IIIB)	H-16 (AG2-2 III)	R-468 (AG3)	RI-64 (AG4)	ST-8 (AG5)	HI-5-11 (AG6)
1 IB	C-527	+[c]	–	–	–	–	–	–
2-2 IIIB	C-113	–[d]	+	–	–	–	–	–
	C-125	–	+	–	–	–	–	–
	C-330	–	+	–	–	–	–	–
2-2 IV	A-14-1	–	–	+	–	–	–	–
	H-16	–	–	+	–	–	–	–
	SH-17	–	–	+	–	–	–	–
3	NR-3	–	–	–	+	–	–	–
	P-7-1	–	–	–	+	–	–	–
	R-468	–	–	–	+	–	–	–
4	GM-11	–	–	–	–	+	–	–
	R101	–	–	–	–	+	–	–
	RI-64	–	–	–	–	+	–	–
5	ST2-1	–	–	–	–	–	+	–
	ST-8	–	–	–	–	–	+	–
	W-3	–	–	–	–	–	+	–
6	HI-5-11	–	–	–	–	–	–	+

[a] The cloned fragment of the DNA plasmid obtained from the isolate indicated was used as the source of hybridization probe
[b] Isolate
[c] The probe hybridized to one or two transcripts in total RNA of the isolate indicated
[d] The probe did not hybridize to total RNA of the isolate indicated

3.2
Linear Plasmids of *Fusarium oxysporum*

Isolates of *F. oxysporum* that cause disease in crucifers are categorized based on their relative ability to cause disease in a particular host species. Isolates causing the cabbage yellows disease, sweet potato stem rot disease, flax wilt disease, and the edible burdock wilt disease are defined as *F. oxysporum* formae speciales (f.sp.) *conglutinans* (Snyder and Hansen 1940), f.sp. *lini*, and f.sp. *arctii*, respectively, while isolates pathogenic to the radish, ornamental crucifer stock, and eggplant are called *F. oxysporum* f.sp. *raphani*, f.sp. *matthioli*, and f.sp. *melongenae*, respectively (Bosland and William 1987; Kendrick and Snyder 1942).

Three 1.9-kb DNA plasmids designated pFOXC1, pFOXC2, and pFOXC3, are found in *F. oxysporum* f.sp. *conglutinans*, f.sp. *raphani*, and f.sp. *matthioli*, respectively (Kistler et al. 1987; Kistler and Leong 1986). The plasmids have different restriction fragment length polymorphism maps, and while pFOXC2 and pFOXC3 cross-hybridize, Southern hybridization analyses suggest that pFOXC1 and pFOXC2 have no sequence similarity (Kistler et al. 1987; Kistler and Leong 1986). A survey of 61 field isolates of *F. oxysporum* from Japan revealed six linear plasmids (Hirota et al. 1992). All examined isolates of *F. oxysporum* f.sp. *arctii*, f.sp. *batatas*, f.sp. *conglutinans*, f.sp. *lini*, f.sp. *melongenae*, and f.sp. *raphani* contain mitochondrial plasmids, termed pFOA, pFOB, pFOC, pFOL, pFOM, and pFOR, respectively (Hirota et al. 1992). pFOC, pFOL, and pFOA have considerable sequence homology among each other, but pFOR, pFOM, and pFOB have no sequence similarity. Both pFOC and pFOXC1 obtained from f.sp. *conglutinans* have essentially the same restriction endonuclease map, and both pFOR and pFOXC2 obtained from f.sp. *raphani* had sequence similarity (Fig. 1) (Kistler and Leong 1986; Kistler et al. 1987; Hirota et al. 1992).

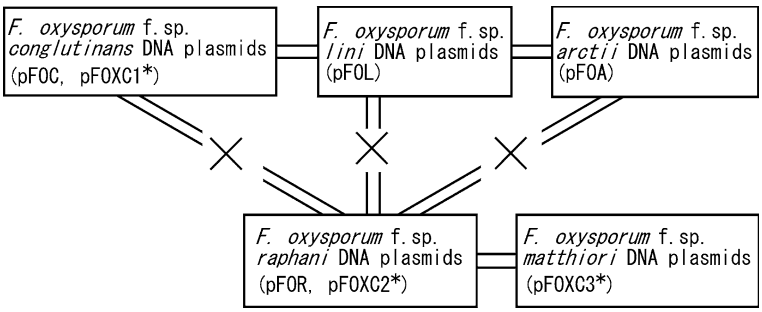


Fig. 1 Relatedness of 1.9-kb DNA plasmids found in various *F. oxysporum* formae speciales; =, considerable sequence homology was observed, = × =, no sequence homology was observed. Plasmids with an asterisk were found by Kistler and Leong (1986) and Kistler et al. (1987)

4
Structure

4.1
Covalently Closed Hairpin Ends Structure

Three linear DNA plasmids were isolated in RI-64 of AG4 of *R. solani*. These plasmids, designated pRS64-1, -2, and -3, had the same size (2.7 kb). Restriction mapping and Southern hybridization analysis of pRS64-1, -2, and -3 revealed the presence of homologous regions. The DNA plasmids were resistant to both 3′-exonuclease and 5′-exonuclease, even after treatment with proteinase K or alkali. The length of both terminal fragments generated by restriction endonuclease digestion was doubled under denaturation conditions, indicating that the linear DNA plasmids have hairpin loops at both termini (Miyashita et al. 1990; Hongo et al. 1994). A computer-based study of the pRS64-2 DNA folding at both termini predicted hairpin loop structures. The hairpin loops of the pRS64 DNA plasmid consist of the left- and right-hand terminal 113 and 105 nucleotides, and are therefore similar in length to the 104-nucleotide loop of vaccinia virus (Baroudy et al. 1982). Unlike the vaccinia virus DNA, however, they are not flip-flopped. The pRS64 plasmid hairpins form a cruciform base-paired structure (Katsura et al. 1997). They are not inverted and complementary (flip-flopped).

The linear DNA plasmid (pRS224-1) from isolate H-16 of AG2-2 of *R. solani* consists of 4986 nucleotides. A computer-based study of the folding of pRS224-1 at both termini predicted hairpin-loop structures. The hairpin

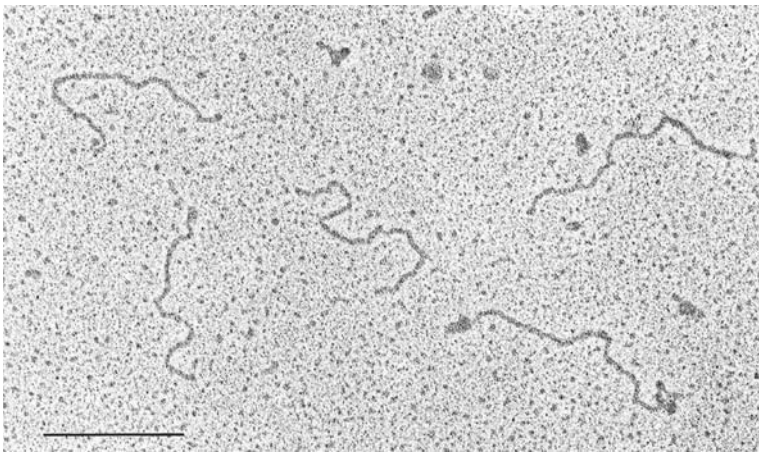


Fig. 2 Electron micrograph of DNA molecules from the isolate RI-64 of *R. solani* by the formamide and cytochrome *c* technique. The *scale bar* represents 1 kb

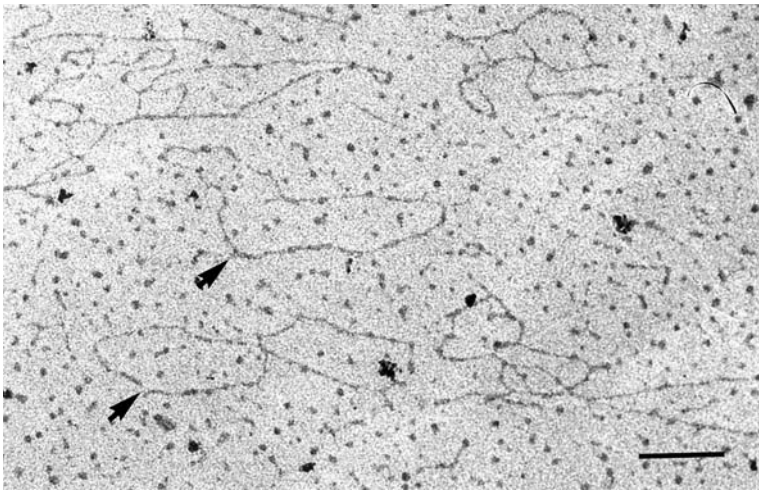


Fig. 3 Structure of plasmid DNA. The plasmid forms a single-stranded circle when completely denatured. *Arrows* show single-stranded circles that were produced. The *scale bar* indicates 1000 nucleotides

loops consisted of the left and right termini of 236 and 264 nucleotides, respectively, and share no sequence homology (Katsura et al. 2001).

Jabaji-Hare et al. (1994) also reported extrachromosomal DNA elements in field isolates of *R. solani* belonging to five AGs (1 to 5), and revealed a 2.4-kb plasmid (pRS104) in AG4 isolates and a 3.6-kb plasmid (pRS188) in AG5 isolates. Both plasmids were analyzed in detail and were determined to be linear with a knob structure at both termini, the same as the plasmids described by Miyashita et al. (1990).

Electron microscopy revealed that the pRS64 and pRS224 plasmid fractions purified from gels consisted of homogeneous populations of linear DNA (Fig. 2) (Miyasaka et al. 1990; Hashiba et al. 2001). The plasmid forms a single-stranded circle when completely denatured (Fig. 3). There were single-stranded DNA circles with a circumference twice the length of the duplex linear DNA (Katsura et al. 2001).

4.2 Clothespin Structure

Restriction digest analysis of pFOXC2 and pFOXC3 DNAs indicates that the plasmids have a "clothespin" structure located upstream of the plasmid open reading frame (ORF) and noncovalently linked ends with two to five copies of a pentameric repeat (5′-ATCTA) at the other terminus (Walther and Kennell 1999). The ability to cut pFOXC2 and pFOXC3 DNAs with restriction endonucleases that require duplexed DNA substrates, together with observa-

tions of the plasmid's structure by electron microscopy (Kistler and Leong 1986), indicates that the plasmids are double-stranded molecules. The (+) or coding strand terminus downstream of the plasmid ORF has an accessible 3′-hydroxyl (inferred by its ability to serve as a substrate for terminal deoxynucleotide transferase and sensitivity to 3′→5′ exonuclease III), whereas the 5′ end of the opposing (–) strand appears to be modified in a manner that results in resistance to 5′→3′ λ-exonuclease and the inability to be end-labeled with [γ^{32}P]ATP (Walther and Kennell 1999).

5
Mitochondrial Location of DNA Plasmids

Most of the linear DNA plasmids of filamentous fungi are thought to reside within mitochondria (Samac and Leong 1989). To investigate the location of DNA plasmids of *R. solani* AG4 and AG2-2, total DNAs or DNA isolated from intact mitochondria were separated by CsCl-bisbenzimide centrifugation. Plasmid bands were detected only in the mitochondrial DNA fractions. The plasmids were resistant to DNase treatment of intact mitochondria, indicating that they were located in the mitochondria. DNA plasmids (pRS64 and pRS224) from *R. solani* are associated with mitochondria (Chen et al. 1992; Kuradate et al. 1996; Wako et al. 1991). Jabaji-Hare et al. (1994) also demonstrated that DNA plasmids (pRS104 and pRS188) from *R. solani* are associated with mitochondria in cellular fractionation experiments.

Mitochondrial DNAs fractionated on CsCl-bisbenzimide density gradients were examined. Linear plasmids were found in the mitochondria in all examined isolates of *F. oxysporum* f.sp. *conglutinans*, f.sp. *raphani*, f.sp. *matthioli*, f.sp. *arctii*, f.sp. *melongenae*, f.sp. *lini*, and f.sp. *batatas* from a global collection (Hirota et al. 1992; Kistler et al. 1987; Kistler and Leong 1986). Our preliminary work indicated that the plasmids were resistant to DNase treatment of intact mitochondria, indicating their mitochondrial location.

6
Chromosome-Associated Plasmid

To gain information on the chromosomal DNA of isolates belonging to AG4 of *R. solani*, we used the orthogonal-field-alternation gel electrophoresis (OFAGE) technique to separate large DNAs from the filamentous fungus. At least six chromosome-sized DNA bands from isolates RI-64 and 1271 of *R. solani* were separated using OFAGE. Thus, the total genome size of *R. solani* is estimated to be more than 11 Mb. The assignment of linkage groups to the resolved chromosomal DNA bands might be accomplished using linkage group-specific probes.

Fungal plasmids share homology with either nuclear DNA (Wright and Cummings 1983) or mitochondrial DNA, or both (Van den Boogert et al. 1982; Samac and Leong 1989). Southern hybridization analysis with nearly full-length clones of pRS64-1 and -2 (excluding the terminal full regions) identified a homologous sequence within a small chromosome-sized DNA band of approximately 1.1 Mb. Sequence homology to pRS64 was also maintained within the chromosomal DNA of isolate 1271 of AG4, which does not possess the plasmid. The plasmid showed no homology to the mitochondrial DNA of isolate 1271 (Wako et al. 1991). This finding suggests that pRS64 are stably integrated into the chromosomal DNA. Study of the plasmid sequence will help to identify their integrated forms in the fungal DNA and reveal how integration over the eons has impacted the evolution of the mitochondrial and nuclear genetics of *Rhizoctonia*.

7
Function

7.1
Replication

7.1.1
Replication of pRS64 Plasmids

We propose a possible model for the replication of pRS64, in which hairpins and terminal homologous regions have key roles. This model is modified from the model of vaccinia virus DNA replication (Baroudy et al. 1983). Replication of pRS64 DNAs might be initiated by site-specific nicking proximal to the loop structure and within the terminal homologous region. This results in the formation of a free 3′-hydroxyl end that can serve as a primer for DNA replication. If DNA replication proceeds from both ends simultaneously, mature molecules can be formed directly (Fig. 4, right). If DNA synthesis is initiated from a single end, replication will proceed around the opposite hairpin, leading to the formation of dimers (Fig. 4, left and center); one is a head-to-head dimer and the other is a tail-to-tail dimer (Miyashita et al. 1990). The finding of dimeric forms of pRS64-1 suggests that the replication of this plasmid is initiated from a single end.

The structures of the two dimer types shown in Fig. 4 are consistent with the model (Fig. 4, left and center); however, some molecules might replicate directly to monomers. In contrast, dimeric forms of pRS64-2 and -3 were not detected. Replication of these plasmid DNAs might be initiated from both ends. Another possibility is that these dimers are resolved into monomers too quickly to be detected (Miyashita et al. 1990). A terminal inverted repeat was found in vaccinia virus, and is thought to function as a recognition

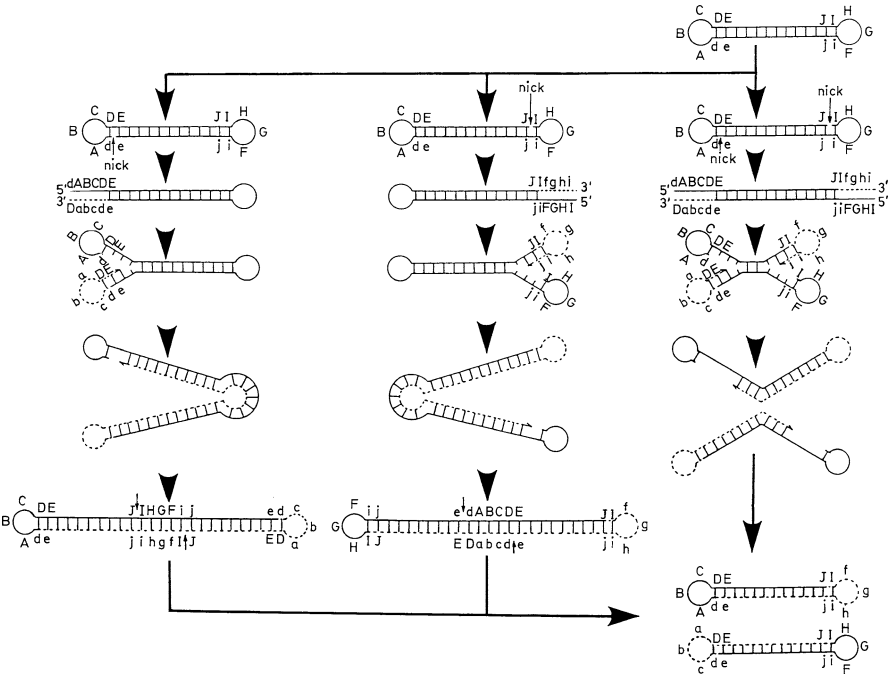


Fig. 4 Possible models for replication of pRS64 DNA plasmids. Complementary sequences are indicated with *capital and lowercase letters*. The terminal hairpins are drawn as loops and designated ABC and FGH. The terminal homologous regions are designated DE and IJ

sequence for a single species of nicking enzyme. In pRS64, two different nicking enzymes might be involved in replication, although we cannot rule out the possibility of a short inverted repeat sufficient for recognition by a single species of nicking enzyme.

7.1.2
Replication of pFOXC2 and pFOXC3 Plasmids

A hypothetical model for the replication of pFOXC2 and pFOXC3 plasmids was constructed (Walther and Kennell 1999; Simpson et al. 2004; Galligan and Kennell 2007, this volume).

7.2
Transcription

Unique poly(A)$^-$RNA, 0.5 kb in length and hybridizable with the pRS64 DNAs, was found in mycelial cells of the isolates RI-64 (containing all three plasmids), R101 (containing plasmid pRS64-1), and GM-11 (containing

pRS64-2) (Chen et al. 1992). Comparison of the nucleotide sequence of the cDNA derived from poly(A)⁻RNA with those of pRS64-1, -2, and -3 showed 100, 73, and 84% homology, respectively. This suggests that the poly(A)⁻RNA hybridizing to pRS64 represents the transcription products of pRS64-1, -2, and -3, and that the genetic information in pRS64 DNAs is expressed (Hongo et al. 1994).

Unique poly(A)⁻RNA, 4.7 and 7.4 kb in length and hybridizing with the pRS224 DNA, are also found in the plasmid-containing *R. solani* H-16 isolate (Katsura et al. 2001). Transcript product-mapping allowed for the prediction of the locations of different expression signals. Of the two transcripts, one was approximately 2.7 kb longer than the 4.7-kb transcript. Transcripts longer than plasmids have also been detected from *R. solani* AG5 and AG6 (Kuradate et al. 1996). The striking finding in this study is that the 7.4-kb transcript is RNA containing the full-length transcript, which is generated from the left terminal region of the complementary strand via a full-length sense strand to the right terminal region of the complementary strand. In contrast, the 4.7-kb transcript is predicted to be transcribed from the center of the sense strand to the right terminal region of the complementary strand (Fig. 5).

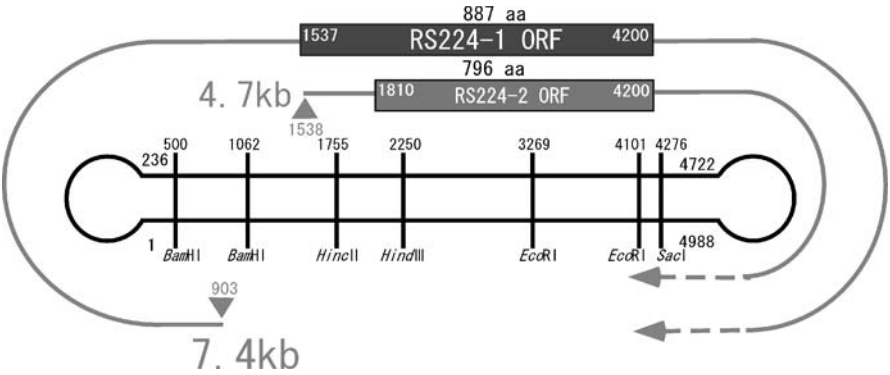


Fig. 5 Mapping of pRS224-1 transcripts. The 7.4- and 4.7-kb RNAs are indicated with *arrows*. The *black box* indicates the ORF. *Letters* represent restriction sites

7.3 Translation

The ORFs found in pRS64-1, -2, and -3 (ORF1-1, ORF2-1, and ORF3-1) encode predicted peptides that are 68 amino acids long. The deduced amino acid sequences had no significant homology with known proteins. Extracts from *E. coli* cells expressing ORF1-1 gave a specific 7-kDa protein, and anti-ORF1-1-encoded protein (RS64) antibodies raised against the synthetic fusion peptide cross-reacted with the specific 7-kDa protein from the mycelia

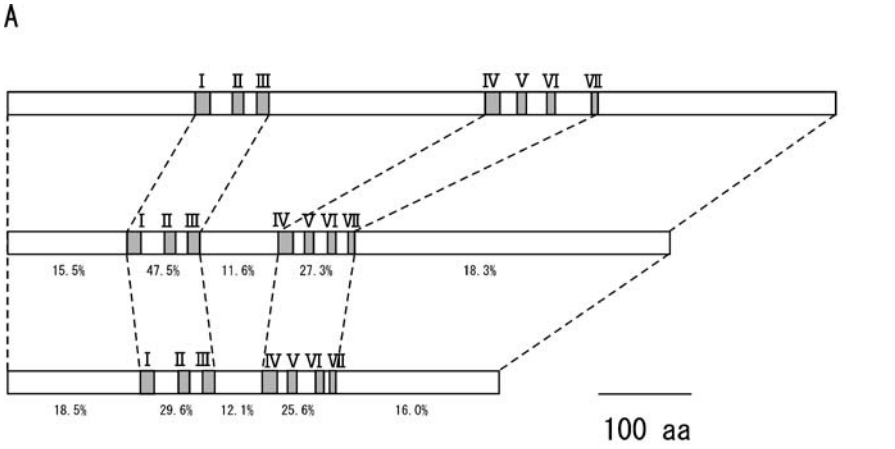
(Hongo et al. 1994). The other ORFs were not considered to be genes because the transcripts corresponding to these ORFs were not detected.

The complete nucleotide sequence of the hairpin-loop linear DNA plasmid (pRS224-1) from the plant pathogenic fungus *R. solani* AG2-2 isolate H-16 was determined, and its unique RNA transcripts were characterized. One ORF (RS224ORF) found in pRS224-1 encoded a peptide predicted to consist of 887 amino acids with an approximate molecular mass of 102 kDa (Katsura et al. 2001).

To determine whether the ORF of pRS224-1 is translated, we constructed pET11a-RS224orf carrying both the T7.Tag gene and the RS224ORF. Transformed *E. coli* cells were induced with IPTG, and the total extract was fractionated using sodium dodecyl sulfate–polyacrylamide gel electrophoresis. The fusion protein (T7.Tag-RS224orf, 92 kDa) was obtained in cells containing the pET11a-RS224orf. We detected a single protein with a molecular mass of approximately 92 kDa, a size slightly smaller than the 102 kDa calculated from the amino acid sequence coded by the ORF (Nagasaka et al. 2003).

Immunoblotting experiment results demonstrated that the antiserum raised against the RS224ORF recognized a specific polypeptide in the total protein extracts of mycelial cells. The putative polypeptide encoded by the ORF had no significant homology with polypeptides listed in various databases. The 92-kDa protein was associated with the mitochondrial fraction, as was the pRS224-1 DNA of *R. solani* which is also located in the mitochondria (Nagasaka et al. 2003). DNA plasmid pRS64 of *R. solani* AG4 is also associated with mitochondria (Chen et al. 1992), but the 7-kDa protein was detected in the cell wall fraction. A computer search using the DNASIS program revealed no significant homology between RS64 and RS224. In addition, RS64 and RS224 were detected in the cell wall and mitochondrial fractions, respectively. Thus, RS224 are clearly different proteins with different functions.

The pRS224-1 ORF contains seven conserved domains characteristic of a reverse transcriptase (RT) (Fig. 6) (Katsura et al. 2001), including the highly-conserved YXDD (Try-xxx-asp-asp) sequence that is proposed to be part of the active site of RT (Varmus and Brown 1989; Xiong and Eickbush 1990). Pair-wise comparison of identical amino acids within the seven consensus sequences that were common to the RT sequence and included the highly conserved YXDD sequence had 57.8% identity to the Mauriceville retroplasmid of *Neurospora crassa* and 54.3% identity to the pFOXC2 of *F. oxysporum* (Fig. 6). We identified two regions between domains I and III, and IV and VII, of the amino acid domains with more than 25.6% amino acid sequence homology (Fig. 6a, boxes with a solid line) and an intervening region with less than 12.1% amino acid sequence homology. RT activity was assayed by the incorporation of [α-^{32}P]dTTP into trichloroacetic acid-precipitable material (Farmerie et al. 1987); however, putative RT activity was not detected in mitochondrial lysates from *R. solani* isolate H-16.



B

	Region I	Region II	Region III
pRS224	201 YIPKPGTNKVRPLGVP	241 QHGFVPRRGTLT	267 ILEVDFKGFFPSV
Mauric	128 YIPK-ANGKQRPLGVP	168 QHAYFPKRGVFT	193 IYEFDLKNFFPSV
pFOXC2	142 YILK-KNGKMRPIGAP	183 QHGYMKKRGAWS	209 GYEFDLKSFFNTV

	Region VI	Region V	Region VI	Region VII
pRS224	512 QFGLPQGGPLSPFLST	546 CLFYADDGFF	578 GIQISFEKS	626 KFLGLEY
Mauric	290 TNGVPQGGSTSCGLAT	318 LIMYADDGIL	343 GVVQEPAKS	365 KFLGLEF
pFOXC2	273 RTGVPQGLSLSPLLST	300 LIMYADDGIY	330 GIEISPEKS	345 KFCGVTI

Fig. 6 Multiple alignment of the coding regions of authentic or presumed RTs from fungal mitochondrial plasmids. **a** The derived amino acid sequences of pRS224 of *R. solani* (pRS224), pFOXC2 of *F. oxysporum* (pFOXC2; GenBank accession no. AF005240), and Mauriceville retroplasmid of *Neurospora crassa* (Mauric; accession no. KO3295) were compared. Homologous regions which occur at the same location in three DNA plasmids are shown as *boxes with a solid line*. The numbers below each ORF indicate the percentage amino acid identity with amino acids in one box of the pRS224 ORF. Seven conserved regions common to RT are shown as *gray boxes*. The *bar* indicates 100 amino acids. **b** Identical amino acids within the seven conserved regions common to RT sequences are shown in *blocks*. The YXDD motif of the RT sequence is *underlined*. Amino acid positions are indicated by the *numbers* used in the EMBL and GenBank accessions

In contrast, two linear mitochondrial plasmids (pFOXC2, pFOXC3) of the plant pathogenic fungus *F. oxysporum* contain a single, long ORF of 527 amino acids that is highly conserved between the two plasmids (Kistler et al.

1997; Walther and Kennell 1999). The circular mitochondrial retroplasmids harbored in the Mauriceville and Varkud strains of *Neurospora* spp. contain an ORF of 710 amino acids (Nargang et al. 1984; Akins et al. 1988). This ORF contains conserved amino acid sequence blocks (I–VII) characteristic of retroviral and other RTs; in addition, pFOXC2, pFOXC3, and the circular mitochondrial retroplasmids encode 62- and 81-kDa proteins, respectively, having RT activity (Nargang et al. 1984; Akins et al. 1988; Kuiper et al. 1990; Walther and Kennell 1999).

The most noticeable difference between the RT encoded by the *R. solani* plasmids and the RT encoded by the two linear mitochondrial plasmids (pFOXC2, pFOXC3) is that it is considerably larger (887 versus 527 amino acids). The size of the region between domains III and IV of the amino acid (domains conserved among RT in the largest ORF) is inordinately large. A 232-amino acid gap in the middle of the ORF was not homologous to any known RT (Nargang et al. 1984; Akins et al. 1988; Kistler et al. 1997; Walther and Kennell 1999).

The plasmids pRS224 and pFOXC are classified as retroplasmids because they each encode an RT and replicate via an RNA intermediate (Kistler et al. 1997; Walther and Kennell 1999). To date, retroplasmids have been found only in mitochondria of six filamentous fungi species: *F. oxysporum* (Walther and Kennell 1999), *R. solani* (Katsura et al. 2001; Miyasaka et al. 1992), *Epichloë typhina* (Mogen et al. 1991), *Neurospora* spp. (Akins et al. 1988), *N. crassa* (Collins et al. 1981), and *Trichoderma harzianum* (Antal et al. 2002). Putative RT activity, however, was not present in mitochondrial lysates of *R. solani* (Nagasaka et al. 2003).

8 Effects of Plasmids

The *R. solani* isolate RI-64 carrying plasmids grew slowly, was weakly pathogenic, produced oxalic acid, and was unable to form sclerotia. It is unlikely, however, that the 2.7-kb plasmids carry four genes that control these four phenotypes independently. To obtain isolates that have lost plasmids, or isolates altered in some characteristics associated with the plasmids, protoplasts of isolate RI-64 were regenerated and characterized. Of 205 isolates from the protoplasts, five showed better growth and formed sclerotia, but carried plasmids (Shimma et al. 1988). When a larger number of isolates was examined, there was no correlation between plasmid and hypovirulence (Miyasaka et al. 1990). Jabaji-Hare et al. (1994) compared the degree of pathogenicity of plasmid-carrying and plasmid-free isolates of *R. solani* and demonstrated that the plasmids do not have a role in pathogenicity.

The formae speciales of *F. oxysporum* that harbor the pFOXC plasmids are classified by their ability to cause disease in specific crucifers. Yet, the

isolation of plasmid-free strains and transmission of pFOXC1 and pFOXC2 into different *F. oxysporum* backgrounds via protoplast fusion experiments demonstrated that the plasmids are not determinants of the host range or pathogenicity (Momol and Kistler 1992).

References

Akins RA, Grant DM, Stohl LL, Bottodff DA, Nargang FE, Lambowicz AM (1988) Nucleotide sequence of the Varkud mitochondrial plasmid of *Neurospora* and synthesis of a hybrid transcript with a 5′ leader derivative from mitochondorial RNA. J Mol Biol 204:1–25

Anderson NA (1982) The genetics and pathology of *Rhizoctonia solani*. Annu Rev Phytopathol 20:329–347

Antal GC Jr, Butler EE (1979) Serological relationships among anastomosis groups of *Rhizoctonia solani*. Phytopathology 69:629–633

Antal Z, Manczinger L, Kredcs L, Kevei F, Nagy E (2002) Complete DNA sequence and analysis of a mitochondrial plasmid in the mycoparasitic *Trichoderma harzianum* strain T95. Plasmid 47:148–152

Barbour AG, Garon CF (1987) Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. Science 237:409–411

Baroudy BM, Venkatasan S, Moss B (1982) Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. Cell 28:315–324

Baroudy BM, Venkatesan S, Moss B (1983) Structure and replication of vaccinia virus telomeres. Cold Spring Harbor Symp Quant Biol 47:723–729

Bosland PW, William PH (1987) An evaluation of *Fusarium oxysporum* from crucifers based on pathogenicity, isozyme polymorphism, vegetative compatibility, and geographic origin. Can J Bot 65:2067–2073

Carling DE, Kuninaga S, Brainard KA (2002) Hyphal anastomosis reactions, rDNA-internal transcribed spacer sequences, and virulence levels among subsets of *Rhizoctonia solani* anastomosis group-2 (AG-2) and AG-BI. Phytopathology 92:43–50

Casjens S (1999) Evolution of the linear DNA replicons of the *Borrelia spirochetes*. Curr Opin Microbiol 2:529–534

Chen CL, Miyasaka A, Miyashita S, Ehara Y, Hashiba T (1992) Genetic relatedness of plasmid-like DNA in isolates from anastomosis group 4 *Rhizoctonia solani*. Ann Phytopathol Soc Jpn 58:286–291

Collins RA, Stohl LI, Cole MD, Lambowictz AM (1981) Characterization of a novel plasmid DNA found in mitochondria of *N. crassa*. Cell 24:443–452

DeLange AM, Reddy M, Scraba D, Upton C, McFadden G (1986) Replication and resolution of cloned poxvirus telomeres in vivo generates linear minichromosomes with intact viral hairpin termini. J Virol 59:249–259

Farmerie WG, Loeb DD, Casavant NC, Hutchison CA, Edgell MH, Swanstrom R (1987) Expression and processing of the AIDS virus reverse transcriptase in *Escherichia coli*. Science 236:305–308

Galligan JT, Kennell JC (2007) Retroplasmids: Linear and circular plasmids that replicate via reverse transcription. Microbiol Monogr 7

Gonzalez A, Talavera A, Almendral JM, Vinuela E (1986) Hairpin loop structure of African swine fever virus DNA. Nucleic Acids Res 14:6835–6844

Griffiths AJF (1995) Natural plasmids of filamentous fungi. Microbiol Rev 59:673–685
Gunge N, Fukuda K, Takahashi S, Meinhardt F (1995) Migration of the yeast linear DNA plasmid from the cytoplasm into the nucleus in *Saccharomyces cerevisiae*. Curr Genet 28:10–288
Hashiba T (1987a) Plasmids responsible for pathogenicity and morphology in *Rhizoctonia solani*. In: Nishimura S, Vance CP, Doke N (eds) Molecular determinants of plant diseases. Japan Scientific Societies and Springer, Tokyo/Berlin, pp 157–168
Hashiba T (1987b) An improved system for biological control of damping-off by using plasmids in fungi. In: Chet I (ed) Innovative approaches to plant disease control. Wiley, New York, pp 337–351
Hashiba T, Homma Y, Hyakumachi M, Matsuda I (1984) Isolation and characterization of a DNA plasmid in the fungus *Rhizoctonia solani*. J Gen Microbiol 130:2067–2070
Hashiba T, Sasaki A, Katsura K (2001) Protoplast fusion and DNA plasmid characterization in *Rhizoctonia solani*. In: Sreenivasaprasad S, Johnson R (eds) Major fungal diseases of rice: recent advances. Kluwer, Dordrecht, pp 223–233
Hinnebusch J, Barbour AG (1991) Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. J Bacteriol 173:7233–7239
Hirota N, Hashiba T, Yoshida H, Kikumoto T, Ehara Y (1992) Detection and properties of plasmid-like DNA in isolates from twenty-three formae speciales of *Fusarium oxysporum*. Ann Phytopathol Soc Jpn 58:386–392
Hirt B (1967) Selective extraction of polyoma DNA from infected mouse cell cultures J Mol Biol 26:365–369
Hongo M, Miyasaka M, Suzuki F, Hashiba T (1994) Expression of the linear DNA plasmid pRS64 in the plant pathogenic fungus *Rhizoctonia solani*. Mol Gen Genet 245:265–271
Jabaji-Hare SH, Burger A, Forget L, Lang BF (1994) Extrachromosomal plasmids in the plant pathogenic fungus *Rhizoctonia solani*. Curr Genet 25:423–431
Katsura K, Suzuki F, Miyashita S, Nishi T, Hirochika H, Hashiba T (1997) The complete nucleotide sequence and characterization of the linear DNA plasmid pRS64-2 from the plant pathogenic fungus *Rhizoctonia solani*. Curr Genet 32:431–435
Katsura K, Sasaki A, Nagasaka A, Fuji M, Miyake Y, Hashiba T (2001) Complete nucleotide sequence of the linear DNA plasmid pRS224 with hairpin loops from *Rhizoctonia solani* and its unique transcriptional forms. Curr Genet 40:195–202
Kempken F (1995) Plasmid DNA in mycelial fungi. In: Kuch U (ed) The mycota. II. Genetics and biotechnology. Springer, Berlin, pp 169–187
Kendrick JB, Snyder WC (1942) Fusarium wilt of radish. Phytopathology 32:1031–1033
Kistler HC, Leong SA (1986) Linear plasmid-like DNA in the plant pathogenic fungus *Fusarium oxysporum* f.sp. *conglutinans*. J Bacteriol 167:587–593
Kistler HC, Bosland PW, Benny V, Leong SA, Williams PH (1987) Relatedness of strains of *Fusarium oxysporum* from crucifers measured by examination of mitochondrial and ribosomal DNA. Phytopathology 77:1289–1293
Kistler HC, Benny U, Powell WA (1997) Linear mitochondrial plasmids of *Fusarium oxysporum* contain genes with sequence similarity to genes encoding a reverse transcriptase from *Neurospora* spp. Appl Environ Microbiol 63:3311–3313
Kuiper MTR, Sabourin JR, Lambowictz AM (1990) Identification of the reverse transcriptase encoded by the Mauriceville and Varkud mitochondrial plasmids of *Neurospora*. J Biol Chem 265:6936–6943
Kuradate K, Namai K, Miyasaka A, Katsura K, Ito K, Hashiba T (1996) Detection of transcripts homologous to DNA plasmids in isolates from six anastomosis groups of *Rhizoctonia solani*. Ann Phytopathol Soc Jpn 62:147–152

Kununaga S, Yokosawa R (1980) A comparison of DNA base compositions among anastomosis groups *Rhizoctonia solani* Kuhn. Ann Phytopathol Soc Jpn 46:150–158

Kuninaga S, Yokosawa R (1985) DNA base sequence homology in *Rhizoctonia solani* Kuhn. VI. Genetic relatedness among seven anastomosis groups. Ann Phytopathol Soc Jpn 51:127–132

Matsuyama N, Moromizato Z, Ogoshi A, Wakimoto S (1987) Grouping *Rhizoctonia solani* Kuhn with non-specific esterase zymogram. Ann Phytopathol Soc Jpn 44:652–658

Meinhardt F, Rohe M (1993) Extranuclear inheritance: linear protein-primed replicating genomes in plants and microorganisms. Prog Bot 54:334–359

Meinhardt F, Schaffrath R, Larsen M (1997) Microbial linear plasmids. Appl Microbiol Biotechnol 47:329–336

Miyasaka A, Chen CL, Hashiba T (1990) Detection and properties of plasmid-like DNA in isolation from nine anastomosis and intraspecific groups of *Rhizoctonia solani*. J Gen Microbiol 136:1791–1798

Miyashita S-I, Hirochika H, Ikeda J-E, Hashiba T (1990) Linear plasmid DNAs of the plant pathogenic fungus *Rhizoctonia solani* with unique terminal structures. Mol Gen Genet 220:165–171

Mogen KL, Siegel MR, Schardl CL (1991) Linear DNA plasmids of the perennial ryegrass choke pathogen, *Epichloë typhina* (Clavicipitaceae). Curr Genet 20:519–526

Momol EA, Kistler HC (1992) Mitochondrial plasmids do not determine host range in crucifer-infecting strains of *Fusarium oxysporum*. Plant Pathol 41:103–112

Nagasaka A, Sasaki A, Sasaki T, Yonezawa M, Katsura K, Hashiba T (2003) Expression and localization of the linear DNA plasmid-encoded protein (RS224) in *Rhizoctonia solani* AG2-2. FEMS Microbiol Lett 225:41–46

Nargang FE, Bell JB, Stohl LL, Lambowictz AM (1984) The DNA sequence and genetic organization of a *Neurospora* mitochondrial plasmid suggest a relationship to intron and mobile elements. Cell 38:441–453

Ogoshi A (1987) Ecology and pathogenicity of anastomosis and intraspecific groups of *Rhizoctonia solani* Kuhn. Annu Rev Phytopathol 25:125–143

Reynolds M, Weinhold AR, Morris TJ (1983) Comparison of anastomosis groups of *Rhizoctonia solani* by polyacryamide gel electrophoresis of soluble proteins. Phytopathology 73:903–906

Rohoziinski J, Girton LE, Van Etten JL (1989) Chlorella viruses contain linear nonpermuted double-stranded DNA genomes with covalently closed hairpin ends. Virology 168:363–369

Rubio V, Tavantzis SM, Lakshman DK (1996) Extrachromosomal elements and degree of pathogenicity in *Rhizoctonia solani*. In: Sneh B et al. (eds) *Rhizoctonia* species: taxonomy, molecular biology, ecology, pathology and disease control, Kluwer, Dordrecht, pp 127–138

Rybchin VN, Svarchevsky AN (1999) The plasmid prophage N15: a linear DNA with covalently closed ends. Mol Microbiol 33:895–903

Samac DA, Leong SA (1989) Characterization of the termini of linear plasmids from *Nectria haematococca* and their use in construction of an autonomously replicating transformation vector. Curr Genet 16:187–194

Shimma Y, Uno I, Hashiba T, Ishikawa T (1988) Characterization of a *Rhizoctonia solani* strain carrying plasmids. J Gen Appl Microbiol 34:111–117

Simpson EB, Ross SL, Marchetti SE, Kennell JC (2004) Relaxed primer specificity associated with reverse transcriptases encoded by the pFOXC retroplasmids of *Fusarium oxysporum*. Eukaryotic Cell 3:1599–1600

Snyder WC, Hansen HN (1940) The species concept in *Fusarium*. Am J Bot 27:64–67
Van den Boogert P, Samalls J, Agsteribbe E (1982) Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondria genomes of *Neurospora crassa*. Nature 298:187–189
Varmus HE, Brown P (1989) Retroviruses. In: Berg DE, Howe MM (eds) Mobile DNA. American Society of Microbiology, Washington DC, pp 53–108
Wako T, Ishikawa T, Hashiba T (1991) Unique DNA plasmid pRS64 associated with chromosomal DNAs of the plant pathogenic fungus *Rhizoctonia solani*. J Gen Microbiol 137:2817–2821
Walther TC, Kennell JC (1999) Linear mitochondrial plasmids of *F. oxysporum* are novel, telomere-like retroelements. Mol Cell 4:229–238
Watanabe B, Matsuda A (1966) Studies on the grouping of *Rhizoctonia solani* Kuhn pathogenic to upland crops. Bull Appl Exp 7:1–131
Wright RM, Cummings DJ (1983) Integration of mitochondrial gene sequence within the nuclear genome during senescence in a fungus. Nature 302:86–88
Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362

Subject Index

1,2-dichloroethan degradation, 73
5-carbamoyl-methylgroup, 207
5-methoxy-carbonylmethyl group, 207

A-factor, 19
A/B toxins, 204
accessory factors, 133
actinomycetes, 101
actinomycin D, 11, 17
actinorhodin, 16
active site of reverse transcriptase, 239
active site tyrosine, 129
adaptation, 214
adenoviruses, 189
Agrobacterium tumefaciens, 109, 118
alkane oxidation, 75
alkene metabolism, 65
alkylbenzene degradation, 78, 79, 81
$\alpha\beta$-precursor, 204
anastomosis groups, 229
antibiotic, 2, 4, 8, 9, 16–21, 23
–, resistance, 3
antigenic variation, 119
apoptotic markers, 210
AraC, 106
arginine, 106
ARS, 44
Arthrobacter
–, *nitroguajacolicus* Rü61a, 90
–, –, pAL1, 90
autonomous replication, 122
autonomously replicating sequence, 44
autoregulatory compound, 106
autoselection, 212
auxin, 108

Bacillus phage phi29, 189
bacteriophages, 21
β-subunit, 206
β-lactam, 106
bidirectional origins, 122
biodegradation, 65
biphenyl
–, degradation, 76
–, degradation polychlorinated biphenyls, 78
bldA, 2, 23
Borrelia
–, *burgdorferi*, 118
–, *hermsii*, 118
botulinum toxin, 206

calcium-dependent antibiotic, 16
Campbell
–, integration, 3, 14
–, type recombination, 20
cap methyltransferase, 201
carboxy, 105
catabolic diversity, 102
CDA, 16
cell wall binding, 205
cephamycin C, 20
chitinase active site, 213
chloramphenicol, 11
chlorocatechol pathways, 82
chloroethene metabolism, 65
circular retroplasmids, 167
class I resistance, 205
class II resistance, 205
Clavibacter, 101
clavulanic acid, 11, 20
clothespin, 228
–, structure, 172
coenzyme M pathway of epoxide metabolism, 65
colicins, 207
colinear regions, 103
combinatorial biosynthesis, 20

conjugal
–, transfer, 54
–, transfer "end first" model, 54
convergent evolution, 207
cross-axis complex, 133
cruciform base-paired structure, 233
cruciform structure, 156
cryptic elements, 195
cut-and-paste transposases, 130
cytokinin, 105
cytoplasmic
–, autonomous elements, 196
–, in vivo recombination, 201
–, nonautonomous elements, 196
–, replication, 198
–, transcription, 198

dagA, 12
deoxynucleotidylation in vitro, 48
DExH/D box helicase, 200
dibenzo-*p*-dioxin degradation, 86
dibenzofuran degradation, 86, 87
diphteria toxin, 206
DNA
–, damage checkpoint, 210
–, polymerase, 228
–, supercoiling, 132
DnaB, 44
double hairpin, 179
double-stranded transfer, 14

ECF (extracytoplasmic function) sigma factors, 21
ECF sigma factor, 21
echinomycin, 11
Elongator
–, complex, 206
–, interacting proteins, 207
end patching, 46
–, DNA polymerase involved, 49
–, DNA template, 49
–, mechanisms, 46
–, recombination model, 47
–, TP-primed synthesis, 47
endocytotic vesicles, 206
endosymbiont, 215
enzyme–DNA intermediate, 126
erythromycin, 19
Escherichia coli, 2–4, 6, 7, 14, 21
ethidium bromide, 3, 7
eubacterial multisubunit RNA polymerase, 193

F-factor, 2, 3, 6, 7
F-prime, 2
ferredoxins, 105
filamentous fungi, 190
flip-flopped structure, 228
fluorene degradation, 87
full-length transcript, 238
fungal pathogens, 229

G1 arrest, 204
γ-butyrolactone, 11
γ-subunit, 204
γ-butyrolactone, 18, 20, 107
general acid catalyst, 129
genetic redundancy, 75
genome conversion, 144, 147, 157
genomic rearrangements, 75, 87
glyoxylate shunt, 108
Gordonia
–, *rubripertincta* B-276, 66
–, –, pNC30, 67

hairpin, 143, 173
hairpin plasmid, 228
hairpin-binding module, 126, 128
heavy metal resistance, 104
Hfr, 2
Holliday junction, 132
homology, 103
horizontal gene transfer, 75
hormone, 102
hp telomere cleavage, 131
hydrophobic pocket, 130
hyperplasia, 106
hypovirulence, 241

IF, 5
ImmD, 208
immunity factors, 212
initiation mechanisms, 177
inserted plasmids, 195
insertion mutagenesis, 104
integrase, 145
integrated hairpin plasmid, 236
Integration, 181
inverted terminal repeats, 228
invertron, 9, 101

invertrons, 64
IS*466*, 12, 13
IS*468*, 14
iteron, 44

kalilo plasmid, 191
Kex1 protease, 203
killer systems, 203
Krebs cycle, 108

lankacidin, 7, 11, 20
lankamycin, 20
lasalocid, 11, 17
linear
–, dsDNA elements, 188
–, retroplasmids, 165, 172
linear plasmid, 142, 144
–, conjugative, 103
–, prophage, 142, 143, 157, 160
Lyme disease, 118
LysM signature, 205
LysR, 106

malate synthase, 108
mannosyl-diinositolphosphoceramide, 205
maranhar plasmid, 192
mating competence, 210
mating-type status, 210
methylenomycin, 2, 8, 10, 11, 16–21, 23
microbial evolution, 100
mitochondrial
–, plasmids, 164
–, yeast linear plasmids, 196
mobile genetic elements, 92
molecular
–, disease, 195
–, fossils, 165
mRNA capping enzyme, 200
mutation induction, 210
Mycobacterium, 101
–, *rhodesiae* JS60, 67
Möbius strip resolution of sister replicons, 52

naphthalene degradation, 84
negative supercoiling, 133
NF, 5, 12, 13
NF-like, 12
niche, 100
NLS, 55
Nocardioides sp. JS614 pNoc614, 67
noncovalently linked ends, 234
nonribosomal peptide synthetase, 107
nuclear localization signal in Terminal Proteins of Streptomyces, 55

Ochrobactrum sp. TD, 72
origin of replication, 120
oriP, 45
oriT, 54
oxytetracycline, 11, 16, 22

P450-type cytochrome monooxygenases, 105
parA, 50
parAB, 15, 16, 18
parB, 50
parS, 50
partitioning, 144, 159
–, *Streptomyces* linear plasmid, 50
partitioning genes, 15
pathogenicity, 100
pathogenicity island, 100
penicillin-binding proteins, 107
pentameric repeat, 234
PFGE, 8, 9, 17
pFOXC2, 228
pFOXC3, 228
pGKL system, 199
phage
–, öKO2, 142
–, N15, 142
–, PY54, 142
phylogenetic divergence, 214
phytotoxins, 107
pKSL, 11
plasma membrane ATPase, 205
plasmid
–, compatibility, 124
–, initiator, 123
–, maintenance, 123
plasmid replication, 103
plasmids
–, mitochondrial, 164
–, variant, 172
poly(A)$^-$RNA, 237
positive
–, DNA supercoiling, 133
–, twist, 134
pox viruses, 201

poxvirus replication, 121
pPZG101, 11, 16, 22
pPZG102, 11
pre–pro sequence, 204
pristinamycin, 20
protein primed replicating linear plasmids, 188
protelomerase, 144, 147, 148, 155, 157, 159
–, activity
–, –, in vitro, 147, 148, 154
–, –, in vivo, 147, 151
–, cleaving–joining activity, 150, 156
–, DNA-binding, 150, 152, 159
–, monomeric, 152
–, processing, 150, 156
–, promoter, 159
protocatechuate degradation, 79, 82, 87
pRS104, 229
pRS188, 229
pRS224, 234
pRS64-1, -2, and -3, 233
pSCL1, 11
Pseudomonas putida AJ, 72
pseudotelomere, 39
pSLA2, 11
pSLA2-L, 2, 11, 17, 19, 20, 22
pSRM, 11
pSV1, 2, 18, 19
pSV2, 11
pulsed field gel electrophoresis, 8
pulsed-field electrophoresis, 103

quinaldine (2-methylquinoline) degradation, 90
quorum-sensing, 19, 20

R-factor, 3
racket
–, frame, 9
–, frame model, 9, 11
Rad53, 210
rep, 44
rep function, 123
replicated telomere, 124
Replication
–, initiation, 193
–, of pRS64, 236
–, via an RNA intermediate, 241
replication, 44, 144, 147, 157
–, θ mode, 158
–, bidirectional, 157
–, initiation, 44
–, origin, 44
resolution, 52
ResT, 125
ResT–ResT
–, communication, 132
–, interactions, 133
restriction fragment length polymorphism, 232
retroplasmid, 239
retroplasmids, 164
–, circular, 167
–, linear, 165, 172
reverse
–, transcriptase, 164
–, transcription, 175, 176
reverse reaction, 131
reverse transcriptase, 239
–, activity in DNA polymerase I, 48
–, activity in topoisomerase I, 48
Rhococcus sp. CIR2, 84
Rhodococcus
–, *equi*, 102
–, *erythropolis*, 103
–, *erythropolis* BD2, 81
–, *erythropolis* BD2 pBD2, 81
–, *erythropolis* PR4, 75
–, *erythropolis* PR4 pREL1, 75
–, *erythropolis* TA421, 81
–, *erythropolis* TA421 pTA421, 81
–, *fascians*, 101
–, *globerulus* P6, 81
–, *opacus*, 9
–, *opacus* 1CP, 82
–, *opacus* 1CP p1CP, 82
–, *opacus* M213, 87
–, *opacus* M213 pNUO1, 87
–, *opacus* SAO101, 86
–, *opacus* SAO101 pWK301, 86
–, *rhodochrous* K37, 81
–, sp. DK17, 79
–, sp. DK17 pDK2 and pDK3, 79
–, sp. I24, 84
–, sp. NCIMB12038, 84
–, sp. NCIMB12038 p2SL1, 84
–, sp. RHA1, 76
–, sp. RHA1 pRHL1, 76
–, sp. RHA1 pRHL2, 76
–, sp. RHA1 pRHL3, 76

–, sp. strains NCIMB12038, P200 and P400, 84
–, spp., 74
rlrA, 45
RNA
–, guanylyltransferase, 201
–, helicases, 107
–, polymerase, 190
–, triphosphatase, 201
–, world, 181

SAP1, 11, 17
satellite plasmid, 167
scab, 106
SCP1, 2, 5–8, 10–19, 21–23
SCP1′, 14
SCP1′-*argA, uraB*, 7
SCP1′-*cysB*, 7
SCP2, 7, 8, 10, 16
secondary
–, membrane receptor, 205
–, metabolism, 2
secondary metabolism, 19, 20, 22
SEDS family, 107
self-immunity, 208
selfish DNA element, 196
senescence, 172, 181, 194
sequence-specific binding, 127
sheared purine pairs, 37
signal molecules, 102
signal peptides, 203
single strand DNA binding protein, 199
single-stranded circle, 234
site-specific nicking, 236
sliding clamp, 193
SLP2, 10, 11, 16
soil bacterium, 109
spore, 4, 5, 15
–, associated proteins, 19, 22, 23
sporulation, 4, 5, 16, 19
Streptomyces, 101
–, *avermitilis*, 11, 17, 18, 21
–, *clavuligerus*, 8, 11, 20
–, *coelicolor*, 1, 4, 5, 7, 8, 10, 11, 14, 16–19, 21, 23
–, *fradiae*, 9, 11
–, *griseus*, 19
–, *lasaliensis*, 9, 11, 17
–, *lividans*, 11, 16
–, *parvulus*, 7, 9–11
–, *rimosus*, 8, 11, 16, 22
–, *rochei*, 7, 9, 11, 17, 19, 20
–, *turgidiscabies*, 101
–, *venezuelae*, 9, 11
–, *violaceoruber*, 9–11, 18
substrate linearization, 147
symptom, 103

T7 RNA polymerase, 193
tac, 42
tap, 40
telO, 148, 150, 153, 155
telomerase occupancy site, 148
telomere
–, fusions, 131
–, resolution, 124, 150, 154, 155, 159
–, resolution site, 148, 156
–, resolvase, 125
telomere-like repeats, 172
telomeres, 9, 24
–, archetypal, 35
–, nonarchetypal, 38
telRL, 148, 150, 152, 155
temperate phage, 142
–, genome, 143–145
–, integration, 145
terminal
–, inverted repeat, 189
–, inverted repeats, 38
–, inverted repeats (TIRs), 8
–, protein, 8, 9, 40
–, –, archetypal, 40
–, –, nickase model, 46
–, –, nonarchetypal, 42
–, –, primer model, 47
–, proteins, 189
–, recognition factor, 199
–, sequences, 74
terminal inverted repeats, 101
Terrabacter sp. DBF63, 87
–, pDBF1, 87
tetracycline, 17, 19
TIR, 14, 22, 38
–, binding protein, 194
TIRs, 9, 16, 25
Tn*4811*, 16
Tn*5714*, 16
Topo IV, 52
topoisomerase IV decatenation, 52
tos, 148, 151

TP, 40
–, interactions, 43
TP-DNA polymerase complex, 189
tpc, 42
Tpg, 40
tpg, 40
*tpg*pseudogene, 42
TPR2
–, domain, 193
transcription
–, initiation, 202
–, reverse, 175
–, termination, 202
transcriptional regulator, 106
transfer origin, 14
translational activator, 106
trichloroethene cometabolism, 81
tRNA
–, methyltransferase, 207
–, modification, 206
$tRNA^{Glu}$, 207
tRNA-like structure, 171
TTA, 23
–, codon, 2, 23
ttrA, 38, 54
tylosin, 11
type I toxin, 209
type IB topoisomerases, 125
type II toxin, 210
tyrosine
–, recombinase, 125
–, recombinase-like catalytic region, 126

UF, 5
undecylprodigiosin, 16
upstream conserved sequence, UCS, 202
UUA codons, 23

vaccinia virus NphI, 200
variant plasmids, 172
viral
–, attributes, 215
–, B-type DNA polymerase, 199
–, B type DNA polymerase, 192
virion structure, 214
virulence factors, 100

Xanthobacter
–, *autotrophicus* GJ10, 73
–, *autotrophicus* GJ10 pXAU1, 73
Xanthobacter sp. Py2, 65
–, pEK1, 65

Yeast linear plasmids, 196
YKEK-motif, 130
YREK motif, 130

Z conformation, 156
zymocin, 203
–, α-like protein, 213