Eukaryotic Membranes and Cytoskeleton

Origins and Evolution

ADVANCES IN EXPERIMENTAL MEDICINE AND BIOLOGY

Recent Volumes in this Series

Volume 600
SEMAPHORINS: RECEPTOR AND INTRACELLULAR SIGNALING MECHANISMS
Edited by R. Jeroen Pasterkamp

Volume 601
IMMUNE MEDIATED DISEASES: FROM THEORY TO THERAPY
Edited by Michael R. Shurin

Volume 602
OSTEOIMMUNOLOGY: INTERACTIONS OF THE IMMUNE AND SKELETAL SYSTEMS
Edited by Yongwon Choi

Volume 603
THE GENUS YERSINIA: FROM GENOMICS TO FUNCTION
Edited by Robert D. Perry and Jacqueline D. Fetherson

Volume 604
ADVANCES IN MOLECULAR ONCOLOGY
Edited by Fabrizio d'Adda di Gagagna, Susanna Chiocca, Fraser McBlane and Ugo Cavallaro

Volume 605
INTEGRATION IN RESPIRATORY CONTROL: FROM GENES TO SYSTEMS
Edited by Marc Poulin and Richard Wilson

Volume 606
BIOACTIVE COMPONENTS OF MILK
Edited by Zsuzsanna Bősze

Volume 607
EUKARYOTIC MEMBRANES AND CYTOSKELETON: ORIGINS AND EVOLUTION
Edited by Gáspár Jékely

Volume 608
BREAST CANCER CHEMOSENSITIVITY
Edited by Dihua Yu and Mien-Chie Hung

Eukaryotic Membranes and Cytoskeleton

Origins and Evolution

Edited by
Gáspár Jékely, Ph.D.
European Molecular Biology Laboratory, Heidelberg, Germany

Springer Science+Business Media, LLC
Landes Bioscience

Springer Science+Business Media, LLC
Landes Bioscience

Printed in the U.S.A.

Springer Science+Business Media, LLC, 233 Spring Street, New York, New York 10013, U.S.A.
http://www.springer.com

Please address all inquiries to the Publishers:
Landes Bioscience, 1002 West Avenue, 2nd Floor, Austin, Texas, U.S.A. 78701
Phone: 512/ 637 6050; FAX: 512/ 637 6079
http://www.landesbioscience.com

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution edited by Gáspár Jékely, Landes Bioscience / Springer Science+Business Media, LLC dual imprint / Springer series: Advances in Experimental Medicine and Biology

ISBN: 978-0-387-74020-1

Library of Congress Cataloging-in-Publication Data

A C.I.P. catalog is available from the Library of Congress.

PREFACE

This book discusses the evolutionary origin and diversification of eukaryotic endomembranes and cytoskeleton from a cell biological and comparative genomic perspective. The main idea behind this book was to try to convince experimental cell biologists to speculate about the evolutionary origin of cell biological processes they are working on and are deeply familiar with. By doing so, I expected to gain fresh insights into the problem of eukaryote origins from scientists who know a lot about how eukaryotic cells function. To my great happiness, many cell biologists accepted the challenge and provided in-depth cell evolutionary analyses or by teaming up with bioinformaticians carried out comparative genomic surveys. Their contributions, together with contributions from paleontologists and evolutionary biologists, provide a diversity of viewpoints and a fresh look on many aspects of eukaryote evolution.

The first two chapters set the stage for discussions about the origins of eukaryotic cell biological features by describing the early fossil record of eukaryotic evolution (Chapter 1) as well as the current status of eukaryote phylogeny and the possible rooting of the eukaryotic tree (Chapter 2). The *explanandum* is therefore presented in these chapters: About 1500 million years ago, in a world inhabited by prokaryotes, a novel type of cellular organization appeared, characterized by a complex network of endomembranes, and a dynamic internal skeleton able to protrude, constrict and move these membranes. Cells having this architecture subsequently diversified into major clades to make up present day eukaryote diversity. All the other chapters in this volume attempt to explain certain aspects of this major transition and the diversification that followed. Chapters 3-7 discuss endomembrane evolution from theoretical, cell biological and comparative genomic perspectives. Chapters 8 and 9 deal with the origin of the actin cytoskeleton and autophagy, whereas Chapters 10 and 11 present models on the evolution of centrosomes and sensory-motile cilia.

Many of the chapters present original research data from comparative genomic surveys. The presence/absence of gene families with central roles in endomembrane and cytoskeleton dynamics in a variety of eukaryotic taxa and an understanding of eukaryote phylogeny allow us to accurately reconstruct the cellular machineries present in the last common ancestor of eukaryotes. Such a reconstruction is fundamental if we are to understand eukaryotic diversification since this is the ancestral cell from which all diversity arose. Comparative genomics can likewise tell us which lineages expanded or reduced certain gene families and the associated cellular

machineries. As an example, Chapter 6 discusses how the expansion of the Rab family of membrane traffic regulators paralleled the complexification of endomembranes during metazoan evolution and during the evolution of certain parasitic lineages.

The earlier cell evolutionary history, before the last common ancestor, can also be inferred in some cases, mostly from sequence and structural comparison of cellular complexes. Chapters 3 and 6, for example, discuss how structural and sequence analyses of vesicle coating complexes reveal their common origin from an ancestral coat complex during the origin of eukaryotic cellular organization. Chapter 8 on the other hand describes how one can infer, based on structural and sequence comparisons, the divergence of cytoskeletal components from ancestral, simpler protein complexes.

The volume is not an attempt to cover all aspects of eukaryotic endomembrane and cytoskeleton evolution in an encyclopedic manner. It examines the problem along some transects of eukaryote cell biology, but several important issues are not touched upon such as the origin of mitosis, syngamy or peroxisomes. The evolution of symbiotic organelles such as chloroplasts and mitochondria is also not covered, despite their fundamental importance. These omissions were necessary due to space constraints and also to minimize overlap with other recent volumes.

Gáspár Jékely, Ph.D.

ABOUT THE EDITOR...

GÁSPÁR JÉKELY is a postdoctoral scientist at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany. His main research interests include the evolution of eukaryotic cellular organization and the early evolution of metazoan brains. He received his Ph.D. from the Eötvös Loránd University in Budapest, Hungary. He did his first Postdoctoral at the EMBL on *Drosophila* cell migration. Following this he spent a semester at the Collegium Budapest as a Junior Fellow working on cell evolution after which he moved back to the EMBL to work on brain evolution. He serves on the editorial board of *Biology Direct*.

PARTICIPANTS

Juliette Azimzadeh
UMR 144-CNRS-Institut Curie
Paris
France

William E. Balch
Departments of Cell and Molecular Biology and The Institute for Childhood and Neglected Disease
The Scripps Research Institute
La Jolla, California
USA

Victor V. Banin
Research Center
Russian State Medical University
Moscow
Russia

Galina V. Beznoussenko
Department of Cell Biology and Oncology
Consorzio Mario Negri Sud
Santa Maria Imbaro (Chieti)
Italy

Michel Bornens
UMR 144-CNRS-Institut Curie
Paris
France

Henner Brinkmann
Canadian Institute for Advanced Research
Evolutionary Biology Program
Department de Biochimie
Centre Robert-Cedergren
Université de Montreal
Montreal, Quebec
Canada

Fatima Cvrčková
Department of Plant Physiology
Faculty of Sciences
Charles University
Prague
Czechia

Joel B. Dacks
Department of Pathology
University of Cambridge
Cambridge
UK

Viacheslav V. Dolgikh
Laboratory of Microbiological Control
All-Russian Institute for Plant Protection
Russian Academy of Agricultural Sciences
St. Petersburg-Pushkin
Russia

Mark C. Field
Department of Pathology
University of Cambridge
Cambridge
UK

Carme Gabernet-Castello
Department of Pathology
University of Cambridge
Cambridge
UK

Cemal Gurkan
Department of Cell Biology
The Scripps Research Institute
La Jolla, California
USA

Timothy Hughes
Bergen Centre for Computational Science
University of Bergen
Bergen
Norway

Emmanuelle J. Javaux
University of Liège
Department of Geology
Liège Sart-Tilman
Belgium

Gáspár Jékely
European Molecular Biology Laboratory
Heidelberg
Germany

Atanas V. Koulov
Department of Cell Biology
The Scripps Research Institute
La Jolla, California
USA

Alberto Luini
Department of Cell Biology and Oncology
Consorzio Mario Negri Sud
Santa Maria Imbaro (Chieti)
Italy

Alexander Mironov
Department of Cell Biology and Oncology
Mario Negri Sud Institute
Santa Maria Imbaro (Chieti)
Italy

David R. Mitchell
Department of Cell and Developmental Biology
SUNY Upstate Medical University
Syracuse, New York
USA

Hervé Philippe
Canadian Institute for Advanced Research
Evolutionary Biology Program
Department de Biochimie
Centre Robert-Cedergren
Université de Montreal
Montreal, Quebec
Canada

Francisco Rivero
Center for Biochemistry and Center for Molecular Medicine
University of Cologne
Cologne
Germany

Tor Erik Rusten
Department of Biochemistry
The Norwegian Radium Hospital
Montebello, Oslo
Norway

Thomas U. Schwartz
Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts
USA

Irina S. Sesorova
Department of Biology
Shuja State Pedagogical University
Shuja
Russia

CONTENTS

PREFACE v

1. THE EARLY EUKARYOTIC FOSSIL RECORD 1

Emmanuelle J. Javaux

Introduction 1
The Geological Record 2
Dating Biological Innovations 10
Conclusions 15

2. THE DIVERSITY OF EUKARYOTES AND THE ROOT OF THE EUKARYOTIC TREE 20

Henner Brinkmann and Hervé Philippe

Woese's Paradigm 20
Progress in Tree Reconstruction Methods and in Sequencing: The Promise of Phylogenomics 21
The Root of the Tree of Life 25
"Primitive" Eukaryotes: A Second Major Paradigm Shift 26
Current Status of the Eukaryotic Phylogeny 27
Molecular Dating 29
From Simple Organism to Simple Processes 31

3. ORIGIN OF EUKARYOTIC ENDOMEMBRANES: A CRITICAL EVALUATION OF DIFFERENT MODEL SCENARIOS 38

Gáspár Jékely

Introduction 38
Physico-Chemical Constraints of Membrane Topogenesis 39
Symbiotic Scenarios for the Origin of the Nucleus and Endomembranes 40
Origin of Endomembranes by de novo Vesicle Formation 44
Autogenous Scenarios 45
Interspecific Interactions at the Origin of Eukaryotic Endomembranes 49

4. ORIGINS AND EVOLUTION OF COTRANSLATIONAL TRANSPORT TO THE ER 52

Thomas U. Schwartz

Introduction 52
Mediators of Cotranslational Transport 53
The Control of Cotranslational Targeting 55
Evolutionary Considerations 57

5. EVOLUTION OF THE ENDOPLASMIC RETICULUM AND THE GOLGI COMPLEX 61

Alexander A. Mironov, Victor V. Banin, Irina S. Sesorova, Viacheslav V. Dolgikh, Alberto Luini and Galina V. Beznoussenko

Introduction 61
Common and Peculiar Features of the Golgi Complex in Different Eukaryotic Cells 64
The Minimal Set of Genes Involved in Intracellular Traffic 65
Models of Evolution of the ER and the Golgi Complex 66

6. AN EVOLUTIONARY PERSPECTIVE ON EUKARYOTIC MEMBRANE TRAFFICKING 73

Cemal Gurkan, Atanas V. Koulov and William E. Balch

Introduction 73
Coat Protein Complexes: Cellular Machineries Driving Vesicle Formation/Fission 74
SNARE Proteins: Cellular Machineries Driving Membrane Docking/Fusion 77
Rab GTPases: Key Regulators of Membrane Trafficking 77

7. RECONSTRUCTING THE EVOLUTION OF THE ENDOCYTIC SYSTEM: INSIGHTS FROM GENOMICS AND MOLECULAR CELL BIOLOGY 84

Mark C. Field, Carme Gabernet-Castello and Joel B. Dacks

Defining Endocytosis 85
The General Structure and Morphological Evolution of Endocytic Systems 86
Key Factors Involved in Endocytic Systems and Their Evolutionary Distribution 88
Perspectives 93

8. ORIGINS AND EVOLUTION OF THE ACTIN CYTOSKELETON 97

Francisco Rivero and Fatima Cvrčková

Introduction 97
The Actin Cytoskeleton in the Cellular Context 100
Actin- and Monomer-Binding Proteins 100

Actin Nucleation Complexes 103
Other Actin-Binding Proteins 104
Effectors and Regulators of the Actin Cytoskeleton 105

9. ORIGIN AND EVOLUTION OF SELF-CONSUMPTION: AUTOPHAGY 111

Timothy Hughes and Tor Erik Rusten

Summary 111
Introduction 111
Degradation of Proteins in Chambered Proteases in Pro- and Eukaryotes 112
Different Modes of Autophagy Degrade Long-Lived Proteins and Organelles in Eukaryotes 112
The Autophagy-Specific Molecular Machinery Involves Two Ubiquitin-Like Conjugation Pathways 113
Is Autophagy a Pan-Eukaryotic Process? 114
Discussion 116

10. ORIGIN AND EVOLUTION OF THE CENTROSOME 119

Michel Bornens and Juliette Azimzadeh

Summary 119
Introduction 119
The Centrosome-Nucleus Connection 120
Centrosome Has Evolved with Multicellularity 120
The Evolution of the Centrin Genes 121
The Centrosome-Nucleus Connection and the Duplication Process 122
Centrosome Has Evolved in Unikonts 125

11. THE EVOLUTION OF EUKARYOTIC CILIA AND FLAGELLA AS MOTILE AND SENSORY ORGANELLES 130

David R. Mitchell

Evidence for the Presence of a 9+2, Motile, Sensory Organelle in the Last Common Eukaryotic Ancestor 130
Evolution of Tubulin, Dynein and Kinesin 133
The Origins of 9+2 Flagella 134
Diversification of Flagellar Structure and Function during Eukaryotic Radiation 136

INDEX 141

CHAPTER 1

The Early Eukaryotic Fossil Record

Emmanuelle J. Javaux*

Abstract

The Precambrian era records the evolution of the domain Eucarya. Although the taxonomy of fossils is often impossible to resolve beyond the level of domain, their morphology and chemistry indicate the evolution of major biological innovations. The late Archean record for eukaryotes is limited to trace amounts of biomarkers. Morphological evidence appears in late Paleoproterozoic and early Mesoproterozoic (1800-1300 Ma) rocks. The moderate diversity of preservable eukaryotic organisms includes cell walls without surface ornament (but with complex ultrastructure), with regularly distributed surface ornamentation, and with irregularly or regularly arranged processes. Collectively, these fossils suggest that eukaryotes with flexible membranes and cytoskeletons existed in mid-Proterozoic oceans. The late Mesoproterozoic-early Neoproterozoic (1300-750 Ma) is a time of diversification and evolution when direct evidence for important biological innovations occurs in the fossil record such as multicellularity, sex, photosynthesis, biomineralization, predation, and heterotrophy. Members of extant clades can be recognized and include bangiophyte red algae, xanthophyte algae, cladophorale green algae, euglyphid, lobose, and filose amoebae and possible fungi. In the late Neoproterozoic, besides more diversification of ornamented fossils, florideophyte red algae and brown algae diversify, and animals take the stage.

The record of biological innovations documented by the fossils shows that eukaryotes had evolved most cytological and molecular complexities very early in the Proterozoic but environmental conditions delayed their diversification within clades until oxygen level and predation pressure increased significantly.

Introduction

The origin of the eukaryotic cell is still not resolved despite the numerous hypotheses proposed since Margulis[1] and earlier (reviewed in ref. 2) and tested using electronic microscopy and molecular biology. Many hypotheses have been proposed, such as fusion of two[3,4] or three[5] prokaryotic cells, endosymbiosis[1] or evolution from a proto-eukaryote LCA (last common ancestor).[6] Confusion comes in part from the chimeric character of the eukaryotic cell, possessing attributes of Archaea and Bacteria, but also some specific characteristics. All the amitochondriate eukaryotes known so far seem to be highly derived rather than recording a primitive step in eukaryotic evolution.[7] Hydrogenosomes and mitosomes are organelles that evolved polyphyletically from mitochondria as secondary anaerobic adaptations.[8] The first eukaryote has been suggested to be a phagotrophic heterotroph with a cilium (unikont) and facultative aerobe.[8] This eukaryote would have had already all the eukaryotic features; a nucleus, an endosplamic reticulum and Golgi apparatus, a cytoskeleton including one or two flagella and a mitochondrion capable of oxidative phosphorylation.[9] Thus, the emerging view is that

*Emmanuelle J. Javaux—University of Liège, Department of Geology, 17 allée du 6 Août B18, 4000 Liège Sart-Tilman, Belgium. Email: EJ.Javaux@ulg.ac.be

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

the last common ancestor of eukaryotes was aerobic, and mitochondria originated immediately following or even during the origin of the nucleus and microtubules[8,10] instead of a stepwise evolution. Plastids were acquired later in evolution by endosymbiosis with a cyanobacterial ancestor of the chloroplast (primary endosymbiosis), giving rise to photosynthetic eukaryotes. Successive endosymbiotic events where an eukaryotic host engulfed another photosynthetic eukaryote gave rise to the diversity of photosynthetic eukaryotes with multiple membranes around their chloroplasts.[11] Although it has been suggested that only a small fraction of eukaryotic diversity was known, and the rest waiting to be discovered, it seems now that surveys of environmental (rather than culture) gene sequences have largely overestimated the real diversity at the kingdom level and that most sequences discovered are related to known groups.[12] Although immense lower-level diversity remains to be characterized, we may have a relatively good picture of extant eukaryotic (high-level) diversity, that includes six kingdoms[8,9] five supergroups[13] or 8 supergroups.[14] (see the chapter by Henner Brinkmann and Hervé Philippe). These kingdoms might have diverged into a unikont group (one flagellum) and a bikont group (two flagella) from the ancestor, based on myosin phylogeny and gene-fusion data.[8,15] The existence of extant representatives predating the unikont/bikont divergence cannot be excluded but recent studies suggest, as mentioned above, that the first eukaryote was already a "complete" eukaryote. Determining the timing of diversification requires a close analysis of the fossil record. Recently, several papers have reviewed the early evolution of the domain Eucarya, including criteria for recognizing early eukaryotes,[16] the record of Neoproterozoic eukaryotes,[17] and trends and controls of early eukaryote diversification and morphological innovations.[18]

In this chapter, I review the geological record of eukaryotes in the Precambrian, and the biological innovations recorded in the morphology and chemistry of the fossils. Molecular phylogenies provide important information or hypotheses about relationships between kingdoms and order of branching. However paleobiological data are essential for testing these trees and for precising the timing of diversification. They may also record ancestral forms (and steps in evolution) that might not have any extant relatives. Regardless of taxonomy, fossils display morphological attributes related to major biological innovations.[18]

The Geological Record

Several lines of evidence can be used to decipher the early record of eukaryotic cells. Genetic material is rarely preserved in the rock record, and thus paleontologists have to rely on other features to identify microfossils as members of the domain Eucarya. Geochemists can search for phylogenetically informative lipids (biomarkers and biopolymers) preserved in kerogens and bitumens, and reconstruct part of the paleodiversity or at least the evolution of biochemical pathways. Measurements of isotopic fractionation are useful mostly for tracking effects of prokaryotic metabolisms in the early rock record (e.g., see refs.19-21), but it could help clarify the biological affinities of eukaryotes in some cases. Fossils provide direct evidence of early cells, and document steps in biological and biochemical innovations.[18] Fossils can display morphological and ultrastructural features showing a degree of complexity and/or particular features unknown in prokaryotic organisms, therefore pointing to a eukaryotic affinity.[16] Indeed, the wall structure and ornamentation, the presence of processes that extend from the vesicle wall, the presence of excystment structures (openings through which cysts liberate their content), the wall ultrastructure and the wall chemistry can clarify the biological affinities of organic-walled microfossils at the level of the domain, and even at the level of class in some cases. The discovery of extant picoeukaryotes[22] or huge bacteria[23] suggests that size itself is not a good criterion for differentiating prokaryotic from eukaryotic microbes. However macroscopic size and regular well-defined morphology of some Proterozoic fossils and carbonaceous compressions suggest their eukaryotic affinities as well. Microfossils with ornamented walls occur in the rock record as far back as the later Paleoproterozoic, documenting early eukaryote evolution and implying a preceding history of evolution. The estimated age for the origin of

eukaryotes varies widely, ranging from the Archean, as suggested by the biomarker record[24,25] (see ref. 34 for an opposite view) and molecular clock[27] to the Paleoproterozoic based on molecular clocks[28] or the Neoproterozoic (this latter hypothesis discards all fossil evidence prior to 850 Ma).[26] The diversification of most eukaryotic kingdoms by the late Mesoproterozoic-early Neoproterozoic is supported by the fossil record (reviewed in ref. 29) and molecular clocks.[28,30]

Chemical Evidence for Eukaryotes

Biomarkers derive from biochemical precursors by reductive or oxidative processes and generally include lipids and pigments. Biomarkers in 2.7 Ga kerogens of the Fortescue Group, Australia, seem to indicate that contemporaneous cells were able to synthesize sterols.[24,25] Most prokaryotes do not synthesize sterols. However sterol pathway seems to share a common ancestry between bacteria and eukaryotes.[31] Some of the genes and enzymes for sterol synthesis are found in some prokaryotes, but many of the essential genes seem absent.[32] Sterol synthesis has been reported in some cyanobacteria, myxobacteria, mycobacteria, and planktomycetes but most cases either result from contamination, incorporation of molecules made by eukaryotic organisms or synthesis directed by genes transferred laterally from eukaryotic cells.[25,32,33] Rare bacteria have been reported to produce methyl-sterols; these, however, differ structurally from eukaryotic sterols (reviewed in ref. 25). Therefore, C-24 alkylated sterols, the particular sterols found in eukaryotes, might still be used as biomarker of eukaryotic cells in the sedimentary record[32] (for an alternative view see ref. 34). Thus, 2.77 Ga biomarkers would seem to set a minimum date for the domain Eucarya. However, if they originated in early Archean oceans, early eukaryotes probably had a long history of restricted distribution to slightly oxygenated niches in early ecosystems when anoxic, sulphidic stratified oceans dominated.[35-38] C 27-29 steranes are preserved in shales of ~1.7 Ga estuary environment at the base of the Chuanlinggou Fm of China.[39] In 1.64 Ga shales from the McArthur Basin, northern Australia, deposited offshore below the wave base but still in the photic zone, biomarkers indicate the presence of green and purple sulphur bacteria but no eukarya.[38] In the ~1.5 Ga Roper Group of northern Australia, distribution of eukaryotes assemblages range from nearshore to outer platform environments but the diversity of eukaryotic microfossils is higher in nearshore peritidal facies.[40] This restricted distribution of early eukaryotic assemblages could be linked to seaward decreasing availability of trace elements metabolically important for eukaryotic algae.[41] However whether Roper eukaryotes were heterotrophic or autotrophic protists is unknown.

Archean and Proterozoic rocks have yielded biomarkers of alveolates (which include dinoflagellates and ciliates, among other groups). Dinosterane, derived from dinosterol produced by dinoflagellates and their ancestors, occurs in successions ranging from the Archean[25] through the PaleoProterozoic,[42] Mesoproterozoic[43-46] and Neoproterozoic.[43,44,47] However the taxonomic distribution of dinosterol is not well constrained. Gammacerane, derived from tetrahymenol produced by ciliates has been found in 1.7 Ga Tuanshanzi Formation of China[42] and ~ 750 rocks of the Chuar Group, Arizona[48] however it may also be derived from bacteria.[49]

Biopolymers are complex carbohydrate molecules making up biological cell walls. Their composition can be determined by microchemical analysis of single microfossils and in some cases, related to specific clades (e.g., ref. 50). On the contrary to biomarkers, direct link can be made between the source organism and the molecules, and there are no contamination problems.[51] Dinosporin-like biopolymer isolated in extant dinoflagellate walls has been detected possibly in Neoproterozoic process-bearing acritarchs without tabulation (a spatial organization of organic plates characterizing dinoflagellate wall).[50,52] However, the taxonomic distribution of this biopolymer is not known. Algaenan, a biopolymer synthesized by green algae (chlorophyte and eustimagtophycea, one dinoflagellate vegetative cell wall, ref. 53) has been recognized in a late Neoproterozoic acritarch.[50] These biopolymers record the evolution of complex eukaryotic synthetic pathways, and possibly support the fossil and biomarker evidence for the evolution of alveolates and chlorophytes in the Proterozoic, if synthesized only by these groups.

Morphological Evidence: Fossils of Early Eukaryotes

Two categories of morphological evidence document the evolution of eukaryotes in the Proteorozoic: body fossils and carbonaceous compressions.

Body Fossils

Archean Record (4 to 2.5 Ga)

No fossils of eukaryotic cells have been recognized so far in the Archean rock record. Only biomarkers may indicate the evolution of a eukaryotic feature, the eukaryotic sterol pathway.[24]

Paleoproterozoic Record (2.5 to 1.6 Ga)

In the late Paleoproterozoic of China, large (up to 240 μm) sphaeromorphs (unornamented organic-walled vesicles) with regular medial splits, similar to the excystment structure of some protists[54] might document early eukaryotes. The ornamented acritarch, *Valeria lophostriata*, appears in the ~1.8 Ga Chuanliggou Formation of China and in the 1.65 Ga Mallapunyah Fm (McArthur Group, Australia) and records the oldest microfossil evidence for the domain Eucarya.[55] *Valeria lophostriata* is a spherical acritarch easily distinguished by its distinctive ornament of concentric striations (Fig. 1: 5-8). SEM observation shows that these striations consist of parallel ridges spaced 1 μm apart that traverse the inner surface of the vesicle (Fig. 1: 6). *Valeria* has a long stratigraphic range, extending from the late Paleoproterozoic (ref. in 18), through the Mesoproterozoic and Neoproterozoic.[56]

Mesoproterozoic Record (1.6 to 1 Ga)

In the early Mesoproterozoic, early eukaryotes start to diversify and occupy more niches, possibly in relation with rising oxygen levels. The type of ornamentation on the organic-walled vesicles includes concentric ridges, polygonal plates, polygonal network, processes (vesicle expansions) of various shapes and dimensions. Preserved wall ultrastructure range from mono-to multi-layered walls, with layers varying in texture and electron density. Reproduction occurs by budding and binary fission. Excystment (opening of a cyst) occurs by medial split and possibly opening at the end of a neck-like extension. Resting and vegetative stages are documented.

Early Mesoproterozoic fossils strongly indicate that eukaryotic organisms of marked cytological and genetic complexity existed 1500-1300 million years ago. Relative to earlier assemblages, early Mesoproterozoic protists show higher diversity and more obvious ecological heterogeneity.[40]

At the end of the Mesoproterozoic and beginning of the Neoproterozoic, multicellular organisms appear, and some of them can be related to extant clades, such as Bangiophyte red algae, Vaucheriales algae, and possibly fungi.

The oldest process-bearing organic-walled microfossil, *Tappania plana*, was first described in the Ruyang Group of China,[57,58] then reported in India[59] and Siberia.[60] In northern Australia, very well preserved populations of *Tappania* occur in the well-dated (1492 ± 3Ma to 1429 ± 31 Ma) Roper Group, northern Australia (ref. 40 and refs. therein). Population of *Tappania plana* comprises 20 to 160 μm vesicles that bear 0 to 20 or more heteromorphic processes (elongated extensions with variable morphology and length-25 to 60 μm) distributed asymmetrically about the vesicle surface (Fig. 1: 1-4). Processes communicate freely with the vesicle interior, have dark slightly expanded closed ends (Fig. 1: 1-3) and may branch (Fig. 1: 4). Specimens may also bear up to 3 bulbous protuberances suggesting reproduction by budding (Fig. 1: 4). Opening at the end of neck-like extensions could be a sophisticated excystment structure (Fig. 1: 1). No prokaryote, at my knowledge, produces at once acid-resistant preservable walls, ornamentation and large size. The vesicle size; appendage diameter, number, distribution and morphology (branching, heteromorphic); and the presence of neck-like extensions in *Tappania* indicate an early eukaryotic cell with a cytoskeleton.[40,16] Futhermore, the irregular morphology and asymmetric distribution of processes and the budding in *Tappania* suggest that *Tappania* might have been an actively growing cell or germinating cyst rather than a metabolically inert spore as most Paleozoic acritarchs (organic-walled microfossils) are assumed to represent.[40]

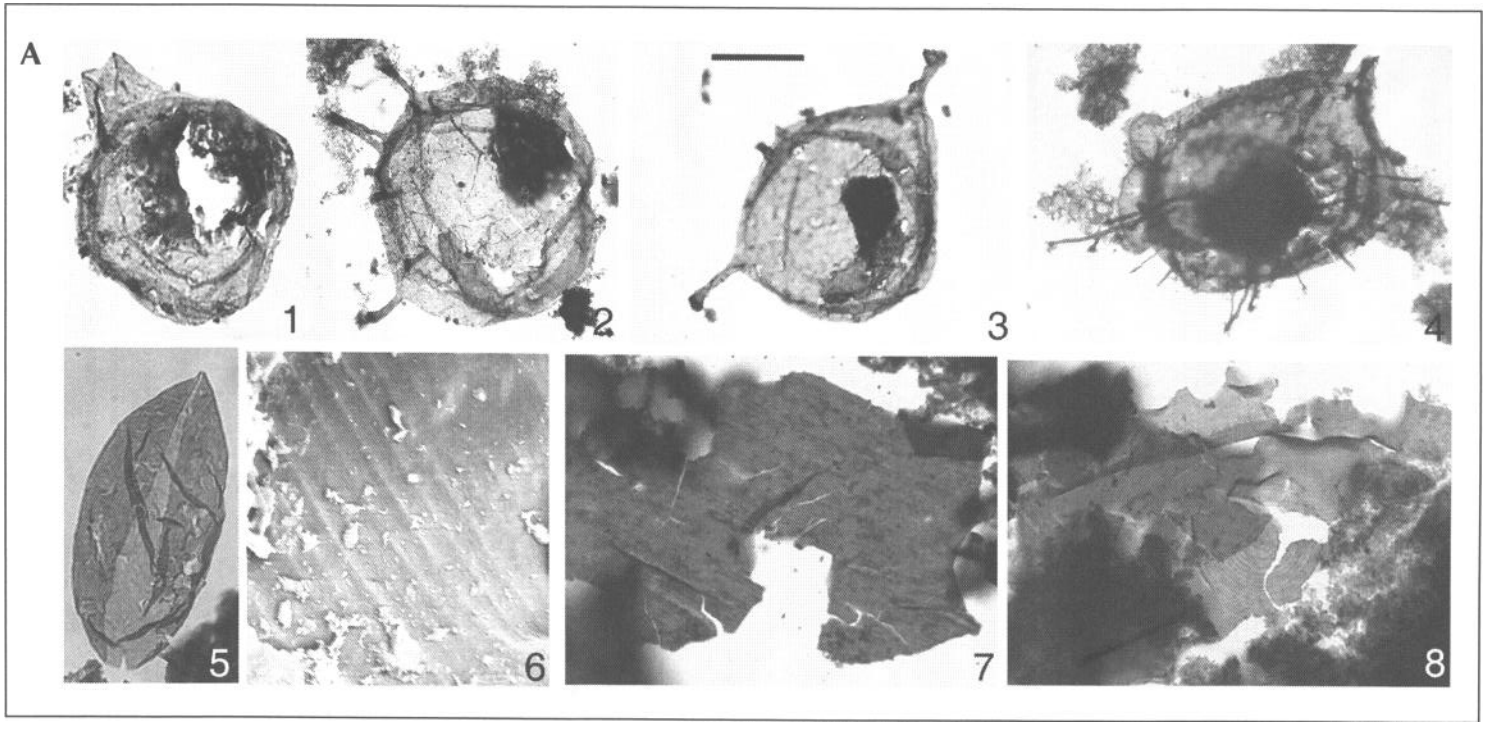


Figure 1. Late Paleoproterozoic-early Mesoproterozoic eukaryotic microfossils. A) 1-4) *Tappania plana* from the 1.5 Ga Roper Group, Australia; 1-4) light microscopy; 1) specimen with possible excystment structure (opening at the end of neck-like extension); 2-4) specimens with heteromorphic processes distributed asymmetrically about the vesicle; 4) specimen with a branched process and bulbous protrusions interpreted as budding structures. 5-8) *Valeria lophostriata*, 1.5 Ga Roper Group, Australia (5-6) and 1.65 Ga Mallapunyah Fm, Australia (7,8). 5) partially enrolled half vesicle, likely resulting from medial split (light microscopy); 6-8) ornamentation of closely spaced parallel ridges on the inner wall surface in SEM (6) and light microscopy (7,8). Scale bar for 1 = 32 μm; 2, 3 = 38 μm; 4, 5 = 52 μm; 6 = 5 μm; 7, 8 = 20 μm. Figure continued on next page.

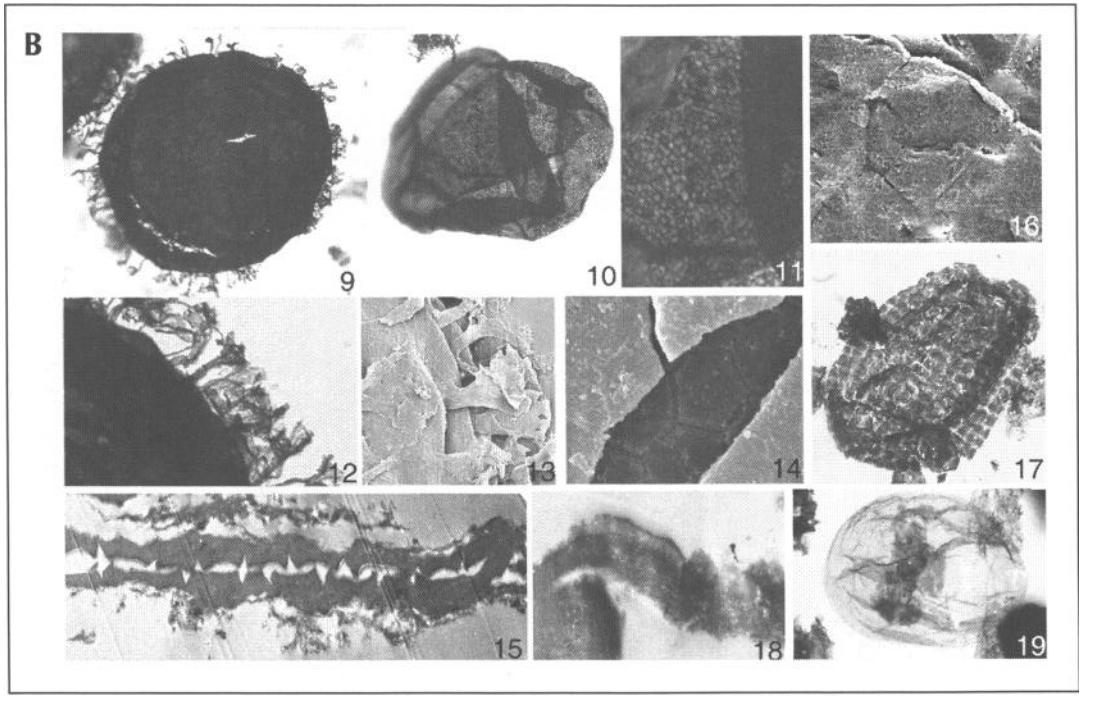


Figure 1, continued. B) 9-15) *Shuiyousphaeridium macroreticulatum* from the ~1.3 Ga Ruyang Group, northern China. 9-12) Light photographs showing specimen with numerous regularly spaced cylindrical processes that flare outward (9,12), specimen without processes (10) and a reticulated surface (11); 13-14) SEM showing detail of flaring furcating processes (13), and inner wall surface of closely packed, beveled hexagonal plates (14); 15) TEM showing the two walls of compressed acritarch and multilayered wall comprising a thick electron-dense homogeneous layer of organic plates surrounded by an outer layer of debris and processes and a thin electron-tenuous layer that lines the inner side of plates; 16-17) *Satka favosa*, Roper Group, Australia; 16) SEM showing detail of juxtaposed polygonal plates; 17) specimen showing wall of polygonal plates (light microscopy); 18-19) *Leiosphaeridia tenuissima*, 18) SEM showing multilayered wall with external layers often removed by degradation; 19) specimen showing wall with thin folds (light microscopy). Scale bar (in Fig. 1A:3) for 11, 12, 13 = 20 μm; 9 = 34 μm; 10 = 23 μm, 14 = 1.7 μm; 15 = 1.3 μm, 16 = 7.5 μm, 17 = 48 μm, 18 = 1.5 μm, 19 = 50 μm.

Another fossil, *Satka favosa*, also occurs in the Roper Group, and in other Mesoproterozoic successions (e.g., see refs. 61,62), has a wall made of interlocking polygonal plates 10-15 μm in maximum dimension (Fig. 1: 16-17). No prokaryotic cells build comparable walls.

Among Roper microfossils, three "species" (form taxa) of leiosphaerids showed multilayered heterogeneous wall ultrastructures under TEM (Fig. 1: 18-19). Such fine structure is common among extant protists that make acid-resistant preservable walls of comparable size and morphology, but distinct from the most likely prokaryotic candidates, envelope-forming cyanobacteria.[55]

Among Proterozoic ornamented microfossils, the large (100-250 μm in diameter) microfossil *Shuiyousphaeridium macroreticulatum* from the Ruyang Group, China, has remarkable morphology characterized by a reticulated surface (Fig. 1: 11) and numerous regularly spaced cylindrical processes that flare outward (Fig. 1: 9,12,13).[40,55,57,58,63] Close SEM examination of its wall shows that the wall's outer surface is covered with ridges that delimit granular polygonal fields. The inner wall surfaces show the reverse of the same ornamentation—a wall structure consisting of closely packed, beveled (2 μm across) hexagonal plates (Fig. 1: 14).[55] TEM images show that the ca. 1.5 μm wall is multilayered. A 391-586 nm thick, electron-dense, homogeneous layer of organic plates lies between an outer layer of debris and sectioned processes and a thin electron-tenuous layer that lines the inner side of the plates (Fig. 1: 15).

Microchemical analyses and wall ultrastructure of *Shuiyousphaeridium macroreticulatum* did not show evidence for a green algal (as suggested by ref. 64) or a fungal affinity (as suggested by ref. 65).[50,55] Recent analysis of whole rock samples containing *Shuiyousphaeridium* yielded dinosterane,[66] although it is unknown if the biomarker came from the microfossils or from other, unpreserved, organisms. This species may also show medial split excystment structures, suggesting a cyst-like morphology, but whether it was a metabolically inert stage of a unicellular (or multicellular ?) organism, or whether it had a phototrophic or heterotrophic metabolism, is unknown, as underlined by Butterfield.[65] In summary, *Shuiyousphaeridium* is a unicellular eukaryote with unknown biological affinity and could represent an extinct group, or the ancestor of extant taxa, possibly but not conclusively related to alveolate ancestors.

Radiometric dates, and chemo- and litho-stratigraphy suggest that Ruyang shales are older than ca. 1250 Ma.[58,66] Other microfossils in the Ruyang assemblage occur also in the well-dated 1492+/-3 Ma Roper Group, of northern Australia.

Prokaryotes can be large, they can have ornamentation, and they can have preservable walls (at least cyanobacteria sheaths),[16] but no prokaryote currently known has all three (large size, ornamentation, preservable acid-resistant walls) at once. And none exhibits the complexity of form from light microscopy, SEM, and in some cases TEM, observed in *Shuiyousphaeridium macroreticulatum, Valeria lophostriata, Tappania plana,* and *Satka favosa.* Many eukaryotes do. Therefore, these microfossils display characters of a eukaryotic grade of organization, and are interpreted as eukaryotes with a sophisticated cytoskeleton.

Towards the end of the Mesoproterozoic, other extraordinary fossils are preserved in cherts of the Hunting Formation, Arctic Canada, and consist of population of abundant *Bangiomorpha pubescens.*[67,68] This fossil includes vertically oriented (15-45 μm wide) uniseriate and (30-67 μm wide) multiseriate unbranched filament up to 2 mm long, surrounded by a translucent outer wall, and attached to a firm substrate by a lobate multicellular holdfast structure (Fig. 2: 1). The ontogeny of *Bangiomorpha* includes single-celled, double-celled stages, and four-to eight-celled stage where the wedge-shaped cells are arranged along a radial symmetry. The holdfast starts to differentiate at the 12-16 cell stage and is attached by a single cell to the remainder of the filament. At some more mature stage, the cells of some filaments undergo longitudinal intercalary division giving rise to multiseriate filaments, followed in some cases by tertiary division leading to close-packed spheroidal cells interpreted as fertilized carpospores.[68]

This pattern of longitudinal intercalary cell (or pie-like) division is only known in modern Bangia (Bangiales Rhodophytes). Stigonematales cyanobacteria do produce a comparable thick sheath and multiseriate filaments, and differentiated multicellular holdfast; but they are also characterized by branching, apical growth and differentiated cells (akinetes and heterocysts),

characters not found in *Bangiomorpha*.[68] No cyanobacteria nor green algae show radial intercalary division (one modern prasiolalean chlorophyte does but the radial cells are not arranged as a fourfold symmetry and they are transient, dividing in three planes into parenchymatous spheres).[68] Taphonomy also points to a noncyanobacterial affinity.[18] Therefore, on the basis of diagnostic fourfold radially symmetrical arrangement of wedge-shaped cells, *Bangiomorpha* is interpreted as a bangiophyte alga, although it differs in some aspects from modern *Bangia* by having a multicellular holdfast rather than rhizoids (but other filamentous non bangiales bangiophytes do have a comparable holdfast).[68]

Bangiomorpha pubescens is the oldest taxonomically resolved eukaryote so far, and records the evolution of complex multicellularity, cell differentiation, and sexual reproduction. It grew attached to a firm substrate, in an upright position in the shallow-water to intertidal environments. The fossils are preserved in cherts of the Hunting Formation, Somerset island, Arctic Canada, correlated with units on nearby Baffin island and Greenland dated at 1198 ± 24 on the basis of chemostratigraphy, litho- and bio-stratigraphy.[29]

At 1 Ga, *Palaeovaucheria*, a xanthophyte from the recently well-dated Lakhanda Formation, Siberia[69-71] indicates the appearance of stramenopiles (which include diatoms, xanthophytes, and brown algae) and of secondary symbiosis (involving a red alga-like endosymbiont). Populations of *Paleovaucheria* display morphological traits characteristic of vaucherian xanthophytes such as branching at right angles, 2 sizes of filaments on the same individual, and terminal pores and septae at filament ends[71] (Fig. 2: 3). Vaucherians recently discovered in cc.700-800 Ma shales in Spitsbergen display a more complete range of vaucherian morphologies[72] (Fig. 2: 4).

Neoproterozoic Record (1 to 0.54 Ga)

The Neoproterozoic records further diversification of multicellular organisms, including fungi, and green, red and vaucheriale algae; and the advent of biomineralization and predation. The end of the era is marked by the first diversification of animals.[73,74]

The 1100-1005 Ma Lakhanda Fm and the 1000-800 Ma Miroyedikha Fm of Siberia include ornamented and process-bearing acritarchs as well as filamentous multicellular microfossils, networks of cell, and possible fungi.[69]

The 850-750 Ma fossils "*Tappania*" discovered by Butterfield[65] in the Wynniatt Fm, Canada, are remarkable in their exquisite preservation and morphological complexity including features such as serial septae in the hollow, branched processes capable of secondary fusion (interpreted as hyphal fusion) (Fig. 2: 5-8) possibly forming a character combination synapomorphic of the higher fungi. These fossils have been compared to the older 1.5 Ga Roper Group and 1.3 Ga Ruyang Group *Tappania* but these have neck-like extensions unknown in the Wynniatt "*Tappania*" and do not show filaments fusion.[40,51,57] For these reasons, the two Mesoproterozoic and Neoproterozoic populations are best regarded as distinct taxa.[18] More detailed work on the older material is needed to uncover the range of morphologies displayed by the Neoproterozoic fossils, as well as microchemical analyses (C isotopes, biopolymers) of both populations.

Proterocladus, a siphonocladalean chlorophyte alga (based on long and irregular cell lengths, branching, intercellular septae and reproductive structures) (Fig. 2: 2) from the ca. 750 Ma Svanbergfjellet Formation of Spitsbergen,[75] suggests that chlorophyte diversification was well advanced by the mid-Neoproterozoic.

Based on aperture and test morphology, vase-shaped microfossils (VSM) have been related to filose and lobose amoebae from >742 ± 6 Ma rocks of the Chuar Group, Arizona (Fig. 2: 9), and other contemporaneous successions.[76,77] Other VSMs with honeycomb-patterned walls have been interpreted as casts of euglyphid amoebae with originally mineralized siliceous scales (Fig. 2: 10-11). These protists provide a firm calibration point for the supergroup opisthokonts, the clade that includes animals, fungi and the amoebozoans,[77-79] not to mention direct evidence for heterotrophic eukaryotes and eukaryotic biomineralization.[76] Hemispherical holes in some of these testate amoebae might record the earliest traces of predation (Fig. 2: 12).[77] Other evidence for biomineralization includes 0.65 Ga siliceous scales of chrysophytes or diatoms.[80,81]

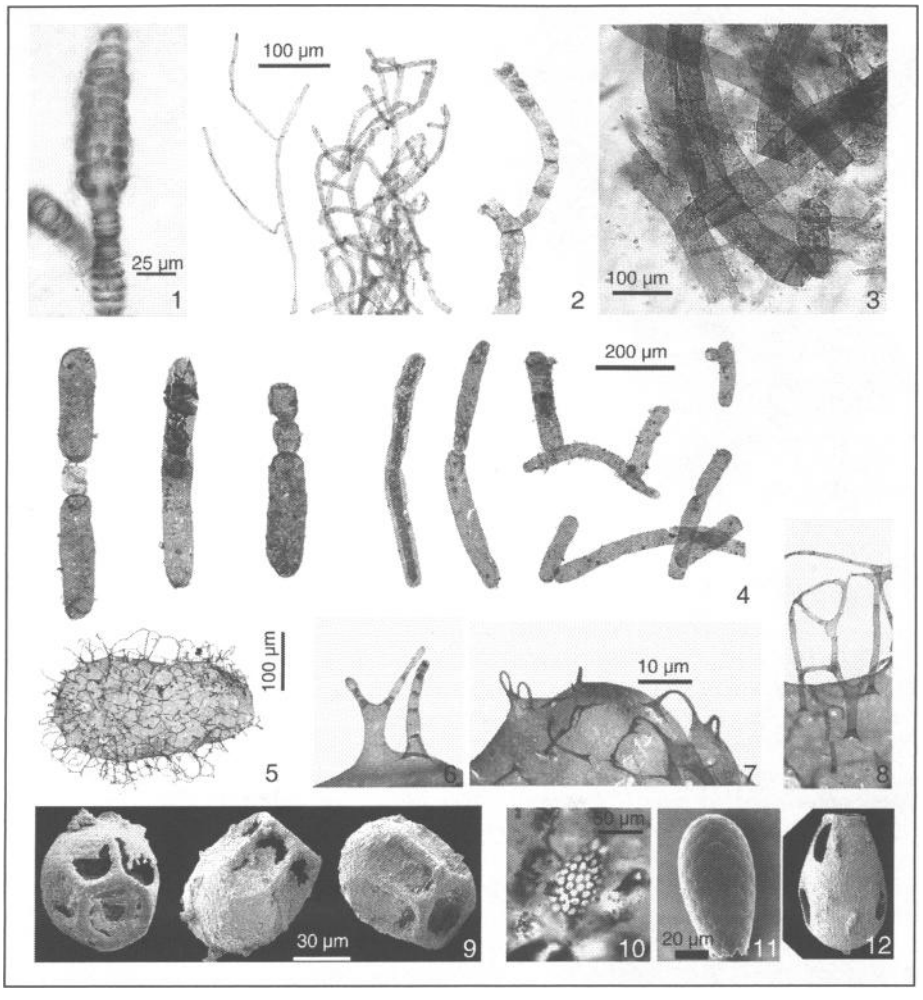


Figure 2. Neoproterozoic eukaryotic fossils relative of extant clades. (1) *Bangiomorpha pubescens*, from the ~1.2 Ga Hunting Formation, arctic Canada, showing radial division of cells within uniseriate filaments, interpreted as Bangiophyte red alga; (2) *Proterocladus* sp., from the 750 Ma Svanbergfjellet Formation, Spitsbergen, interpreted as Cladophorale green alga; 3-4) *Paleovaucheria clavata* from the ~1 Ga Lakhanda Formation, Siberia (3), and the ~750 Ma Svanbergfjellet Formation, Spitsbergen (4) interpreted as a vaucheriacean alga; 5-8) "*Tappania plana*" from the ~800 Ma Wynniatt Formation, arctic Canada, a complex multicellular form with septate (6), anastomosing processes (8) (light microscopy) interpreted as possible fungi; 9,10,12) vase-shaped microfossils form the ~750 Ma Chuar Group, Arizona, testate amoebae, 9) *Melanocyrillium hexodiadema*(SEM), interpreted as a lobose testate amoeba, 10) *Melicerion poikilon*, interpreted as a euglyphid testate amoeba; 11) modern analog *Euglypha tuberculata*; 12) VSM with holes in the test, possibly due to predation. Scale bar in 7 is also for 6 and 8; in 11 also for 12. Images 1,2,4,5-8, courtesy of N.J. Butterfield; images 9,10,12 courtesy of S.H. Porter, image 11 courtesy of R. Meisterfeld.

After the Sturtian (710-725 Ma), the Marinoan (635-600 Ma) and the Gaskiers (580 Ma) glaciations, the diversity of acritarchs with symmetrically distributed processes increases in the

Ediacarian (632-550 Ma)[82,83] and microscopic animals are preserved in phosphorites,[84] carbonates[85] and sandstones[74] around the world. The Doushantuo Formation of China, recently dated at ca. 598 ± 2 Ma,[86] hosts multicellular green, red and, possibly brown algae, as well as animal embryos, possible stem group cnidarians, and putative sponges.[73,84,87,88] Fossils interpreted as florideophyte red algae have differentiated medullary and cortical tissues with reproductive structures similar to carposporangia and spermatangia of living red algae.[73] Other Doushantuo fossils are attributed to *Porphyra*-like bangiophytes, to xanthophyte algae and possible green algae. The embryos are 500 μm spheres with an ornamented envelope that contains 1, 2, 4, 8 or more closely packed internal bodies with faceted sides and decreasing in size as their number increases, suggesting cells dividing by successive binary divisions in cleaving embryos.[73]

Just before the Cambrian, the late Ediacarian acritarch assemblages include only large (100s μm) unornamented sphaeromorphs, then diversity rises again in the Cambrian.[18]

Carbonaceous Compressions

Carbonaceous compressions are micrometer-thick remains of millimetre to centimetre-sized bodies. Their abundance increases through the Proterozoic, especially in the Neoproterozoic. Millimetric coiled septate filaments called *Grypania* are preserved as carbonaceous compressions in Mesoproterozoic rocks from China and the western United States.[89,91,92] These have been interpreted as eukaryotic multicellular organisms, possibly algae, based on their macroscopic size, helical morphology and the presence of transverse septae. Older fossils in the 1.87 Ga Negaunee Iron Formation, Michigan[93,94] have been related to *Grypania* based on a broad morphological resemblance (coiled shape) however the interpretation has been questioned, as there is no preserved microscopic detail.[95]

Prior to 1.2 Ga, besides *Grypania* mentioned above (also called *Sangshuania*), several forms have been reported in Paleoproterozoic shales from China[96,97] and interpreted as seaweeds. However these structures have been reinterpreted as irregular mat fragments (see ref. 18 and refs. therein). Bedding-plane structures (*Horodyskia*) resembling strings of beads occur in Mesoproterozoic rocks from Australia[98] and Montana[99,100] but their interpretation remains problematic, ranging from seaweeds, colonial metazoans, prokaryotic association, or nonbiological structures.

Neoproterozoic rocks yield diverse centimetric blades and closed tubes such as *Tawuia*, sausage-shaped compressions, and *Shouhsienia*, a ovate to spatulate form, and the blade-shaped *Longfengshania*.[101] The Ediacarian records casts and compressions of animals and perhaps other eukaryotic kingdoms.[74]

Summary of Early Eukaryotes Diversity

In summary, from 1.8-1.3 Ga, the diversity of organic-walled microfossils is moderate and includes smooth and ornamented sphaeromorphs, and vesicles with asymmetrically distributed processes. From 1.3 to 0.7 Ga, the diversity of protists increases and the assemblages also includes multicellular algae and fungi, and testate amoebae. The late Mesoproterozoic-early Neoproterozoic fossil record thus includes members of all extant supergroups: the Opisthokonts, the Amoebozoa, the plants, the chromalveolates, and the Rhizaria; except the Excavates. After the glaciations, the Ediacarian (632-550 Ma) records a burst in diversity of acritarchs with symmetrically distributed processes and macroscopic compressions, and animals are preserved. The late Ediacarian (550-542 Ma) yield only large smooth vesicles. In the Cambrian, diversity of ornamented and process-bearing acritarchs increases tremendously, in parallel with animal diversification.[18]

Dating Biological Innovations

The morphology of the fossils described above permits in rare cases, when they present taxonomically specific traits, to relate them to extant clades. In most cases, only the domain,

Eucarya or Bacteria, can be determined based on defined criteria.[16] However, the morphology of the microfossils can help pinpoint the evolution of major biological innovations.[18] This information sheds light on early eukaryote evolution, and could permit calibration of molecular phylogenies based on genes involved in these functional or biochemical innovations.

Table 1 summarizes through time the morphological and biomarker evidence for early eukaryotes (some of them are taxonomically resolved based on unique characters), and the biological innovations and their implications. Starting in the Archean through the Proterozoic, several important biological milestones are documented in the fossil record. Fossils and biomarkers provide minimum age for these evolutionary events, since phenotypes might be expressed later than genomic changes or might not be preserved in the geological record.

The Biosynthesis of Eukaryotic Sterols

The earliest evidence for the evolution of eukaryotic-like sterol pathway is dated at 2.77 Ga. Debates exists whether these sterols are restricted to eukaryotes (see above). Sterols make the cell membrane more flexible, a necessary innovation to permit, with the evolution of a dynamic cytoskeleton, phagocytosis. Phagocytosis was a crucial step for allowing later evolution of grazing, predation, and photosynthetic eukaryotes (by endosymbiosis of a chloroplast ancestor). Interestingly, the synthesis of sterols requires a minimum amount of oxygen.[25] This synthesis does not mean necessarily the evolution of the nucleus. However, it is also possible, but unproven, that early eukaryotic cells were aerobe and had already nucleus and mitochondria.[8]

The Synthesis of Resistant Biopolymers

Microfossils are extracted from shales using strong acids (HCL, HF) that dissolve the rocks and leave the organic-walled fossil intact. Prior to 1.8 Ga, there is no record so far of this type of fossils, but microfossils are rather preserved by mineralization in chert, phosphates, carbonates, from which they can rarely be extracted.

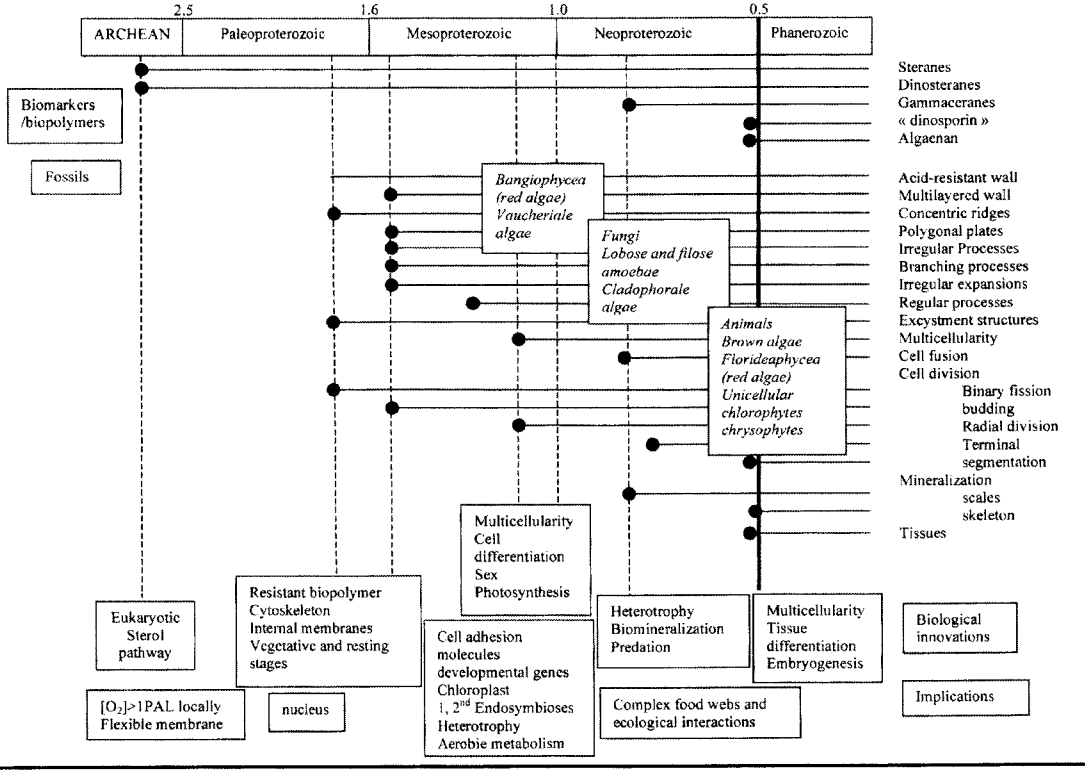
This feature may indicate that either the fossil record is incomplete, the older fossiliferous shales still await to be discovered, or that early cells did not produce acid-resistant walls prior to 1.8 Ga. In any case, by 1.8 Ga, cells were able to synthesize biopolymers probably resistant to a range of physicochemical conditions or to survive low nutrient concentration, as extant walled organisms do today. These biopolymers resist also to strong acids in the lab and permit micropaleontologists to extract the fossils from rocks.

Cytoskeleton and Internal Membranes

Following Cavalier-Smith's[26] (p.37) description of eukaryotic attributes; "cysts with spines or reticulate surface sculpturing would probably have required both an endomembrane system and a cytoskeleton, the most fundamental features of the eukaryotic cell, for their construction". Ornamented walls with polygonal organic plates and/or processes appear in the early Mesoproterozoic. The oldest process-bearing microfossil is *Tappania plana* at ~1.5 Ga. The irregular morphology and asymmetric distribution of processes, and the presence of budding structures, suggest that *Tappania* was an actively growing vegetative cell or germinating structure rather than a metabolically inert spore.[40] Some specimens with possible excystment structures would have been cysts. Prior to 1.5 Ga, the only ornamented fossils are vesicles ornamented with regularly-spaced concentric ridges (*Valeria lophostriata*). Such wall structure is common among extant protists that make preservable walls of comparable size and morphology, but distinct from the most likely prokaryotic candidates, envelope-forming cyanobacteria or other bacteria, to my knowledge.[55]

Recent findings show that Bacteria contain cytoskeletal elements.[102] Distant bacterial homologues of tubulin (FtsZ) and actin (MreB and ParM) not only resemble their eukaryotic counterparts structurally but also show similar functional characteristics. With the discovery of crescentin, a cell-shape-determining protein that resembles eukaryotic intermediate filament proteins, the third major cytoskeletal element has now been identified in bacteria as well.[103]

Table 1. Biological innovations in early eukaryotes: inference from the fossil record. The biomarker, biopolymer and fossil evidence in the Proterozoic provide direct evidence for biological innovations in early eukaryotes, with implications for the evolution of the eukaryotic cell. Dates are minimum ages. See detailed explanations and references in the text.



However, motor proteins involved in intracellular transport along the filamentous components of the eukaryotic cytoskeleton are unknown in Bacteria. What is determining (and limiting) the morphological plasticity of Bacteria (compared to protist)? The molecular mechanisms underlying the generation and maintenance of bacterial cell shape remain largely unresolved.[104] Known bacteria do not produce complex morphologies like protists do. Bacteria can reach relatively large size but they cannot produce large ornaments (these are often proteinic or nannoscale polysaccharide structures) and preservable walls all at once.[16]

Therefore, the complexity shown by some of the Proterozoic fossils, present in extant protists but unknown in prokaryotes, demonstrates the evolution of the eukaryotic cytoskeleton by ~1.8-1.5 Ga.

Life Cycle: Vegetative and Resting Stages

Excystment structures consisting of simple medial split in populations of leiospheres (simple sphaeromorphs) appear in the late Paleoproterozoic (~1.8 Ga). These structures may document the existence of resting stages, thereby implying the evolution of a life cycle of at least two (vegetative and resting) stages. Vegetative cells or germinating cysts are represented in early Mesoproterozoic populations of *Tappania plana*, which show budding, variable morphology and also possible complex excystment structures (opening at the end of a neck-like expansion). In the Neoproterozoic, an ornamented sphaeromorph (*T. laufeldii*) shows circular opening interpreted as excystment structures called pylomes.[105] Every protozoan phylum is able to produce resting cysts, suggesting that it could have been present in the ancestral eukaryote, possibly for protection against starvation.[106]

Multicellularity and Cell Differentiation

Multicellularity occurs in prokaryotes and eukaryotes. Multicellularity arose first in aquatic environment, possibly resulting from cells failing to separate after division, such as in filamentous bacteria and algae.[107,108] The next step involves cellular differentiation that permits division of labor. Stigonematales cyanobacteria produce multiseriate filaments, differentiated multicellular holdfast, branching filament, apical growth, and differentiated cells. These cyanobacteria produce specialized cells (heterocysts) for fixing nitrogen and photosynthesizing cells in the same filament. They can also make resistant spores called akinetes, which are preserved in 2.1 Ga rocks (review in ref. 29). These microfossils provide then a minimum age for cell differentiation and multicellularity. Thus, prokaryotes can exhibit multicellularity and cell differentiation, but their complexity is limited. Eukaryotic multicellularity permits a larger increase in organism size, higher-order emergent structures and development of tissue to accomplish more complex functions.[68] The size increase can give advantages such as better protection against predation, or increased speed and efficiency of colony locomotion (e.g., Volvox algae, dictyostelids slime molds), allowing migration to a more favourable environment.[108] With tremendous possibilities in morphological complexity and size, multicellular eukaryotes produced a new biological environment with complex ecological interactions.[68]

The 1.8-1.4 *Grypania* (macroscopic coiled filament) might record a coenocytic-grade of multicellularity. But the oldest example of complex eukaryotic multicellularity and cell differentiation is a population of ~1.2 Ga bangiophyte red algae (*Bangiomorpha*), dividing by radial intercalary division where the radial cells are arranged as a fourfold symmetry, and producing holdfast, uniseriate and multiseriate unbranched filaments and packed cells interpreted as carpospores.[68] The anchoring structure and upright position of this multicellular alga implies a new dimension in early ecosystems: tiering, that is, vertical diversification of ecological niches (rather that simply horizontal distribution of environments linked mostly to varying physicochemical conditions). Later record of eukaryotic multicellularity includes 1 Ga and 0.85 Ga vaucheriale algae, 0.85 Ga possible fungi, 0.75 Ga cladophorale algae and 0.75 Ga taxonomically unresolved forms such as *Valkyria* that has six cell types.[75] The 1 Ga Lakhanda Fm in Siberia includes several multicellular filamentous algae and cellular networks of unknown affinities.[69]

Large fossilized microfossils and some macroscopic compressions, could record coenocytic (multinucleate but noncellular) organisms, which could be considered as a primitive or derived form of multicellularity. Tissue-grade multicellularity or parenchymatous construction appears at the end of the Proterozoic, as shown by exquisitely preserved ~600 Ma florideophyte red algae.[73] Compressions, ichnofossils (traces) and fossils of multicellular organisms, including metazoans, abound in Ediacarian rocks in latest Proterozoic.

Sex

As mentioned above, based on comparative morphology with extant relative bangiophytes, including the possible presence of carpospores, *Bangiomorpha* also records the evolution of sexual reproduction by 1.2 Ga.[68] Since they are unicellular organisms reproducing sexually, sex antedated multicellularity.[67] Sex and multicellularity are independent innovations, that occurred in several groups, probably by the late Mesoproterozoic-early Neoproterozoic. Sex (or syngamy, nuclear fusion and meiosis) might have evolved very early in eukaryote evolution, as soon as the evolution of a flexible cell membrane and of an internal cytoskeleton would have facilitated membrane fusion and cell merger, the basis for syngamy.[109]

Photosynthesis

Earliest evidence for photosynthesis is given again by *Bangiomorpha.* This fossil is a multicellular red alga, implying a preceding evolution of unicellular photosynthetic eukaryotes, and thus the evolution of phagocytosis for endosymbiosis of the chloroplast ancestor.[11] Based on biomarker evidence,[24] the antiquity of eukaryotes and cyanobacteria may be 2.77 Ga, making theoretically possible the evolution of chloroplast by endosymbiosis. Geological evidence for atmosphere oxygenation and undisputable fossil evidence for cyanobacteria (see review in ref. 2) indicate that this endosymbiotic event could have occurred from 2.32- 2.1 Ga. Therefore, prior to the 1.2 Ga red algae, the rock record preserves several organic-walled microfossils since about 1.8 Ga, probably including both heterotrophic and photosynthetic protists. Later photosynthetic eukaryotes include 1-0.85 Ga Xanthophyte algae, 750 ma Cladophorale algae[69,72,75] and 600 Ma acritarch *Tanarium* with a wall made of the algal biopolymer algaenan.[50]

Heterotrophy

Earliest direct evidence come from early Neoproterozoic (0.85 Ga) fungi from the Wynniatt Formation, Canada[65] and lobose and filose testate amoebae from the Chuar Group, Arizona.[76] The record of photosynthetic eukaryotes since 1.2 Ga implies preceding evolution of heterotrophic organisms, most probably preserved among Paleo- and Mesoproterozoic protists.

Biomineralization

Biomineralization is first recorded by casts of 0.75 Ga euglyphid amoebae (siliceous scales), then 0.65 Ga siliceous scales of chrysophytes or diatoms. In the late Neoproterozoic, animals with carbonate skeleton form large reefs, and sponge spicules and calcified tubes are preserved (see review in ref. 110).

Predation

Hemispherical holes in 0.78-0.75 Ga amoebae might record traces of predation by other microorganisms.[77] This would indicate complex microbial food webs. From 600-632 Ma, Ediacarian animals added another scale to predation.

Tissue Differentiation

Another degree of complexity is reached with the arrangement of cells into 3-D parenchymatous tissues, as recorded in ca. 0.6 Ga floridophyte red algae[111] and in Ediacarian metazoan compressions.[74]

Embryogenesis

Phosphatized embryos are preserved in 0.6 Ga rocks and record several steps of egg segmentation up to the stereoblastula stage.[73] These fossils imply the evolution of complex developmental pathway.

Conclusions

Major biological innovations in eukaryote evolution occurred in the Precambrian. Fossils and biomarkers record these innovations, showing that clade origination and diversification were Proterozoic events. The Phanerozoic sees diversification within the clades, and major innovations such as eukaryote invasion of land and intelligence.

Early eukaryotes had developed many complex and characteristic cellular and molecular mechanisms by 1.2 Ga, but reached high diversity of macroscopic forms only at the end of the Proterozoic era. The causes of this relatively late or delayed diversification are probably multiple, including the chemistry of early atmosphere and oceans and lack of predation pressure until the Ediacarian when increased oxygenation permitted the evolution of animals (possibly already appeared) and consequently the complexification of ecosystems (see discussions in refs. 17,18,112).

Innovative analyses of the fossils using combined microscopic and micro-chemical techniques (see refs. 50,51,55 and refs. therein, and refs. 113,114) will clarify the biological affinities of early eukaryotes. Increased resolution of the geochemical and stratigraphical record will precise the physicochemical evolution of early environments where eukaryotes appeared and diversified. Molecular and cellular biology, looking at genes and ultrastructural elements involved in these major biological innovations, will provide improved phylogenies to test with the fossil record.

Collaborative research, involving geologists, paleontologists, chemists and biologists, will undoubtedly improve our understanding of early eukaryote evolution.

Acknowledgments

I am thankful to Gáspar Jékely for inviting me to write this chapter, and for Andrew Knoll to recommend my participation. Thanks also to Nick Butterfield, Susannah Porter and Ralph Meisterfeld who provided pictures for Figure 2.

References

1. Margulis L. Symbiosis in Cell Evolution. WH Freeman: San Francisco, 1981.
2. Knoll AH. Life on a Young Planet. Princeton: Princeton University Press, 2003:277.
3. Martin W, Müller M. The hydrogen hypothesis for the first eukaryote. Nature 1998; 392(6671):37-41.
4. Moreira D, Lopez-Garcia P. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. J Mol Evol 1998; 47(5):517-530.
5. Hartman H, Federov A. The origin of the eukaryotic cell: A genomic investigation. Proc Natl Acad Sci USA 2002; 99(3):1420-1425.
6. Xu Y, Glansdorff N. Was our ancestor a hyperthermophilic procaryote? Comp Biochem Physiol A Mol Integr Physiol 2002; 133:677-688.
7. Dyall SD, Johnson PJ. Origins of hydrogenosomes and mitochondria: Evolution and organelle biogenesis. Curr Opin Microbiol 2003:404-411.
8. Cavalier-Smith T. Only six kingdoms of life. Proc Biol Sci 2004; 271:1251-1262.
9. Simpson AGB, Roger AJ. The real kingdoms of eukaryotes. Curr Biol 2004; 14(17):693-696.
10. Katz LA. Changing perspectives on the origin of eukaryotes. Trends Ecol Evol 1998; 13(12):493-497.
11. McFadden GI. Primary and secondary endosymbiosis and the origin of plastids. J Phycol 2001; 37:951-959.
12. Berney C, Fahrni J, Pawlowski P. How many novel eukaryotic 'kingdoms'? pitfalls and limitations of environmental DNA surveys. BMC Biol 2004; 2:13
13. Keeling PJ, Burger G, Durnford DG et al. The tree of eukaryotes. Trends Ecol Evol 2005; 20(12):670-676.
14. Baldauf SL. The deep roots of eukaryotes. Science 2003; 300:1703-1706.

15. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436:1113-1118.
16. Javaux EJ, Knoll AH, Walter MR. Recognizing and interpreting the fossils of early eukaryotes. Orig Life Evol Biosph 2003; 33:75-94.
17. Porter SM. Early eukaryotic diversification. In: Lipps J, Waggoner B, eds. Neoproterozoic-Cambrian Biological Revolutions. Paleontological Society Papers, 2004:10:35-50.
18. Knoll AH, Javaux EJ, Hewitt D et al. Eukaryotic organisms in Proterozoic Oceans. Proc R Soc Lond B Biol Sci (in press).
19. Rosing MT. C-13-depleted carbon microparticles in >3700-Ma sea-floor sedimentary rocks from western Greenland. Science 1999; 283:674-676.
20. Shen Y, Buick R, Canfield DE. Isotopic evidence for microbial sulphate reduction in the early Archaean era. Nature 2001; 410:77-81.
21. Hayes JM, Kaplan IR, Wedeking KW. Precambrian organic chemistry, preservation of the record. In: Schopf JW, ed. Earth's Earliest Biosphere. Princeton University Press, 1983:93-134.
22. Moreira D, Lopez-Garcia P. The molecular ecology of microbial eukaryotes unveils a hidden world. Trends Microbiol 2002; 10(1):31-38.
23. Schulz HN, Brinkhoff T, Ferdelman TG et al. Dense populations of a giant sulfur bacterium in Namibian shelf sediments. Science 1999; 284:493-495.
24. Brocks JJ, Logan GA, Buick R et al. Archean molecular fossils and the early rise of eukaryotes. Science 1999; 285:1033-1036.
25. Brocks JJ, Buick R, Summons RE et al. A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 - 2.45 billion year old Mount Bruce Supergroup, Hamersley Basin, Western Australia, Geochim Cosmochim Acta 2003; 67(22):4321-4335.
26. Cavalier-Smith T. The neomuran origin of archaebacteria: The negibacteria root of the universal tree and bacteria megaclassification. International Int J Syst Evol Microbiol 2002; 52:7-76.
27. Hedges SB, Blair JE, Venturi ML et al. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol 2004; 4, (Art. No. 2).
28. Yoon HS, Hackett JD, Ciniglia C et al. A molecular time line for the origin of photosynthetic eukaryotes. Mol Biol Evol 2004; 21:809-818.
29. Javaux EJ. Extreme life on Earth-past, present and possibly beyond. Res Microbiol 2006; 157:37-48.
30. Douzery EJP, Snell EA, Bapteste E et al. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? Proc Natl Acad Sci USA, 2004; 101:15386-15391.
31. Pearson A, Budin M, Brocks J. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus. Proc Natl Acad Sci USA 2003; 100:15352-15357.
32. Volkman JK. Sterols and other triterpenoids: Source specificity and evolution of biosynthetic pathways. Org Geochem 2005; 36:139-159.
33. Volkman JK. Sterols in microorganisms. Appl Microbiol Biotechnol 2003; 60:495-506.
34. Raymond J, Blankenship RE. Biosynthetic pathways, gene replacement and the antiquity of life. Geobiology 2004; 2(4):199-203.
35. Canfield DE. A new model for Proterozoic ocean chemistry. Nature 1998; 396:450-453.
36. Shen Y, Knoll AH, Walter MR. Evidence for low sulphate and deep water anoxia in a mid-Proterozoic marine basin. Nature 2003; 423:632-635.
37. Kah LC, Lyons TM, Frank TD. Low marine sulphate and protracted oxygenation of the Proterozoic biosphere. Nature 2004; 431:834-838.
38. Brocks JJ, Love GD, Summons RE et al. Biomarker evidence for green and purple sulfur bacteria in an intensely stratified Paleoproterozoic ocean. Nature 2005; 437:866-870.
39. Li C, Peng P, Sheng GY et al. A molecular and isotopic geochemical study of Meso- to Neoproterozoic (1.73-0.85 Ga) sediments from the Jixian section, Yanshan Basin, North China. Precambrian Res 2003; 125(3-4):337-356.
40. Javaux EJ, Knoll AH, Walter MR. Morphological and ecological complexity complexity in early eukaryotic ecosystems. Nature 2001; 412:66-69.
41. Anbar A, Knoll AH. Proterozoic ocean chemistry and evolution: A bioinorganic bridge? Science 2002; 297:1137-1142.
42. Peng PG, Sheng JF, Yan Y. Biological markers in 1.7 billion year old rock from the Tuanshanzi Formation, Jixian strata section, North China. Org Geochem 1998; 29:1321-1329.
43. Moldowan JM, Jacobsen SR, Dahl J et al. Molecular fossils demonstrate Precambrian origins of dinoflagellates. In: Zhuralev AY, Riding R, eds. The Ecology of the Cambrian Radiation. New York: Columbia University Press, 2001:475-493.
44. Summons RE, Walter MR. Molecular fossils and microfossils from proterozoic sediments. Am J Sci 1990; 290-A:212-244.

45. Meng FW, Zhou CM, Yin LM et al. The oldest known dinoflagellates: Morphological and molecular evidence from Mesoproterozoic rocks at Yongji, Shanxi Province. Chin Sci Bull 2005; 50:1230-1234.
46. Pratt LM, Summons RE, Hieshima GB. Sterane and triterpane biomarkers in the Precambrian Nonesuch Formation, North American Midcontinent Rift. Geochem Cosmochim Acta 1991; 55:911-916.
47. Summons RE, Thomas J, Maxwell JR et al. Secular and environmental constraints on the occurrence of dinosterane in sediments. Geochem Cosmochim Acta 1992; 56:2437-2444.
48. Summons RE, Brassell SC, Eglinton G et al. Distinctive hydrocarbon biomarkers from fossiliferous sediment of the late proterozoic Walcott Member, Chuar Group, Grand-Canyon, Arizona. Geochem Cosmochim Acta 1988; 52(11):2625-2637.
49. Kleeman G, Poralla K, Englert G et al. Tetrahymenol from the phototrophic bacterium Rhodopseudomonas palustris: First report of a gammacerane triterpene from a prokaryote. J Genet Microbiol 1990; 136:2551-2553.
50. Marshall CP, Javaux EJ, Knoll AH et al. Combined micro-Fourier transform infrared (FTIR) spectroscopy and Micro-Raman spectroscopy of Proterozoic acritarchs: A new approach to palaeobiology. Precambrian Res 2005; 138:208-224.
51. Javaux EJ, Marshall CP. A new approach in deciphering early protist paleobiology and evolution: Combined microscopy and microchemistry of single Proterozoic acritarchs. Rev Palaeobot Palynol (in press).
52. Arouri K, Greenwood PF, Walter MR. A possible chlorophycean affinity of some Neoproterozoic acritarchs. Org Geochem 1999; 30:1323-1337.
53. Versteegh GJM, Blokker P. Resistant macromolecules of extant and fossil microalgae. Phycological Res 2004; 52:325-339.
54. Yan Y, Liu Z. Significance of eukaryotic organisms in the microfossil flora of Changcheng System. Acta Micropalaeontologica Sinica 1993; 10:167-180.
55. Javaux EJ, Knoll AH, Walter MR. TEM evidence for eukaryotic diversity in mid-Proterozoic oceans. Geobiology 2004; 2:121-132.
56. Hofmann HJ. Global distribution of the Proterozoic sphaeropmorph acritarch Valeria lophostriata (Jankauskas). Acta Micropaleontologica Sinica 1999; 16:215-224.
57. Yin L. Acanthomorphic acritarchs from Meso-Neoproterozoic shales of the Ruyang Group, Shanxi, China. Rev Palaeobot Palynol 1997; 98:15-25.
58. Xiao S, Knoll AH, Kaufman AJ et al. Neoproterozoic fossils in Mesoproterozoic rocks? Precambrian Res 1997; 84:197-220.
59. Prasad B, Asher R. Acritarch biostratigraphy and lithostratigraphic classification of Proterozoic and Lower Paleozoic sediments (Pre-Unconformity Sequence) of Ganga Basin, India. Paleontographica Indica 2001; 5:1-151.
60. Nagovitsin KE. Mikrofossilii i stratigrafiya verchnego Rifeya Yugo-Zapadnoi chasti Siberskoi Platformi. Ph.D. Thesis. RAS Siberian Branch, Institute of Geology, 2001:222.
61. Hofmann HJ, Jackson. Shale facies microfossils from the Proterozoic Bylot Supergroup, Baffin Island, Canada. J Paleontol 1994; 68(4):Memoir 37-39.
62. Jankauskas TV. Mikrofossilii dokembriya SSSR (Precambrian microfossils of the USSR). Nauka Leningrad 1989:1-190.
63. Yan Y, Zhu S. Discovery of acanthomorphic acritarchs from the Baicaoping Formation in Yongi, Shanxi, and its geological significance. Acta Palaeontologica Sinica 1992; 9:267-282.
64. Kaufman AJ, Xiao S. High CO2 levels in the Proterozoic atmosphere estimated from analyses of individual microfossils. Nature 2003; 425:279-282.
65. Butterfield NJ. Probable Proterozoic fungi. Paleobiology 2005; 31:165-182.
66. Yin LM, Yuan XL, Meng FW et al. Protists of the Upper Mesoproterozoic Ruyang Group in Shanxi Province, China. Precambrian Res 2005; 141(1-2):49-66.
67. Butterfield NJ, Knoll AH, Swett K. A bangiophyte red alga from the Proterozoic of Arctic Canada. Science 1990; 250:104-107.
68. Butterfield NJ. Bangiomorpha pubescens n. gen., n. sp.: Implications for the evolution of sex, multicelluarity and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. Paleobiology 2000; 26:386-404.
69. Herman N. Organic world one billion years ago. Leningrad, Nauka: 1990.
70. Rainbird RH, Stern RA, Khudoley AK et al. U-Pb geochronology of Riphean sandstone and gabbro from southeast Siberia and its bearing on the Laurentia-Siberia connection. Earth and Planetary Sciences Letters 1998; 164:409-420.

71. Woods KN, Knoll AH, Herman TN. Xanthophyte algae from the Mesoproterozoic/Neoproterozoic transition: Confirmation and evolutionary implications. Geological Society of America Abstracts with Programs 1998; 30:A232.
72. Butterfield NJ. A vaucherian alga from the middle Neoproterozoic of Spitsbergen: Implications for the evolution of Proterozoic eukaryotes and the Cambrian explosion. Paleobiology 2004; 30:231-252.
73. Xiao S, Zhang Y, Knoll AH. Three-dimensionally preservation of algae and animal embryos in a Neoproterozoic phosphate. Nature 1998; 391:553-558.
74. Narbonne GM. The Ediacara biota: Neoproterozoic origin of animals and their ecosystems. Ann Rev Earth and Planetary Sciences Letters 2005; 33:421-442.
75. Butterfield NJ, Knoll AH, Swett N. Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. Fossils and Strata 1994; 34:1-84.
76. Porter SM, Knoll AH. Testate amoebae in the Neoproterozoic era: Evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. Paleobiology 2000; 26:360-385.
77. Porter SM, Meisterfeld R, Knoll AH. Vase-shaped microfossils from the Neoproterozoic Chuar Group, Grand Canyon: A classification guided by modern testate amoebae. J Paleontol 2003; 77:409-429.
78. Baldauf SL, Roger AJ, Wenk-Siefert I et al. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 2000; 290:972-977.
79. Bapteste E, Brinkmann H, Lee JA et al. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. Proc Natl Acad Sci USA 2002; 99(3):1414-1419.
80. Allison CW, Awramik SM. Organic-walled microfossils from earliest Cambrian or latest Proterozoic Tindir Group rocks, northwest Canada. Precambrian Res 1989; 43:253-294.
81. Kaufman AJ, Knoll AH, Awramik SM. Biostratigraphic and chemostratigraphic correlation of Neoproterozoic sedimentary successions - Upper Tindir Group, northwestern Canada, as a test case. Geology 1992; 20:181-185.
82. Moczydlowska M. Acritarch biostratigraphy of the lower cambrian and the precambrian-cambrian boundary in southeast Poland. Fossils and Strata 1991; 29:1-127.
83. Grey K. Ediacaran palynology of Australia. Australian Association Palaeontologists Memoir 2005; 31:1-432.
84. Xiao S, Knoll AH. Phosphatized embryos from the Neoproterozoic Doushantuo Formation. J Paleontol 2000; 74:767-788.
85. Grotzinger JP, Watters W, Knoll AH. Calcareous metazoans in thrombolitic bioherms of the terminal Proterozoic Nama Group, Namibia. Paleobiology 2000; 26:334-359.
86. Barfod GH, Albarède F, Knoll AH et al. New Lu-Hf and Pb-Pb age constraints on the earliest animal fossils. Earth and Planetary Sciences Letters 2002; 201(1):203-212.
87. Xiao S, Knoll AH, Yuan X. Miaohephyton, a possible brown alga from the terminal Proterozoic Doushantuo Formation, China. J Paleontol 1998; 72:1072-1086.
88. Xiao S, Yuan X, Steiner M et al. Macroscopic carbonaceous compressions in a terminal Proterozoic shale: A systematic reassessment of the Miaohe biota, South China. J Paleontol 2002; 76:345-374.
89. Walter MR, Du R, Horodyski RJ. Coiled carbonaceous megafossils from the middle proterozoic of Jixian (Tianjin) and Montana. Am J Sci 1990; 290A:133-148.
90. Walter MR, Oehler JH, Oehler DZ. Megascopic algae 1300 million years old from the Belt Supergroup, Montana: A reinterpretation of Walcott's Helminthichnites. J Paleontol 1976; 50:872-881.
91. Kumar S. Megafossils from the Mesoproterozoic Rohtas Formation (The Vindhyan Supergroup), Katni Area, Central India. Precambrian Res 1995; 72(3-4):171-184.
92. Sarangi S, Gopalan K, Kumar S. Pb-Pb age of earliest megascopic, eukaryotic alga bearing Rohtas Formation, Vindhyan Supergroup, India: Implications for Precambrian atmospheric oxygen evolution. Precambrian Res 2004; 132(1-2):107-121.
93. Han TM, Runnegar B. Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee Iron Formation. Science 1992; 257:232-235.
94. Schneider DA, Bickford ME, Cannon WF et al. Age of volcanic rocks and syndepositional iron formations, Marquette Range Supergroup; implications for the tectonic setting of Paleoproterozoic iron formations of the Lake Superior region. Can J Earth Sci 2002; 39:999-1012.
95. Samuelsson J, Butterfield NJ. Neoproterozoic fossils from the Franklin Mountains, northwestern Canada: Stratigraphic and palaeobiological implications. Precambrian Res 2001; 107:235-251.
96. Zhu S, Chen H. Megascopic multicellular organisms from the 1700-million-year-old Tuanshanzi Formation in the Jixian area, North China. Science 1995; 270:620-622.
97. Zhu SX, Sun SF, Huang XG et al. Discovery of carbonaceous compressions and their multicellular tissues from the Changzhougou Formation (1 800 Ma) in the Yanshan range, North China. Chin Sci Bull 2000; 45(9):841-847.

98. Grey K, Williams IR. Problematic bedding-plane markings from the Middle Proterozoic Manganese Subgroup, Bangemall Basin, Western Australia. Precambrian Res 1990; 46:307-327.
99. Horodyski R. Problematic bedding-plane markings from the Middle Proterozoic Appekunny Argillite, Belt Supergroup, northwestern Montana. J Paleontol 1982; 56:882-889.
100. Yochelson EL, Fedonkin MA. A new tissue-grade organism 1.5 billion years old from Montana. Proc Biol Soc Wash 2000; 113:843-847.
101. Hofmann HJ. Proterozoic carbonaceous films. In: Schopf JW, Klein C., eds. The Proterozoic biosphere: A multidisiciplinary study. Cambridge: Cambridge University Press, 1992:349-357.
102. Graumann PL. Cytoskeletal elements in bacteria. Curr Opin Microbiol 2004; 7(6):565-571.
103. Moller-Jensen J, Lowe J. Increasing complexity of the bacterial cytoskeleton. Curr Opin Cell Biol 2005; 17(1):75-81.
104. Cabeen MT, Jacobs-Wagner C. Bacterial cell shape. Nat Rev Microbiol 2005; 3(8):601-610.
105. Knoll AH. Archaean and proterozoic palaeontology. In: Jansonius J, McGregor DC, eds. Palynology: Principles and Applications, Vol. 1. American Association of Stratigraphic Palynologists Foundation. Publishers Press Salt Lake City, 1996:51-80.
106. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 2002; 52:297-354.
107. Bonner T. The origins of multicellularity. Integrative Biology 1998:27-36.
108. Bonner T. On the origin of differentiation. J Biosci 2003; 28(4):523-528.
109. Cavalier-Smith T. Origins of the machinery of recombination and sex. Heredity 2002; 88:125-141.
110. Knoll AH. Biomineralization and evolutionary history. Rev Mineralog Geochem 2003; 54:329-356.
111. Xiao S, Knoll AH, Yuan X et al. Phosphatized multicellular algae in the Neoproterozoic Doushantuo Formation, China, and the early evolution of florideophyte red algae. Am J Botany 2004; 91:214-227.
112. Peterson KJ, Butterfield NJ. Origin of the Eumetazoa: Testing ecological predictions of molecular clocks against the Proterozoic fossil record. Proc Natl Acad Sci USA 2005; 102:9547-9552.
113. Talyzina NM. Ultrastructure and morphology of Chuaria circularis (Walcott, 1899) Vidal and Ford (1985) from the Neoproterozoic Visingsö Group, Sweden. Precambrian Res 2000; 102:123-134.
114. Schopf JW, Kudryavtsev AB. Three-dimensional Raman imagery of Precambrian microscopic organisms. Geobiology 2005; 3:1-12.

CHAPTER 2

The Diversity of Eukaryotes and the Root of the Eukaryotic Tree

Henner Brinkmann and Hervé Philippe*

Abstract

More than 15 years ago, on the basis of phylogenetic analyses of a handful of anciently duplicated genes and of rRNA, Carl Woese proposed both a eubacterial rooting of the Tree of Life and a stepwise evolution of the eukaryotic cell. An important part of Woese's paradigm was the assumption that the so-called Archezoa were considered to be genuinely primitive because they were lacking mitochondria and several other organelles characteristic for most eukaryotes. Since then, enormous progress have been accomplished in sequencing technology and in phylogenetic reconstruction. In particular, it is now clear that a tree reconstruction artefact, known as Long Branch Attraction, is responsible for the early emergence of the fast evolving Archezoa in the eukaryotic tree. The corollary hypothesis that all extant eukaryotes are ancestrally mitochondrial is strongly supported by the discovery of rudimentary mitochondrial organelles in all analysed Archezoa. Today a consensus that divides the extant eukaryotes into six major groups is replacing Woese's paradigm, which needs, however, further confirmation. Recently, a molecular dating study based on a large phylogenomic dataset with a relaxed molecular clock and multiple time intervals yielded in a surprisingly recent time estimate of 1085 Mya for the origin of the extant eukaryotic diversity. Therefore, extant eukaryotes seem to be the product of a massive radiation that happened rather late, at least in terms of prokaryotic diversity. In multiple cases evolution has proceeded via secondary simplification of a complex ancestor, instead of the constant march towards rising complexity generally assumed. Therefore it is time to reevaluate the origin and evolution of eukaryotes, in light of the newly established phylogeny, by further integrating secondary simplification as an equal partner to complexification.

Woese's Paradigm

The first molecular studies meant to uncover the evolutionary relationships among eukaryotes date back to the late eighties and were essentially based on phylogenetic analyses using the ubiquitous and highly conserved small subunit ribosomal RNA (SSU rRNA).[1,2] The eukaryotic SSU rRNA tree, generally rooted by the distantly related Archaea, displays several basal and deeply diverging lineages that progressively emerge followed by about ten (often multi-cellular) apical lineages called the "crown groups". Since the relative order of emergence of these latter groups was unresolved with short internal branches connecting them, it was suggested that the crown groups originated in a massive radiation.[3]

*Corresponding Author: Hervé Philippe—Canadian Institute for Advanced Research, Evolutionary Biology Program, Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Succursale CentreVille, 2900 Edouard Montpetit, Montréal, Québec H3C3J7, Canada, Email: herve.philippe@umontreal.ca

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

The unicellular organisms (often parasites) that emerged at the base of the rRNA tree are usually simple and lack organelles (in particular mitochondria) or systems characteristic for the vast majority of eukaryotes (e.g., Golgi apparatus).[4] Therefore they were considered to be genuinely "primitive", relicts of an ancient world essentially devoid of oxygen and representing ancestral organisation levels in the progressive complexification of the eukaryotic cell. This tree was in good agreement with the "Archezoa" hypothesis.[5] Archezoa were defined as primarily amitochondrial and peroxisome lacking eukaryotes, which were postulated to have evolved under anaerobic conditions before the great oxygenation event,[6] around 2,300 mya. Furthermore, the large distances observed within the eukaryotic SSU rRNA, especially for the Archezoa, led to the idea of a diversity of eukaryotes similar or even greater than the diversity of prokaryotes.[4]

The theory of the endosymbiotic origin of mitochondria and plastids, originally proposed by Mereschkowsky[7] was revived in the early 70s[8] and increasingly supported by molecular data. The identification of the bacterial groups from which they originated was first achieved through rRNA-based analyses that demonstrated a specific affinity between plastids and cyanobacteria[9] and between mitochondria and alpha Proteobacteria (then called purple bacteria).[10]

Although rRNA analyses showed the existence of the three domains of life (Archaea, Bacteria and Eukaryotes),[11] rRNA genes were not suited to address the question of their relationships, since there is no outgroup to root rRNA phylogenies. The discovery that a few conserved genes exist across all three domains of life as two distinct yet homologous copies allowed circumventing this problem. Indeed, being the descendants of an ancient gene duplication that had happened before the time of the "Last universal common ancestor" (LUCA) of the three domains, and these two paralogous copies could be used as reciprocal outgroup sequences. In 1989, two studies based on translation elongation factors EF-Tu and EF-G[12] and F and vacuolar H^+-ATPases[13] supported a sister-group relationship between Archaea and Eukaryotes, with the Bacteria emerging first. A few subsequent studies confirmed this initial result[14-17] but see.[18] Arguing in favour of an origin of eukaryotes from a "primitive" Archaeon, both the shape and the bacterial rooting of the Tree of Life rapidly became widely accepted.[1] However, rather than a simple gradual transformation from an archaeon to an eukaryote most scenarios are fusion-based hypotheses.[19-22] We call this view of the Tree of Life Woese's paradigm, it has been presented in many textbooks.

The persisting predominance of Woese's paradigm even in light of current data is rather surprising, since it is founded on: 1) a low number of genes (less than ten) that provide limited guarantee to represent the entire genome, 2) a limited number of species (quite often a single representative per major lineage, except for animals, fungi and plants), and 3) simple tree reconstruction methods, mainly distances computed with an unrealistic model of sequence evolution. The two latter points were established very early on as seriously decreasing the accuracy of phylogenetic inference.[23-25] In this chapter, we will give a rapid overview of progress in phylogenetic inference and show which parts of Woese's paradigm are affected by recent improvements and their impacts on biological deductions. Then we will present our current knowledge of the eukaryotic phylogeny and the most important open questions. Finally we will discuss some hypotheses that have been inspired by the new phylogenetic framework.

Progress in Tree Reconstruction Methods and in Sequencing: The Promise of Phylogenomics

The aim of tree reconstruction methods is to extract phylogenetic signal from a set of homologous characters. The basic unit of the phylogenetic signal is simply a substitution (i.e., a mutation fixed in the population) from one nucleotide to another. If a substitution has occurred in an ancestral species, all descendant species share a unique nucleotide that allows uniting them into a monophyletic group. The Maximum Parsimony method strictly follows this principle by selecting the tree that minimises the number of apparent substitutions.

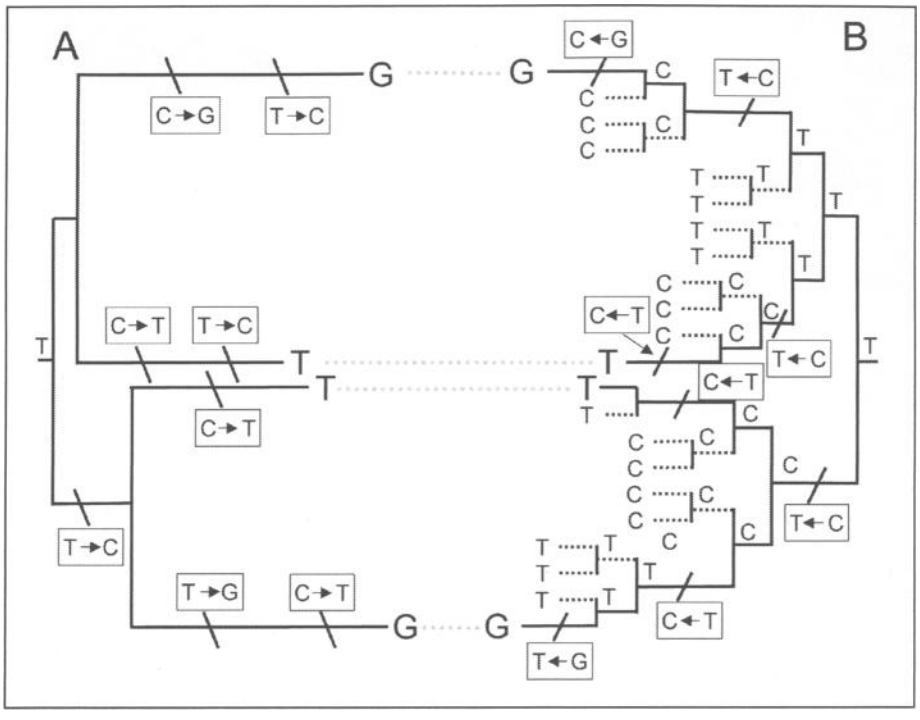


Figure 1. Long Branch Attraction and taxon sampling. A fast-evolving position is shown in (A). Because the top and bottom species evolve faster, they are more likely to converge by change towards the same nucleotide. Since based on four taxa it is impossible to deduce multiple internal substitutions, the inferred tree will likely group together the two fast evolving lineages because of the Long Branch Attraction artefact.[29] However, if more species are considered, the evolutionary history of this fast evolving position can be correctly deduced (B). This is why a dense taxon sampling by breaking long branches renders the phylogenetic inference less sensitive to artefacts.[24]

The difficulty of phylogenetic inference comes from the fact that many homologous positions in an alignment have undergone not a single but several substitutions. Therefore, distantly related species will eventually harbour identical character states (Fig. 1) which are not inherited from a common ancestor, a phenomenon called homoplasy. The species rich tree (Fig. 1B) allows the complete reconstruction of the evolutionary history, where as many as eight substitutions have occurred. In contrast, the incorrect tree that would be inferred based on four species assumes a single homoplastic substitution from T to G (tree not shown) that actually never took place. This kind of erroneous interpretation of multiple substitutions by tree reconstruction methods generates nonphylogenetic signal, i.e., support for an incorrect tree topology. The two most common tree reconstruction artefacts will now be introduced.

The first and most easily understandable artefact is due to the heterogeneity of nucleotide or amino acid composition. When two nonspecifically related lineages have independently acquired a similar nucleotide composition (for example, extremely A + T rich), they accidentally share the same nucleotide at numerous positions of the alignment due to their similar composition. These homoplasious positions are interpreted as evidence for substitutions in a hypothetical, however never existing, common ancestor. In cases when this nonphylogenetic, compositional signal is outnumbering the genuine phylogenetic signal, the inferred tree incorrectly groups the species sharing the same composition. This compositional artefact was rapidly identified in the case of rRNA phylogeny for *Archeoglobus*[26] and *Thermus*.[27] It consists

in an elevated G + C content of the rRNA of thermophilic organisms[28] and lead to an artefactual clustering of GC rich sequences. For example *Thermus thermophilus* is artefactually grouped with the extreme thermophilic organisms *Thermotoga* and *Aquifex* instead of being together with the mesophilic *Deinococcus*. However, the addition of the nonthermophilic species, *Thermus ruber*, allows to recover the correct position of this genus,[27] thus overcoming the compositional signal. This approach circumventes the artefactual clustering of thermophilic organisms in the rRNA tree.

The second artefact is the well-known Long Branch Attraction (LBA) artefact, in which two unrelated species with an elevated evolutionary rate are erroneously clustered together.[29] Although more difficult to capture, the "rate signal" is similar to the compositional signal, since both are the result of a biased substitution process. In the simplified example of Felsenstein (1978)[29], two species A and C evolve faster than two other species B and D, with A and B being closely related. Because B and D evolve slowly, they have retained numerous ancestral nucleotide character states, whereas A and C independently accumulate many substitutions. Due to the limited set of nucleotides (only four character states), A and C will often, by chance, acquire the same nucleotide at a given position, whereas B and D will essentially remain in the ancestral state. These homoplastic positions can be interpreted as evidence for a nonexisting common ancestor of A and C. The rate signal predominates over the phylogenetic signal when rate (branch length) differences are more pronounced and the internal branch is short, i.e., fast rates (long branches) imply many (homoplastic) changes and a short internal branch reflects few informative changes. Although first described in 1978, the magnitude of LBA-related problems was for a long time largely overlooked. However, they represent major challenges to molecular phylogenetic inference.[30] It is worth to mention a particularly important case of LBA, the one where a distant outgroup attracts all fast evolving ingroup species, which therefore artefactually emerge in basal positions.[31]

The cause of these two artefacts is a severe underestimate of the true evolutionary distance, i.e., the vast majority of the multiple substitutions remain undetected by most phylogenetic methods. Therefore, the first approach to deal with multiple substitutions involves the improvement of phylogenetic methods. We will introduce basic facts of the evolutionary process and consider their impact on phylogenetic inference. The simple MP approach does not try to take into account the fact that two substitutions on long branches may be much more probable than a single substitution on a short internal branch, since it gives the same weight to all substitutions. Basically, maximum likelihood (ML) and Bayesian methods try to explicitly account for this possibility by defining a probabilistic framework.[32] More specifically, for a given topology, the probability of observing the data is computed by analysing all possible substitutional histories that explain the observed distribution of characters. To obtain an accurate estimate of this probability, one has to define a mathematical model of sequence evolution. The aim is to be able to predict when multiple substitutions have a greater probability to occur by chance on the correct topology than a single substitution on an incorrect topology i.e., to avoid that these substitutions are wrongly interpreted, thus leading to an increase of the nonphylogenetic signal.

The accuracy of ML is well above the one of MP in the case of LBA, for simulated[33] as well as real data,[34] even if a very simple and unrealistic model of evolution is used i.e., all substitutions are equally likely, positions have the same probability of accepting a substitution,[35] the only heterogeneity being that the various branches have different lengths. Several other heterogeneities of the evolutionary process have been incorporated into the models. In particular, (1) different probabilities of substitutions from a nucleotide/amino acid to another (e.g., inferred from the data in the GTR model,[36] or empirically determined for the WAG matrix),[37] (2) different probabilities of substitutions across sites, generally modelled by a Γ distribution,[38] (3) different profiles of substitution probabilities across sites,[39,40] or (4) variable nucleotide composition among lineages.[41-43] Despite these progresses that have solved several problems,[44] tree reconstruction artefacts still persist, probably because all these heterogeneities are not integrated

into a single, highly complex, model and because the real evolutionary process is much more heterogeneous. Probabilistic methods (ML and Bayesian) are widely used in current practice. Nevertheless, this accuracy depends on the underlying evolutionary model.[45] For example, heterotachy, i.e., rate variation of a given site throughout time, is known to generate artefacts when ignored,[46,47] but is not yet efficiently modelled.

A complementary approach to the use of better phylogenetic methods is the improvement of data sets in terms of species and gene sampling. The first and common strategy is simply to consider more species for the same overall biodiversity.[24] These additional species will naturally reveal multiple substitutions that are hidden when long branches are unbroken (Fig. 1B). Many simulation studies[34,48-50] and empirical studies[34,51-53] have confirmed that phylogenies based on numerous species are more accurate. In the case of animals, the inclusion of the slow-evolving nematodes reveals that they are closely related to arthropods,[54] instead of being a basal bilaterian phylum as rRNA analyses previously suggested.[55] Therefore, the ideal conditions for phylogenetic inference are a dense species sampling with the sequences under study forming a continuum, thus naturally excluding any long branches. However, in reality, beside economical reasons that limit the number of sequenced species, there are many cases in which the evolutionary process has shaped the true phylogeny in a way that long branches do exist. This is applicable for a high number of taxa like mammals, birds, tuatara, lungfishes, coelacanths, angiosperms and eukaryotes. Therefore, improving species sampling cannot be the only solution.

A second option became accessible thanks to major progresses in sequencing technology that also largely revolutionised the domain of phylogenetic inference. Large-scale genomic and EST sequencing projects provide a huge amount of raw material (in terms of orthologous genes). The use of large data sets (> 100 proteins) in molecular phylogenetics is known as phylogenomics. The phylogenomic approach has several advantages, the most important being the drastic reduction of the sampling (random) error, a major issue in analyses based on a single or few proteins. The resolving power of the phylogenetic inference is largely improved, being able to solve with high statistical support questions that were not tractable in the recent past.[56-59] Unfortunately, a 100% bootstrap value does not necessarily mean that the result is correct, because of the limitation of tree reconstruction methods. Indeed nonphylogenetic signals (e.g., compositional or rate) can accumulate under certain conditions and become predominant, especially if the phylogenetic signal is sparse. In the case of eukaryotes this is currently exacerbated by the fact that only a few completely sequenced model organisms are available for phylogenomic analyses. For instance, several studies based on up to 800 genes[60-62] significantly rejected the Ecdysozoa (moulting animals) hypothesis, in favour of the classical Coelomata (animals with a real body cavity) hypothesis. This happened mainly because the outgroup used was too distantly related, thus leading to an artefactual attraction of the fast evolving nematodes to the base of animals.[30,63,64]

An additional option to increase the phylogenetic accuracy is more drastic and perhaps less elegant, but appears to be indispensable in certain cases.[65] Since the reconstruction methods as aforementioned are not able to efficiently handle multiple substitutions, one can remove the part of the data that has a high level of multiple substitutions. The first data removal approach was applied to solve problems with compositional bias. It consists in recoding the rRNA sequences in a way that only the rarely occurring transversions (change from a puRine (R = A + G) to a pYrimidine (Y = C + T) base) and not the frequent transitions (changes among the two purine and pyrimidine bases) are used to infer the phylogeny.[26] This R + Y coding is also efficient for DNA sequences[66] and a similar approach has been developed for amino acid sequences.[67] One can also remove from the analysis the fastest evolving species, or the ones that have the most biased composition. For instance, removing the thermophilic *Thermus thermophilus* and using only the mesophilic *T. ruber* reinforces the support for the correct branching pattern.[27] Finally, the fastest evolving positions can be excluded. The simplest version is the exclusion of the third codon position. Nevertheless, it is impossible to know a priori the rate starting from which a position is too rapid to be correctly analysed. Therefore, in practice, fast evolving

sites are progressively removed until the resolution becomes limited; topological changes are strong indicators that tree reconstruction artefacts (LBA or any others) affect the inference when all positions are used and the topology based on slowly evolving positions should be preferred. Several site removal approaches have been developed and successfully applied (for review see ref. 44).

Although the accuracy of phylogenomics is currently restricted by the limited species sampling, it is ideal for applying the two other approaches discussed above to improve the phylogenetic accuracy.[44] First, when more information is available, more parameters can be accurately estimated, thus allowing the use of more complex, more realistic models of sequence evolution.[68] Second, a large part of the alignments that contain an excess of nonphylogenetic signal can be removed without any significant loss in resolution. Although, phylogenomics will demonstrate its full potential in the next few years, the conjugate progresses in phylogenetic inference methods and in sequencing technology have already seriously challenged Woese's paradigm. This is not very surprising because this paradigm was originally based on a single gene, with only a few species and simple reconstruction methods, a combination enhancing the probability of tree reconstruction artefacts.

The Root of the Tree of Life

The localisation of the root of the Tree of Life is one of the most difficult problems in phylogeny, if not the most difficult, because it concerns an extremely ancient event. Accordingly, the bacterial rooting as a part of Woese's paradigm is a good illustration of an unjustified faith. Indeed, it was inferred based on very few paralogous pairs of genes, i.e., EF-Tu/EF-G,[12] F- and V- ATPases,[13] Ile/Val tRNA synthetases (tRS),[15] hisA/hisF,[18] Trp/Tyr tRS,[69] and the SRP54/Srα pair,[17] as well as an internal duplication in the carbamoylphosphate synthetase (CPS) gene.[16] Except for the hisA/hisF pair, all these paralogs support the bacterial rooting. However, the number of alignable positions is very low (~100 per gene) in each study and the species sampling was scarce (often a single Archaea). Not surprisingly, genomics and progress in tree reconstruction methods seriously challenged these results.[70-72] First, the inference of the species phylogeny from the gene phylogeny was shown to be impossible for Ile/Val tRS, CPS, and to a lesser extent F- V-type ATPase, because of the large extent of horizontal gene transfer (HGT).[70]

More importantly, the bacterial rooting of the Tree of Life is the archetypal case of the LBA artefact (Fig. 2). In comparison to the already elevated distance observed for a given ortholog across the three domains, the distance separating two paralogous copies is really enormous. The outgroup is thus expected to attract long branches of the ingroup to the base of the tree, especially with the rudimentary tree inference methods used at that time. Further, Bacteria have by far the longest branch in all cases except hisA/hisF (Fig. 2). In consequence, Bacteria are attracted by the distantly-related paralogous copy in a systematic way. Such an interpretation provides an explanation to the strong statistical support observed for the bacterial rooting despite the use of only ~100 positions and the antiquity of the node. Interestingly, in trees where eukaryotes have the longest branch, like hisA/hisF, an eukaryotic rooting is accordingly inferred,[18] thus confirming that LBA plays a major role.

The central question is therefore to know whether (1) the bacterial rooting is an artefact due to an accelerated evolutionary rate of Bacteria or (2) if the bacterial rooting is actually correct, with the long branch simply reflecting a long period of independent evolution and a possible rate acceleration. To adequately address this question, we chose to focus on SRP54/SRα, because these two paralogs are the most similar pair among the anciently duplicated genes.[72] In addition, we used the SF method that analyses the most slowly evolving positions and is known to eschew the LBA artefact when even complex probabilistic methods fail.[30] In these conditions, an alternative rooting of the Tree of Life is obtained, with the prokaryotes (Bacteria and Archaea) forming a monophyletic group to the exclusion of eukaryotes.[72] Unfortunately, because of the reduction of the number of positions imposed by the SF method, this eukaryotic rooting is not significantly supported.

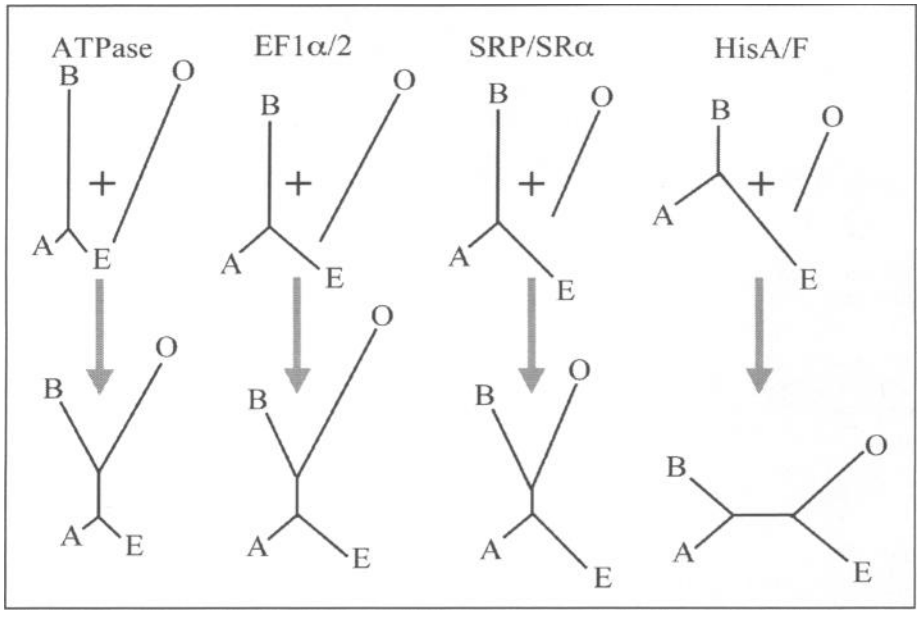


Figure 2. Long Branch Attraction and the root of the Tree of Life. The three domains of Life (Archaea, Bacteria and Eukaryota) are each represented by their initial letter. O means outgroup and corresponds to the second paralog of the anciently duplicated gene. The branch length is proportional to the average rate within each domain. Note that the scale is not the same for the different genes, since our concern is on the relative branch lengths of A, B, E, and O. The gene name corresponds to the unrooted tree. For the three leftmost pairs, the branch of Bacteria is by far the longest with respect to Archaea and Eukaryota and is therefore very likely attracted by the even longer branch of the outgroup, resulting into the classical bacterial rooting.[1] Importantly, for the hisA/F gene pair, the eukaryotic branch is the longest, and accordingly an eukaryotic rooting is recovered.

Another approach, though quite radical, to solve this central question is to remove the source of LBA artefacts by discarding the outgroup. In fact a nonreversible model allows to find the root of a tree.[73] Although the elimination of a distant outgroup is highly advantageous, its benefits are somewhat weakened by the strong assumptions that have to be made about the evolutionary process. Such an approach applied to the secondary structural elements of both large and small subunit rRNAs recovered an eukaryotic rooting.[74] However, the lack of reference studies makes the estimation of the reliability of the method difficult.

In conclusion, the bacterial rooting recovered in the early nineties very likely resulted from a nonphylogenetic signal due to poor species sampling and use of simplistic methods, as demonstrated by refined studies.[70-72] Moreover, the lack of evidence for an archaeal rooting suggests that such a hypothesis can be ruled out. While the phylogenetic signal seems to design the eukaryotic rooting as our best working hypothesis, the weakness of the statistical support prompts to additional studies.

"Primitive" Eukaryotes: A Second Major Paradigm Shift

All methodological improvements discussed above pointed towards an artefactual nature of the base of the rRNA tree,[75-80] strongly suggesting that Archezoa and other so-called "early" emerging taxa belong to the "crown". Indeed, even based on rRNA, better methods (e.g., a gamma distribution to model rate across site variation) are able to locate some amitochondriate

eukaryotes (e.g., trichomonads) at the top of the tree, albeit without significant support.[81] The amitochondriate Microsporidia is the most famous example being convincingly demonstrated to be closely related to fungi using protein sequences.[34,82]

An important implication of this paradigm shift is that the last common ancestor of extant eukaryotes must have been much more complex than previously thought. The prominent example is the presence of a mitochondrium in this ancestor. Indeed, even if convoluted scenarios involving HGTs can not be formally ruled out, the discovery of typical mitochondrial genes in all archezoan groups examined, *Entamoeba*,[83] trichomonads,[84,85] microsporidia,[86] and diplomonads,[87] is more parsimoniously interpreted as evidence for the ancestral presence of mitochondria in these organisms. Accordingly, the proteins encoded by these genes are localised in double-membrane bound organelles known as hydrogenosomes[88] or in the so-called mitosomes.[89-91] These organelles are therefore most likely remnants of mitochondria having lost their respiratory function. Generally devoid of a genome (except the ciliate *Nyctotherus*),[92] hydrogenosomes possess an anaerobic energy metabolism based on hydrogen production and are polyphyletically distributed in diverse groups (e.g., trichomonads, ciliates, fungi). Mitosomes are not known to be involved in energy metabolism and the reason for their evolutionary conservation probably lies in other functions fulfilled by mitochondria, in particular iron-sulphur protein maturation.[91]

In summary, two independent lines of evidence invalidate the Archezoa hypothesis, thus seriously challenging Woese's paradigm. Furthermore, the relocation of all former "intermediate groups" within the crown creates a huge gap between the root and extant eukaryotes and greatly reduces the evolutionary distances among eukaryotes. This suggests that the diversification of eukaryotes is more recent than previously assumed, and that their evolutionary diversity is much less important than the one of prokaryotes. In fact, because mitochondria originated from within alpha-proteobacteria and therefore, in terms of elapsed time of evolution, the diversity of extant eukaryotes is inferior to the alpha-proteobacterial diversity.

Current Status of the Eukaryotic Phylogeny

Despite the amazing increase of available data, the resolution of the eukaryotic phylogeny turns out to remain difficult, as predicted by the Big Bang hypothesis.[77] Reasons are a relatively rapid diversification and to a lesser extent the obligation to resort to the distant Archaea as an outgroup. A consensus (Fig. 3), in which six super-ensembles are recognised,[93-96] is nevertheless emerging. It is mainly based on a few multigene analyses, associated with the use of a handful of rare evolutionary changes (i.e., insertions/deletions -indels- and gene duplication/fusion/replacement). The attention accorded to rare evolutionary events is based on the (often unjustified) assumption that such characters provide a highly reliable information, that can be captured even when a simplistic inference method (i.e., maximum parsimony) is implicitly used.[31,97]

Opisthokonta, grouping animals, fungi and several protist clades (e.g., choanoflagellates, nucleariids, ichthyosporeans, corallochytreans) is the most reliable super-ensemble. It is supported by numerous approaches, including indels[98] and multigene inference.[99-102] While the uniflagellated reproductive stage constitutes a single morphological character shared by most animals and some fungi (i.e., chytrids), there seems to be no morphological features uniting all opisthokont lineages.[101]

Plantae contain the three lineages of primary photosynthetic organisms, i.e., glaucophytes, green plants and red algae. Although a solid demonstration was lacking, the monophyly of Plantae has been generally accepted,[103] most likely because of the broad consensus for a single origin of primary plastids. Recently, phylogenomic analyses provided sound support for the monophyly of both primary plastids and primary photosynthetic eukaryotes.[104] This strongly argues for a single primary endosymbiosis with a cyanobacterium (but see ref. 105 for a second much more recent independent primary endosymbiosis in the cercozoan *Paulinella chromatophora*).

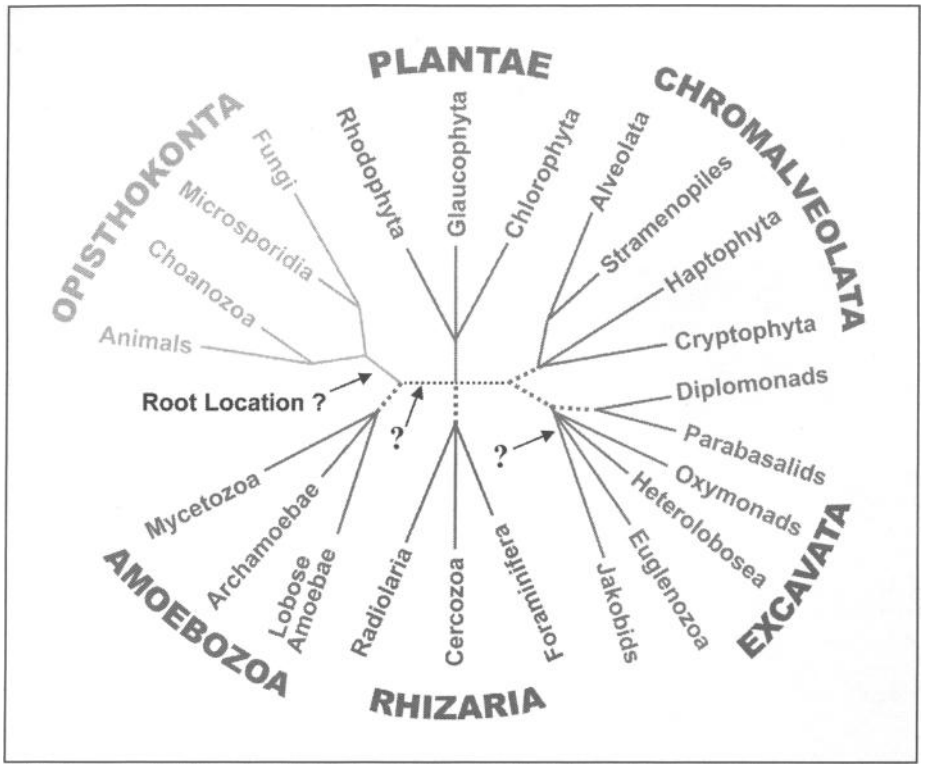


Figure 3. The current consensus of eukaryotic phylogeny. The six proposed super-ensembles of eukaryotes including their major constituting groups are shown in different colours. Not solidly established connections are indicated by dotted lines. The arrows and the question marks indicate possible locations of the root of the eukaryotic tree.

Amoebozoa unites most of the amoeba, including slime molds, lobose amoeba and anaerobic Archamoebae, but neither filose amoeba nor Actinopoda. Evidence for the monophyly of Amoebozoa is rather limited. Indeed, the fusion of cox1 and cox2 genes in mitochondrial genomes of *Acanthamoeba* and *Dictyostelium*[106] supports this grouping, but needs to be examined in additional species. The monophyly of Conosa (i.e., slime molds + pelobiont + entamoebids) is found by a phylogenomic approach[107] and more recently, we recovered the monophyly of Amoebozoa with an extended biodiversity (9 species) using ~100 genes (unpublished results).

Chromalveolata contains four diverse groups, alveolates (including dinoflagellates, Apicomplexa and ciliates), cryptophytes, haptophytes and stramenopiles (including oomycetes, brown algae, golden algae, diatoms, among many others). Actually, the Chromalaveolata hypothesis[108] implies a single secondary endosymbiosis involving a red alga at the origin of the group. Two specific gene duplications, concerning glyceraldehyde-3-phosphate dehydrogenase[109] and fructose bisphosphate aldolase,[110] support their monophyly. However, the gene for the plastid-specific phosphoribulokinase[111] found in all chromalveolates surprisingly seems to be of green algal origin. However, the absence of any specific association with red algae for all three aforementioned markers is puzzling. Furthermore, a phylogeny based on six concatenated genes does not recover the monophyly of chromalveolates,[112] whereas the analysis of 141 genes strongly places alveolates and stramenopiles as sister-groups[100] but haptophytes and cryptophytes

were missing. Confirmation of the clade chromalveolates is urgently needed as it has important implications on the spread of photosynthesis in eukaryotes.

If chromalveolates are a genuine monophyletic group, this would imply the occurrence of several losses of photosynthetic functions (e.g., in Apicomplexa), but also of all plastid structures (e.g., in ciliates and in oomycetes). Nevertheless, the monophyly of chromalveolates can be partially tested by using plastid-encoded genes. Such phylogenies (usually excluding very fast-evolving apicomplexa and dinoflagellates) lead to contradictory and weakly supported results: either monophyly in a species-rich and gene-poor study[113] and a rather species-rich and gene-rich study,[100] or polyphyly in a species-poor and gene-rich study.[114] The question remains therefore open.

Evidence for the monophyly of the last two super-ensembles, Rhizaria and Excavata, is much more tenuous. It consists mainly on weakly supported phylogenies based on a single or a few genes that unite some, but usually never all, members of these groups,[115-118] as well as indels.[119] Whereas excavates have been proposed on morphological grounds, there seems to be no morphological feature that unites all of them.[120] Indeed, the particular type of ventral feeding groove giving its name to the group is only shared by the "core" excavates i.e., diplomonads, retortamonads, jakobids, heteroloboseans, *Carpediemonas*, *Trimastix*, and *Malawimonas*. In addition, Euglenozoa, oxymonads and parabasalids are associated with excavates by a combination of morphological and molecular data.[121] Finally, no morphological character is known to be specific to all Rhizaria.[122]

Even if the monophyly of these six assemblages is confirmed, the important question of the location of the root remains, because Archaea are too distantly related for a reliable rooting by current methods.[34] Given the small distance of proteobacteria to eukaryotes compared to Archaea, so far poorly studied eukaryotic genes of mitochondrial origin are a promising avenue of research that may ultimately allow to confidently root extant eukaryotic diversity.

Instead, attention was essentially focused on a few rare evolutionary changes. For instance, an enolase indel shared with prokaryotes supports a basal position of *Trichomonas*.[123] Later shown to have been independently lost or gained among prokaryotes, this particular indel is now considered to be unreliable.[124,125] One highly debated feature is the gene fusion of dihydrofolate reductase (DHFR) and thymidylate synthase (TS) genes.[76,126] This fusion is found in all groups that are assumed to be ancestrally with two cilia (bikonts), whereas the remaining groups (unikonts) have separate genes (Opisthokonta and Bacteria). However, several species (in particular some amoebozoan, diplomonads and trichomonads) lack the orthologous genes, thus preventing their placement based on the DHFR + TS fusion. Two other characters, a gene fusion of the first three enzymes of the pyrimidine synthesis pathway[127] and an internal duplication in the phosphofructokinase gene, are restricted to opisthokonts and Amoebozoa.[128] Furthermore, a recent study based on the taxonomic distribution of myosin domain combinations claims that "five innovations (new domain combinations) strongly support unikont monophyly and the primary bikont/unikont bifurcations".[129]

While these slowly evolving characters point towards a root between Amoebozoa + Opisthokonta and Cercozoa + Chromalveolata + Plantae + Excavata; the replacement of the proteobacterial RNA polymerase by a T3/T7-like polymerase in mitochondria of all eukaryotes except jakobids (e.g., *Reclinomonas*) strongly supports an incompatible rooting within excavates, with jakobids emerging first.[130] This demonstrates that these rare characters can also be affected by convergence, HGT, or reversion. The location of the root of the eukaryotic tree remains an open question and the last word will have to wait for congruence of phylogenetic inference and rare genomic changes.

Molecular Dating

Time estimates based on molecular data vary widely. For example, the origin of Plantae is either inferred around 1,600-1,500 mya[113,131] or around 1,000 mya.[132] Beside problems due to phylogenetic inference explained above, dating methods have to deal with uncertainty of the

fossil record and with rate variation across lineages.[133,134] Interestingly, important methodological progresses have been recently achieved.[135-137] In particular, calibration dates can now be provided in terms of intervals, which allows to account for paleontological uncertainty, whereas the use of a fixed time point previously created the illusion of precision in molecular dating.[133] In addition, instead of excluding genes that do not have a clockwise behaviour (a difficult task given the weaknesses of existing tests),[138,139] rate variation is modelled by so-called relaxed molecular clock techniques.[136,137,140]

To obtain the most accurate time estimates, it is primordial to use: (1) a large number of genes to reduce the stochastic error, (2) a large number of species to reduce the phylogenetic error, (3) multiple calibration intervals to reduce the paleontological error, and (4) relaxed clock methods to reduce the dating error. Very few studies have fulfilled all these four criteria,[132,141] with the only one focused on eukaryotic evolution being the work of Douzery et al. (2004) based on 36 species and 129 genes. Virtually identical results are obtained with 131 species and 107 genes (unpublished results). In that study, the mean age estimate for the separation of the most basal node (i.e., between opisthokonts and bikonts; Amoebozoa being used as an outgroup) is 1,085 million years ago (Mya) with a 95% confidence interval of (950,1,259). The primary plastid endosymbiosis occurred between 825 and 1,162 Mya (mean: 1,010 Mya) while the secondary endosymbiosis with a red alga likely happened at the origin of chromalveolates between 767 and 1,072 Mya (mean: 872 Mya). These estimates suggest that the events leading to the major groups of extant eukaryotes occurred in a relatively short time span (less than 200 My) around 1,000 Mya, in agreement with the Big Bang hypothesis.[77]

According to the fossil record (see chapter of Emmanuelle Javaux) eukaryotes were already well diversified about 1,500 Mya.[142] However, none of these early eukaryotic fossils can be clearly associated with any of the extant groups.[143] Evidence for an older origin of eukaryotes includes the existence of eukaryotic bio-markers (macromolecules) as far back as 2,700 Mya.[144] Since this conclusion heavily depends on the assumption that the considered biological macromolecules are specific to eukaryotes, it needs to be taken with caution. At any case, there is a gap of at least ~400 My (but potentially up to ~1,600 My) between the first occurrence of eukaryotes (or eukaryote-specific biomarkers) and the last common ancestor of extant eukaryotes (LCAEE). Corresponding to a prolonged early (and essentially unknown) phase of eukaryotic evolution, this very long yet unbroken basal branch explains the difficulty of rooting the eukaryotic tree. Is the existence of a branch devoid of any detectable speciation event simply due to chance (contingency) or is it associated with specific factors, either external (e.g., major environmental changes), and/or internal (e.g., major biological innovations)? Being the simplest, the first interpretation constitutes the null hypothesis. Indeed, the well-known phenomenon of coalescence implies that the common ancestor of the extant diversity of genes (or lineages) can never be traced back to their origin, as examplified by the recent common ancestor of current humans.[145] This result is due to gene/lineage losses that can be obtained under a simple model of a random birth and death process.[146] Nonetheless, the other options are much more tempting and are in consequence often favoured. For example, Philippe and Adoutte[77] proposed that the diversification of eukaryotes was intimately linked to the mitochondrial endosymbiosis and that the great increase of the atmospheric oxygen level that started about 1,000 Mya[147] was a major trigger of the evolution of aerobic eukaryotes. Schematically, the mitochondrial endosymbiosis likely occurred well before the diversification of extant eukaryotes. This view is supported by the observation that a large number of genes of mitochondrial origin is found in the nucleus of all extant eukaryotes (especially in comparison to plastids). While their original function had not necessarily to be aerobic respiration (e.g., anaerobic photosynthesis, oxygen detoxification or hydrogen metabolism), mitochondria would have provided a major selective advantage to their host in the context of a steeply increasing oxygen level. In addition to trigger the diversification of mitochondriate eukaryotes, the mitochondrial endosymbiosis and the rise of oxygen level would have simultaneously favoured the extinction of the genuinely amitochondriate eukaryotes, thus explaining the current shape of

the eukaryotic tree. Nevertheless, the simple coalescent hypothesis has never been rejected and formally remains the null hypothesis for explaining the long unbroken basal branch of eukaryotes.

From Simple Organism to Simple Processes

The basic and widespread philosophy underlying Woese's paradigm is that simple organisms (e.g., prokaryotes and Archezoa) represent genuinely primitive intermediates in the progressive construction of complex eukaryotic cells. However, this conclusion is based on a phylogenetic framework now known to be erroneous. The rejection of the naïve assumption "simple organism = primitive organism" is well illustrated by microsporidia, which are simple but highly derived, rather than primitive (see chapter by Alexander Mironov). Moreover, taking the new eukaryotic phylogeny for granted (Fig. 3), any feature present in some opisthokonts (e.g., animals) and in some bikonts (e.g., plants) is necessarily ancestral, i.e., inherited from the LCAEE. This implies that the LCAEE was amazingly more complex than previously thought.[77,148] Among others, the LCAEE already possessed mitochondria, an efficient cytoskeleton associated with several intracellular transport systems, an endo-membrane system interconnected by a complicated vesicular transport machinery, including an endoplamic reticulum, a Golgi apparatus, a standard nucleus, efficient secretory and uptake pathways, recycling of food-vacuoles, peroxisomes, spliceosomal introns, and flagella-dependent motility.[77,122,148,149] Therefore, simple extant eukaryotes have evolved from a complex LCAEE mainly by loss and secondary simplification.

As early as in 1943, secondary simplification was already recognised as a fundamental mode of evolution by André Lwoff in his insufficiently known book "L'évolution physiologique. Etude des pertes de fonctions chez les microorganismes" (The physiological evolution: Study of functional losses in microorganisms). Lwoff[150] not only demonstrates that simplification is a major evolutionary process, but also explains why scientists are so reluctant to accept the importance of secondary simplification. The current notion of evolutionary change is strongly associated with human concepts of progress and of evolution towards higher complexity.[151] This is linked to an unfortunate analogy between the upward trend characteristic of cultural evolution and natural evolution. Hence, conflicting with the powerful ideology of never-ending progress, evolution by secondary simplification has always been largely discarded. Nevertheless, recent molecular studies of eukaryotes have confirmed Lwoff's ideas about secondary simplification, as best illustrated by the numerous losses of oxygenic respiration in mitochondria or of plastid-based photosynthesis.

The rejection of simplification, combined with the strong underlying influence of the *Scala Naturae* of Aristotle, lead most researchers to confuse simple states with simple processes. From an evolutionary perspective, the division into simple steps that may eventually lead toward a complicated characteristic/structure (e.g., eye development) is primordial to its realisation. However, extant species, irrespective of whether they are simple or complex, do not allow any a priori deduction of their complexity relative to their ancestors; such deductions can only be made through a phylogenetic tree. Indeed, because evolution fundamentally proceeds through tinkering,[152] it is likely that a new function will be first achieved by the modification of an existing system, thus resulting in a rather complex and less efficient solution, which may eventually be secondarily simplified and become more effective.[153] We therefore strongly urge to abandon the inherent bias towards complexification and to reevaluate the origin and evolution of eukaryotes open minded.

Acknowledgements

We wish to thank the CRC and Génome Québec for financial support given to HP. Furthermore, we are grateful to Denis Baurain and Naiara Rodríguez-Ezpeleta for helpful comments made on an earlier version of the manuscript. We greatly appreciated the encouragement and patience of Gáspár Jékely.

References

1. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 1990; 87(12):4576-4579.
2. Sogin ML, Elwood HJ, Gunderson JH. Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proc Natl Acad Sci USA 1986; 83(5):1383-1387.
3. Knoll AH. The early evolution of eukaryotes: A geological perspective. Science 1992; 256(5057):622-627.
4. Sogin ML. Early evolution and the origin of eukaryotes. Curr Opin Genet Dev 1991; 1(4):457-463.
5. Cavalier-Smith T. Molecular phylogeny. Archaebacteria and Archezoa. Nature 1989; 339(6220):100-01.
6. Bekker A, Holland HD, Wang PL et al. Dating the rise of atmospheric oxygen. Nature 2004; 427(6970):117-120.
7. Mereschkowski C. Ueber Natur und Ursprung der Cromatophoren im Pflanzenreiche. Biologisches Centralblatt 1905; 25:593-604.
8. Margulis L. Origin of Eukaryotic cells. New Haven: Yale University Press, 1970.
9. Tomioka N, Sugiura M. The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue-green alga, Anacystis nidulans. Mol Gen Genet 1983; 191(1):46-50.
10. Yang D, Oyaizu Y, Oyaizu H et al. Mitochondrial origins. Proc Natl Acad Sci USA 1985; 82(13):4443-4447.
11. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Natl Acad Sci USA 1977; 74(11):5088-5090.
12. Iwabe N, Kuma K, Hasegawa M et al. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 1989; 86(23):9355-9359.
13. Gogarten JP, Kibak H, Dittrich P et al. Evolution of the vacuolar H+-ATPase: Implications for the origin of eukaryotes. Proc Natl Acad Sci USA 1989; 86(17):6661-6665.
14. Baldauf SL, Palmer JD, Doolittle WF. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc Natl Acad Sci USA 1996; 93(15):7749-7754.
15. Brown JR, Doolittle WF. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA 1995; 92(7):2441-2445.
16. Lawson FS, Charlebois RL, Dillon JA. Phylogenetic analysis of carbamoylphosphate synthetase genes: Complex evolutionary history includes an internal duplication within a gene which can root the tree of life. Mol Biol Evol 1996; 13(7):970-977.
17. Gribaldo S, Cammarano P. The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. J Mol Evol 1998; 47(5):508-516.
18. Charlebois RL, Sensen CW, Doolittle WF et al. Evolutionary analysis of the hisCGABdFDEHI gene cluster from the archaeon Sulfolobus solfataricus P2. J Bacteriol 1997; 179(13):4429-4432.
19. Gupta RS, Singh B. Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. Curr Biol 1994; 4(12):1104-1114.
20. Zillig W. Eukaryotic traits in Archaebacteria. Could the eukaryotic cytoplasm have arisen from archaebacterial origin? Ann NY Acad Sci 1987; 503:78-82.
21. Martin W, Müller M. The hydrogen hypothesis for the first eukaryote. Nature 1998; 392(6671):37-41.
22. Moreira D, López-García P. Symbiosis between methanogenic archaea and delta-Proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. J Mol Evol 1998; 47(5):517-530.
23. Olsen G. Earliest phylogenetic branching: Comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harb symp Quant Biol 1987; LII:825-837.
24. Hendy MD, Penny D. A framework for the quantitative study of evolutionary trees. Syst Zool 1989; 38:297-309.
25. Lecointre G, Philippe H, Le HLV et al. Species sampling has a major impact on phylogenetic inference. Mol Phylogenet Evol 1993; 2(3):205-224.
26. Woese CR, Achenbach L, Rouviere P et al. Archaeal phylogeny: Reexamination of the phylogenetic position of Archaeoglobus fulgidus in light of certain composition-induced artifacts. Syst Appl Microbiol 1991; 14(4):364-371.
27. Embley TM, Thomas RH, Williams RAD. Reduced thermophilic bias in the 16S rDNA sequence from Thermus ruber provides further support for a relationship between Thermus and Deinococcus. Syst Appl Microbiol 1992; 16:25-29.
28. Galtier N, Lobry JR. Relationships between genomic G + C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J Mol Evol 1997; 44(6):632-636.
29. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 1978; 27:401-410.

30. Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 2005; 6(5):361-375.
31. Philippe H, Laurent J. How good are deep phylogenetic trees? Curr Opin Genet Dev 1998; 8(6):616-623.
32. Felsenstein J. Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 1981; 17(6):368-376.
33. Huelsenbeck JP. Is the Felsenstein zone a fly trap? Syst Biol 1997; 46(1):69-74.
34. Brinkmann H, Giezen M, Zhou Y et al. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst Biol 2005; 54(5):743-757.
35. Jukes TH, Cantor CR. Evolution of protein molecules. In: Munro HN, ed. Mammalian Protein Metabolism. New York: Academic Press, 1969:21-132.
36. Lanave C, Preparata G, Saccone C et al. A new method for calculating evolutionary substitution rates. J Mol Evol 1984; 20(1):86-93.
37. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 2001; 18(5):691-699.
38. Yang Z. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol 1993; 10(6):1396-1401.
39. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 2004; 21(6):1095-1109.
40. Pagel M, Meade A. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Syst Biol 2004; 53(4):571-581.
41. Galtier N, Gouy M. Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol Biol Evol 1998; 15(7):871-879.
42. Yang Z, Roberts D. On the use of nucleic acid sequences to infer early branchings in the tree of life. Mol Biol Evol 1995; 12(3):451-458.
43. Foster PG. Modeling compositional heterogeneity. Syst Biol 2004; 53(3):485-495.
44. Philippe H, Delsuc F, Brinkmann H et al. Phylogenomics. Annu Rev Ecol Evol Syst 2005; 36:541-562.
45. Felsenstein J. Inferring phylogenies. Sunderland, MA, USA: Sinauer Associates Inc., 2004.
46. Lockhart PJ, Larkum AW, Steel M et al. Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis. Proc Natl Acad Sci USA 1996; 93(5):1930-1934.
47. Kolaczkowski B, Thornton JW. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 2004; 431(7011):980-984.
48. Hillis DM. Inferring complex phylogenies. Nature 1996; 383(6596):130-131.
49. Kim J. General inconsistency conditions for maximum parsimony: Effects of branch lengths and increasing numbers of taxa. Syst Biol 1996; 45(3):363-374.
50. Graybeal A. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst Biol 1998; 47(1):9-17.
51. Philippe H, Douzery E. The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. Journal of Mammalian Evolution 1994; 2:133-152.
52. Mitchell A, Mitter C, Regier JC. More taxa or more characters revisited: Combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). Syst Biol 2000; 49(2):202-224.
53. Lin YH, McLenachan PA, Gore AR et al. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol Biol Evol 2002; 19(12):2060-2070.
54. Aguinaldo AM, Turbeville JM, Linford LS et al. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 1997; 387(6632):489-493.
55. Field KG, Olsen GJ, Lane DJ et al. Molecular phylogeny of the animal kingdom. Science 1988; 239(4841 Pt 1):748-753.
56. Madsen O, Scally M, Douady CJ et al. Parallel adaptive radiations in two major clades of placental mammals. Nature 2001; 409(6820):610-614.
57. Murphy WJ, Eizirik E, Johnson WE et al. Molecular phylogenetics and the origins of placental mammals. Nature 2001; 409(6820):614-618.
58. Qiu YL, Lee J, Bernasconi-Quadroni F et al. The earliest angiosperms: Evidence from mitochondrial, plastid and nuclear genomes. Nature 1999; 402(6760):404-407.
59. Soltis PS, Soltis DE, Chase MW. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 1999; 402(6760):402-404.
60. Wolf YI, Rogozin IB, Koonin EV. Coelomata and not ecdysozoa: Evidence from genome-wide phylogenetic analysis. Genome Res 2004; 14(1):29-36.

61. Blair JE, Ikeo K, Gojobori T et al. The evolutionary position of nematodes. BMC Evol Biol 2002; 2(1):7.
62. Philip GK, Creevey CJ, McInerney JO. The Opisthokonta and the Ecdysozoa may not be clades: Stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. Mol Biol Evol 2005; 22(5):1175-1184.
63. Philippe H, Lartillot N, Brinkmann H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol Biol Evol 2005; 22(5):1246-1253.
64. Dopazo H, Dopazo J. Genome-scale evidence of the nematode-arthropod clade. Genome Biol 2005; 6(5):R41.
65. Jeffroy O, Brinkmann H, Delsuc F et al. Phylogenomics: The beginning of incongruence? Trends in Genetics, (in press).
66. Delsuc F, Phillips MJ, Penny D. Comment on "Hexapod origins: Monophyletic or paraphyletic?" Science 2003; 301(5639):1482, (author reply 1482).
67. Hrdy I, Hirt RP, Dolezal P et al. Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature 2004; 432(7017):618-622.
68. Steel M. Should phylogenetic models be trying to 'fit an elephant'? Trends Genet 2005; 21(6):307-309.
69. Brown JR, Robb FT, Weiss R et al. Evidence for the early divergence of tryptophanyl- and tyrosyl-tRNA synthetases. J Mol Evol 1997; 45(1):9-16.
70. Philippe H, Forterre P. The rooting of the universal tree of life is not reliable. J Mol Evol 1999; 49(4):509-523.
71. Lopez P, Forterre P, Philippe H. The root of the tree of life in the light of the covarion model. J Mol Evol 1999; 49:496-508.
72. Brinkmann H, Philippe H. Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Mol Biol Evol 1999; 16(6):817-825.
73. Yap VB, Speed T. Rooting a phylogenetic tree with nonreversible substitution models. BMC Evol Biol 2005; 5(1):2.
74. Caetano-Anolles G. Evolved RNA secondary structure and the rooting of the universal tree of life. J Mol Evol 2002; 54(3):333-345.
75. Philippe H, Germot A, Moreira D. The new phylogeny of eukaryotes. Curr Opin Genet Dev 2000; 10(6):596-601.
76. Philippe H, Lopez P, Brinkmann H et al. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc R Soc Lond BS 2000; 267(1449):1213-1221.
77. Philippe H, Adoutte A. The molecular phylogeny of Eukaryota: Solid facts and uncertainties. In: Coombs G, Vickerman K, Sleigh M, Warren A, eds. Evolutionary Relationships Among Protozoa. Dordrecht: Kluwer, 1998:25-56.
78. Baldauf SL, Roger AJ, Wenk-Siefert I et al. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 2000; 290(5493):972-977.
79. Roger AJ. Reconstructing early events in eukaryotic evolution. Am Nat 1999; 154:S146-S163.
80. Embley TM, Hirt RP. Early branching eukaryotes? Curr Opin Genet Dev 1998; 8(6):624-629.
81. Silberman JD, Clark CG, Diamond LS et al. Phylogeny of the genera Entamoeba and Endolimax as deduced from small- subunit ribosomal RNA sequences. Mol Biol Evol 1999; 16(12):1740-1751.
82. Keeling PJ, Fast NM. Microsporidia: Biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol 2002; 56:93-116.
83. Clark CG, Roger AJ. Direct evidence for secondary loss of mitochondria in Entamoeba histolytica. Proc Natl Acad Sci USA 1995; 92(14):6518-6521.
84. Roger AJ, Clark CG, Doolittle WF. A possible mitochondrial gene in the early-branching amitochondriate protist Trichomonas vaginalis. Proc Natl Acad Sci USA 1996; 93(25):14618-14622.
85. Germot A, Philippe H, Le Guyader H. Presence of a mitochondrial-type 70-kDa heat shock protein in Trichomonas vaginalis suggests a very early mitochondrial endosymbiosis in eukaryotes. Proc Natl Acad Sci USA 1996; 93(25):14614-14617.
86. Germot A, Philippe H, Le Guyader H. Evidence for loss of mitochondria in Microsporidia from a mitochondrial- type HSP70 in Nosema locustae. Mol Biochem Parasitol 1997; 87(2):159-168.
87. Roger AJ, Svard SG, Tovar J et al. A mitochondrial-like chaperonin 60 gene in Giardia lamblia: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Natl Acad Sci USA 1998; 95(1):229-234.
88. Bui ET, Bradley PJ, Johnson PJ. A common evolutionary origin for mitochondria and hydrogenosomes. Proc Natl Acad Sci USA 1996; 93(18):9651-9656.
89. Tovar J, Fischer A, Clark CG. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica. Mol Microbiol 1999; 32(5):1013-1021.

90. Williams BA, Hirt RP, Lucocq JM et al. A mitochondrial remnant in the microsporidian Trachipleistophora hominis. Nature 2002; 418(6900):865-869.
91. Tovar J, Leon-Avila G, Sanchez LB et al. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature 2003; 426(6963):172-176.
92. Boxma B, de Graaf RM, van der Staay GW et al. An anaerobic mitochondrion that produces hydrogen. Nature 2005; 434(7029):74-79.
93. Keeling PJ, Burger G, Durnford DG et al. The tree of eukaryotes. Trends in Ecology and Evolution 2005; 20(12):670-676.
94. Adl SM, Simpson AG, Farmer MA et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 2005; 52(5):399-451.
95. Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. Curr Biol 2004; 14(17):R693-696.
96. Patterson DJ. The diversity of eukaryotes. Am Nat 1999; 154:S96-S124.
97. Rokas A, Holland PW. Rare genomic changes as a tool for phylogenetics. Trends in Ecology and Evolution 2000; 15(11):454-459.
98. Baldauf SL, Palmer JD. Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins. Proc Natl Acad Sci USA 1993; 90(24):11558-11562.
99. Baldauf SL. A search for the origins of animals and fungi: Comparing and combining molecular data. Am Nat 1999; 154:S178-S188.
100. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC et al. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. Current Biology 2005; 15(14):1325-1330.
101. Steenkamp ET, Wright J, Baldauf SL. The protistan origins of animals and fungi. Mol Biol Evol 2005.
102. Philippe H, Snell EA, Bapteste E et al. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Mol Biol Evol 2004; 21(9):1740-1752.
103. Palmer JD. The symbiotic birth and spread of plastids: Hows many times and whodunit? J Phycol 2003; 39:4-11.
104. Rodríguez-Ezpeleta N, Brinkmann H, Burey S et al. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae and glaucophytes. Current Biology 2005.
105. Marin B, EC MN, Melkonian M. A plastid in the making: Evidence for a second primary endosymbiosis. Protist 2005; 156(4):425-432.
106. Lonergan KM, Gray MW. Expression of a continuous open reading frame encoding subunits 1 and 2 of cytochrome c oxidase in the mitochondrial DNA of Acanthamoeba castellanii. J Mol Biol 1996; 257(5):1019-1030.
107. Bapteste E, Brinkmann H, Lee JA et al. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. Proc Natl Acad Sci USA 2002; 99(3):1414-1419.
108. Cavalier-Smith T. Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote familly tree. J Eukaryot Microbiol 1999; 46(4):347-366.
109. Harper JT, Keeling PJ. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. Mol Biol Evol 2003; 20(10):1730-1735.
110. Patron NJ, Rogers MB, Keeling PJ. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. Eukaryot Cell 2004; 3(5):1169-1175.
111. Petersen J, Teich R, Brinkmann H et al. A "Green" phosphoribulokinase in complex algae with red plastids: Evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts and dinoflagellates. Journal of molecular evolution 2005; 60, (in press).
112. Harper JT, Waanders E, Keeling PJ. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int J Syst Evol Microbiol 2005; 55(Pt 1):487-496.
113. Yoon HS, Hackett JD, Ciniglia C et al. A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 2004; 21(5):809-818.
114. Martin W, Rujan T, Richly E et al. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA 2002; 99(19):12246-12251.
115. Simpson AG, Roger AJ, Silberman JD et al. Evolutionary history of "early-diverging" eukaryotes: The excavate taxon Carpediemonas is a close relative of Giardia. Mol Biol Evol 2002; 19(10):1782-1791.
116. Cavalier-Smith T. The excavate protozoan phyla Metamonada Grasse emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): Their evolutionary affinities and new higher taxa. Int J Syst Evol Microbiol 2003; 53(Pt 6):1741-1758.

117. Nikolaev SI, Berney C, Fahrni JF et al. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci USA 25 2004; 101(21):8066-8071.
118. Arisue N, Hasegawa M, Hashimoto T. Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. Mol Biol Evol 2005; 22(3):409-420.
119. Bass D, Moreira D, Lopez-Garcia P et al. Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. Protist 2005; 156(2):149-161.
120. Simpson AG. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol 2003; 53(Pt 6):1759-1777.
121. Hampl V, Horner DS, Dyal P et al. Inference of the phylogenetic position of oxymonads based on nine genes: Support for metamonada and excavata. Mol Biol Evol 2005; 22(12):2508-2518.
122. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 2002; 52(Pt 2):297-354.
123. Keeling PJ, Palmer JD. Parabasalian flagellates are ancient eukaryotes. Nature 2000; 405(6787):635-637.
124. Bapteste E, Philippe H. The potential value of indels as phylogenetic markers: Position of trichomonads as a case study. Mol Biol Evol 2002; 19(6):972-977.
125. Keeling PJ. Polymorphic insertions and deletions in parabasalian enolase genes. J Mol Evol 2004; 58(5):550-556.
126. Stechmann A, Cavalier-Smith T. Rooting the eukaryote tree by using a derived gene fusion. Science 2002; 297(5578):89-91.
127. Nara T, Hshimoto T, Aoki T. Evolutionary implications of the mosaic pyrimidine-biosynthetic pathway in eukaryotes. Gene 2000; 257(2):209-222.
128. Stechmann A, Cavalier-Smith T. The root of the eukaryote tree pinpointed. Curr Biol 2003; 13(17):R665-666.
129. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436(7054):1113-1118.
130. Cermakian N, Ikeda TM, Miramontes P et al. On the evolution of the single-subunit RNA polymerases. J Mol Evol 1997; 45(6):671-681.
131. Hedges SB, Blair JE, Venturi ML et al. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol 2004; 4:2.
132. Douzery EJ, Snell EA, Bapteste E et al. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? Proc Natl Acad Sci USA 2004, (10 19).
133. Graur D, Martin W. Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. Trends Genet 2004; 20(2):80-86.
134. Bromham L, Penny D. The modern molecular clock. Nat Rev Genet 2003; 4(3):216-224.
135. Sanderson MJ. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol 1997; 14(12):1218-1231.
136. Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 1998; 15(12):1647-1657.
137. Huelsenbeck JP, Larget B, Swofford D. A compound poisson process for relaxing the molecular clock. Genetics 2000; 154(4):1879-1892.
138. Bromham L, Penny D, Rambaut A et al. The power of relative rates tests depends on the data. J Mol Evol 2000; 50(3):296-301.
139. Philippe H, Adoutte A. What can phylogenetic patterns tell us about the evolutionary processes generating biodiversity? In: Hochberg M, Clobert J, Barbault R, eds. Aspects of the Genesis and Maintenance of Biological Diversity. Oxford University Press, 1996:41-59.
140. Wiegmann BM, Yeates DK, Thorne JL et al. Time flies, a new molecular time-scale for brachyceran fly evolution without a clock. Syst Biol 2003; 52(6):745-756.
141. Springer MS, Murphy WJ, Eizirik E et al. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci USA 2003; 100(3):1056-1061.
142. Javaux EJ, Knoll AH, Walter MR. Morphological and ecological complexity in early eukaryotic ecosystems. Nature 2001; 412(6842):66-69.
143. Javaux EJ, Knoll AH, Walter M. Recognizing and interpreting the fossils of early eukaryotes. Orig Life Evol Biosph 2003; 33(1):75-94.
144. Brocks JJ, Logan GA, Buick R et al. Archean molecular fossils and the early rise of eukaryotes. Science 1999; 285(5430):1033-1036.
145. Rohde DL, Olson S, Chang JT. Modelling the recent common ancestry of all living humans. Nature 2004; 431(7008):562-566.
146. Zhaxybayeva O, Gogarten JP. Cladogenesis, coalescence and the evolution of the three domains of life. Trends Genet 2004; 20(4):182-187.

147. Canfield DE, Teske A. Late proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. Nature 1996; 382:127-132.
148. Dacks JB, Doolittle WF. Reconstructing/deconstructing the earliest eukaryotes: How comparative genomics can help. Cell 2001; 107(4):419-425.
149. Mitchell DR. Speculations on the evolution of 9 + 2 organelles and the role of central pair microtubules. Biol Cell 2004; 96(9):691-696.
150. Lwoff A. L'évolution physiologique. Étude des pertes de fonctions chez les microorganismes. Paris: Hermann et Cie, 1943.
151. Gould SJ. Full house: The spread of excellence from plato to darwin. Harmony Books, 1996.
152. Jacob F. Evolution and tinkering. Science 1977; 196(4295):1161-1166.
153. Forterre P, Philippe H. Where is the root of the universal tree of life? BioEssays 1999; 21:871-879.

CHAPTER 3

Origin of Eukaryotic Endomembranes:
A Critical Evaluation of Different Model Scenarios

Gáspár Jékely*

Abstract

All cells can be assigned to one of two categories based on the complexity of cellular organization, eukaryotes and prokaryotes. Eukaryotes possess, among other distinguishing features, an intracellular dynamic membrane system through which there is a constant flow of membranes scaffolded by an internal cytoskeleton. Prokaryotes, however, can have internal membranes, entirely lack a system that resembles eukaryotic endomembranes in terms of dynamics, complexity and the multitude of functions. How and why did the complex endomembrane system of eukaryotes arise? Here I give a critical overview of the different cell biological model scenarios that have been proposed to explain endomembrane origins. I argue that the widely held symbiotic models for the origin of the nuclear envelope and other endomembranes are cell biologically and evolutionarily highly implausible. Recent findings about the origin of nuclear pore complexes also severely challenge such models. I also criticize a scenario of de novo vesicle formation at the origin of the endomembrane system. I contrast these scenarios to traditional and revised autogenous models according to which eukaryotic endomembranes evolved by the inward budding of a prokaryotic cell's plasma membrane. I argue that such models can best satisfy the major constraints of membrane topology, membrane heredity and straightforwardly account for selection pressures while being consistent with genomic findings.

Introduction

Much has been written about the origin of the eukaryotic cell. Interestingly, despite the fact that it is a genuine problem of evolutionary cell biology, a problem of the origin of novel cellular structures, processes and networks, most attention has been focused on phylogenetic aspects, in the most simplistic cases single gene trees. A quick look at the eukaryotic and prokaryotic organellar network (Fig. 1) makes one realize that in order to understand the origin of the novel nodes and connections in the eukaryotic network we need a profound understanding of cell biology. Most scenarios on the origin of eukaryotes are not sufficiently detailed in terms of cell biology, but carry numerous hidden assumptions about how cells function and evolve. A careful analysis of such implicit assumptions from a cell biological perspective can help to decide between the numerous and disparate models of eukaryogenesis.

This chapter and most other chapters in this book deal with the biology and origin of eukaryotic endomembranes. There is overwhelming evidence and a general consensus that an internal and dynamic endomembrane system comprising a nuclear envelope (NE), endoplasmic reticulum (ER), Golgi system, endosomes, phagosomes, lysosomes, autophagosomes,

*Gáspár Jékely—European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. Email:jekely@embl.de

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

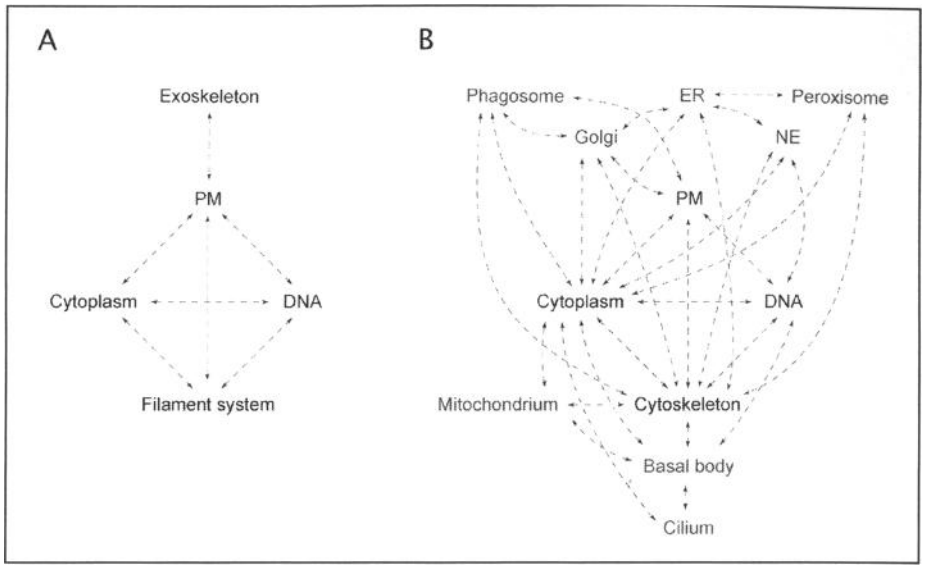


Figure 1. Comparison of the prokaryotic and eukaryotic organellar network. Realistic models of eukaryogenesis have to account for the origin of all novel nodes and the connections between them.

peroxisomes, and mitochondria were present in the last common ancestor of eukaryotes. On the contrary, there is no consensus on how, why and in what order these organelles evolved. What were the underlying molecular and cellular mechanisms? What selection pressures were operating? What was the significance of symbiosis?

Central to all models of endomembrane origins is the problem of the origin of the nucleus. Based on how different model scenarios envisage nuclear origins I will here distinguish and discuss three major model types. (i) Symbiotic scenarios posit that the nucleus evolved from a symbiont (an archaebacterium or enveloped virus) and its envelope is a vestige of either the symbiont's membrane or the host's engulfing membrane. (ii) De novo membrane genesis scenarios propose that the NE and other endomembranes formed by spontaneous lipid vesicle assembly. (iii) Autogenous scenarios state that the NE and other endomembranes evolved via the inward budding of a prokaryotic ancestor's plasmamembrane.

In contrast to the origin of the nucleus and other endomembranes the symbiotic origin of mitochondria and chloroplasts is universally accepted.[1] The evolution of these organelles and other complex endomembrane structure by secondary symbiotic events during eukaryote history will not be dealt with here. The reader is referred to a recent book extensively covering these important subjects.[2] This omission is necessary in order to have a more extensive coverage of other aspects of membrane evolution. It has to be noted, though, that mitochondria in particular had a major and early influence on eukaryote cell evolution, both in terms of genome structure and endomembrane organization. Later symbiotic events had similar wide-ranging effects in specific eukaryote lineages and greatly contributed to the origin of major eukaryotic groups.

Physico-Chemical Constraints of Membrane Topogenesis

Before discussing different models of endomembrane origins it is important to summarize the major constraints of membrane evolution. Membranes are composed of amphipathic molecules that in most cases form bilayers with a hydrophobic interior and two hydrophilic

surfaces. The hydrophobicity of the inner core guarantees that no membrane edges can be exposed to the aqueous solvent. Any membrane compartment in cells has therefore a continuous surface. Topological discontinuity between membranes can arise by the budding off of vesicles or the scission of tubules and sacks. The topologically segregated compartments then form distinct membrane surfaces where the orientation of membrane lipids and proteins is inherited from the mother membrane. Membrane proteins are inserted into the lipid bilayer from the cytoplasmic side by the action of a cytoplasm-oriented protein machinery, the protein conducting channel (see the chapter by T. Schwartz). The self-sustained asymmetry of membrane proteins, besides membrane continuity and discontinuity, is a crucial aspect of membrane topology. Membrane evolution is strongly constrained by the physicochemistry of lipid bilayers and by membrane topology. When considering the evolutionary origin of discontinuous membrane domains these aspects of membrane heredity cannot be neglected.[3] As the reader will see, these constraints, when properly considered, can help to decide between conflicting scenarios.

Symbiotic Scenarios for the Origin of the Nucleus and Endomembranes

The recognition of the mosaic nature of the eukaryotic genome (consisting of genes of either archaebacterial or eubacterial affinity) stimulated the development of a new generation of symbiotic models for the origin of eukaryotes. Although symbiotic scenarios have been proposed long before,[4,5] the recent expansion of genomic data allowed more precise suggestions regarding the nature of the host and its symbiont(s). Some of the symbiotic models have direct relevance to the origin of eukaryotic endomembranes and will be discussed from this perspective. I will argue that the elaboration of these models is not without major cell biological problems.

Sogin[6] hypothesized that eukaryotes evolved from an anuclear proto-eukaryotic lineage distinct from archae- and eubacteria. According to this model the nucleus originated when this proto-eukaryote engulfed an archaebacterium.

Gupta and Golding[7] proposed that a Gram-negative eubacterium that lacked a cell wall engulfed an archaebacterium and endomembranes evolved from the membrane infolds of the host. The plasma membrane of the symbiont that became redundant was eventually lost.

Moreira and López-García[8] developed a similar model in terms of cell biology but with a special emphasis on the mutually advantageous metabolic coupling (syntrophy) of a δ-proteobacterium (similar to Myxobacteria) and a methanogenic archaebacterium. In their scenario, following extensive membrane loss and fusion events the δ-proteobacterial membrane is eventually transformed into the ER and the NE.

Wächtershäuser[9] proposed that eukaryotes evolved when a eubacterium formed a symbiotic merger with a hypothetical precell (a cell not yet completely segregated genetically from other cells). In this model the inner NE is derived from the symbiont's membrane and the outer NE from the engulfing eubacterial membrane. NPCs are thought to have evolved by de novo insertion into these two membranes and the ER is thought to have appeared from the NE because of the faster growth of the outer envelope.

Horiike et al[10] also interpreted results from homology-hit analysis as supporting the symbiotic origin of the nucleus although they don't propose any model scenario on how and why such symbiosis could have happened.

The above models have some key features in common. They posit the nonphagotrophic uptake of a prokaryote (or a precell) by another prokaryote and the loss or radical transformation of the symbiont's plasma membrane to form the NE. Besides, they also require the complete transfer of the host's genome into the symbiont's genome. Below I point out some cell biological difficulties with such transitions and also discuss recent findings that seriously challenge these types of symbiotic models.

Nonphagotrophic Uptake of the Symbiont

The internalization of a bacterium by another prokaryotic cell devoid of a dynamic cytoskeleton and endomembrane system is highly problematic. It is immensely more probable to acquire internal symbionts (the putative proto-nucleus, mitochondria and chloroplasts) for phagotrophic cells that have already evolved endomembrane dynamics. This argument is not challenged by sporadic examples of prokaryotic symbionts within prokaryotes. Even if such internalization is not entirely impossible under certain conditions (e.g., in bacteria that are themselves endosymbionts of eukaryotic cells)[11] the odds are clearly in favor of a phagotrophic host. Phagocytosis-early models of eukaryogenesis fell out of fashion when it became realized that all extant amitochondriate protists once harbored mitochondria. As these amitochondriate phagotrophs do not represent the primitive condition, as formerly assumed, they cannot be considered as 'missing links' in eukaryogenesis. However, this by no means implies that amitochondriate phagotrophs had never existed in the eukaryotic stem lineage and that phagotrophy models are discredited, as often argued (see e.g., refs. 12,13).

The newly emerging rooting of the eukaryotic tree (see refs. 14,15 and the chapter by H. Brinkmann and H. Philippe) indicates that the last common ancestor of eukaryotes was a phagotroph, not an osmotroph.[16] Every model of eukaryogenesis therefore has to account for the origin of phagotrophy. However, none of the symbiotic scenarios is sufficiently developed to explain why a prior endosymbiosis triggered the development of phagotrophy. If the order of origins is reversed, the problem disappears. Phagotrophy can easily account for the acquisition of symbionts.

Loss of the Symbiont's Membrane

One of the major problems with a symbiotic origin of the nucleus is topological. To arrive at the present topology of the NE from a symbiotic event is extremely difficult and one has to posit improbable membrane losses, folding and fusion events. Figure 1 shows four possible transitions, each of them highly dubious regardless of whether the symbiont's plasma membrane contributed or not to NE formation. If it did, membrane topology and the origin of nuclear pore complexes becomes an insurmountable problem (see below). If it did not, it had to be lost, but it is doubtful that the cell could have survived such a drastic and sudden event. Membrane loss couldn't have occurred gradually and loosing the membrane would also have meant the immediate loss of all membrane-associated functions including transport, transmembrane protein insertion, and DNA segregation. Although membrane loss can occasionally occur during cell evolution (e.g., at the origin of Gram positive bacteria or during the evolution of euglenoids, where the periplastid membrane was lost)[3] such events should severely constrain eukaryogenic models. The extreme conservative nature of membranes and the problems of membrane heredity are rarely appreciated and in some scenarios membranes fuse and disappear with incredible liberty and without any topological constraints (see e.g., Fig. 3 in ref. 7).

Transfer of the Host's Genome into the Symbiont's Genome

Although the transfer from a symbiont's genome to the host genome is well documented (e.g., in the case of mitochondria and chloroplasts), gene transfer is much less efficient in the other direction.[17] As Martin and Russell argue convincingly,[18] gene transfer can best occur when the symbiont (present in multiple copies) lyses and liberates DNA into the cytoplasm that can integrate into the host's genome. The reverse process, the lysis of the host would mean the immediate death of the cell and prevent any gene transfer to the symbiont. Even if gene transfer was possible (e.g., via mRNAs and reverse transcription) any scenario of complete genome transfer in conjunction with the loss or rearrangement of the symbiont's plasma membrane is highly unlikely. Genome transfer in principle can be completed either before or after the loss of the symbiont membrane. In the former case (or if the symbiont's membrane is never lost, Fig. 2B) the host cytoplasm has to be maintained entirely by the novel fusion genome, still surrounded with an archaebacterial membrane. To achieve this, the cell has to evolve a

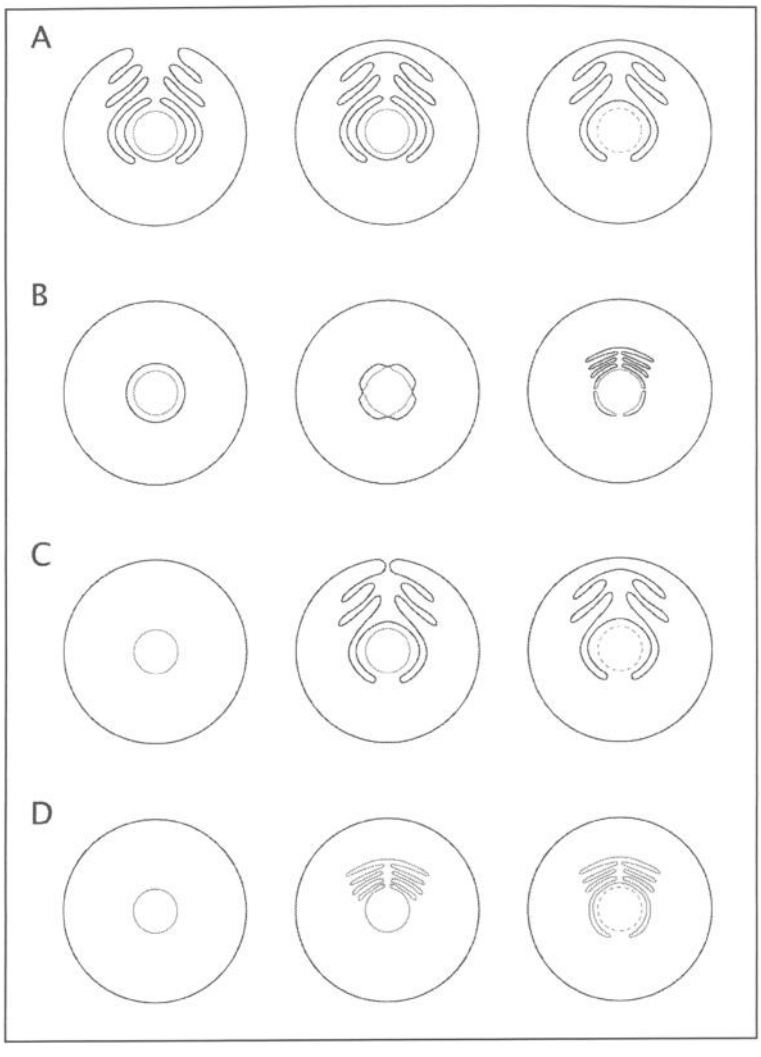


Figure 2. Four possible models for the symbiotic origin of the nucleus. Each scenario is highly problematic and requires drastic topology-breaking membrane rearrangements (see also Table 1). A) The host engulfs the symbiont that later looses its plasmamembrane. The engulfing membranes then have to break and refuse to create the NE and ER membranes. B) The host engulfs the symbiont and fuses its engulfing membrane with the symbiont's membrane. The formation of nuclear pores and the ER requires the breaking of both membranes, their subsequent refusion and outward budding. C) The symbiont is fully internalised and later surrounded by membranes from the host plasmamembrane. The symbiont's membrane is subsequently lost. D) The symbiont is fully internalised and buds off membrane tubules to form the ER. For NE evolution the membrane has to break and refuse with itself.

sophisticated system of mRNA and ribosome export and protein import across the symbiont's membrane. It has to evolve NPCs and a nuclear transport system. However, it is almost impossible to do so from the symbiont's archaebacterial membrane, due to the constraints of membrane topology (Fig. 2). The origin of NPCs from vesicle coats[19] also contradicts such scenarios

Table 1. Cell biologically difficult evolutionary transitions encountered in the different symbiotic model scenarios shown in Figure 1 as contrasted to autogenous scenarios (see text for details)

	Symbiotic; A	Symbiotic; B	Symbiotic; C	Symbiotic; D	Autogenous
Loss of Gram negative host's outer membrane	*Yes*	*Yes*	*Yes*	*Yes*	No
Nonphagocytic uptake of the symbiont	*Yes*	*Yes*	*Yes*	*Yes*	No
Loss or unlikely rearrangement of symbiont's plasma membrane	*Yes*	*Yes*	*Yes*	*Yes*	No
Total transfer of host's genome into the symbiont	*Yes*	*Yes*	*Yes*	*Yes*	No
Fusion and partial loss of host's engulfing membrane	*Yes*	*Yes*	No	No	No
SecY in opposite orientation	No	No	No	*Yes*	No
Fusion of archaebacterial and eubacterial membrane	No	*Yes*	No	No	No
Problems concerning NPC and secretory membrane origins	*Yes*	*Yes*	*Yes*	*Yes*	No

(see below). We are therefore left with the other possibility: the symbiont membrane had to be lost before the completion of the genome transfer. This results in a cytoplasm with two genomes, two types of DNA segregation machines, and two sets of ribosomes. It is highly dubious how such a chimeric cytoplasm can ever be viable and evolvable.

Origin of the Nuclear Pore Complex

Nothing in recent years was so illuminating with respect to endomembrane origins than the discovery that NPCs and vesicle coat complexes are evolutionarily related.[19,20] This recognition introduced a novel, inescapable constraint that immediately and elegantly refuted a number of symbiotic models. It was found based on structural predictions and comparative analyses that components of the Nup107-160 (Nup84 in yeast) subcomplex of the NPC are related to subunits of COPI, COPII, and clathrin vesicle coats. These proteins all share a unique combination and arrangement of two distinct structural modules: an N-terminal β-propeller and a C-terminal α-solenoid. Based on the structural similarity it has been suggested that the corresponding complexes evolved from a prototypic membrane-curving module, the protocoatomer, in an early eukaryote.[19]

The protocoatomer model not only links all vesicle coats and the NPC through a common ancestor but also argues for the evolutionary continuity of the corresponding membrane domains; i.e., the ER, the Golgi, and the nuclear envelope. According to the protocoatomer model a primary secretory/endocytic compartment and its actively budding coated vesicles predated the origin of the nucleus. NPCs can therefore be understood as defective vesicle coats that prevented vesicle fusion around the secondarily forming nuclear compartment.[21]

The origin of NPCs from vesicle coats directly refutes models where NPCs are thought to have evolved by the insertion of channels or pores into a preexisting double nuclear membrane (e.g., ref. 9). Other variants of symbiotic scenarios can also hardly account for the early origin of membrane trafficking and the link between NPCs and vesicle coats. These models would then have to posit that the membranes surrounding the symbiont were actively budding off coated vesicles and had secretory or endocytic function while internalizing a foreign cell. However, if these assumptions are made, we are back to phagotrophic models that symbiotic models originally tried to challenge.

Interestingly, the intraflagellar transport complex that assembles and maintains eukaryotic cilia has also been shown to be a protocoatomer derivative.[22,23] This clearly refutes symbiotic scenarios that posit a common origin of cilia and the nucleus from Spirochete ancestors.[24]

Table 1 summarizes the cell biological difficulties encountered in different symbiotic model scenarios and contrasts them to autogenic ones (see below). Taken together, the week aspect of symbiotic models for the origin of eukaryotic endomembranes is in their cell biology. The idea of such hypothetical cellular mergers had rather been stimulated by genomic data, than cell biological considerations. They therefore elegantly explain the origin of the chimeric eukaryotic genome (that autogenous scenarios also do), but not that of eukaryotic endomembranes.

Origin of Endomembranes by de novo Vesicle Formation

Martin and Müller proposed that eukaryotes evolved as a result of a hydrogen exchange-driven symbiosis between an archaebacterium and an α-proteobacterium, the ancestor of mitochondria.[12] As an extension to this 'hydrogen hypothesis' Martin proposed that eukaryotic endomembranes evolved de novo in the archaebacterial cytoplasm as the α-proteobacterial genes for fatty acid ester lipid biosynthesis got transferred into the host's genome.[25] The activity of these genes led to the sudden appearance of a novel lipid synthesis pathway in the host's cytoplasm (bounded by the archaebacterial isoprenyl ether lipid membrane). The freshly synthesized fatty acid ester lipids spontaneously formed lipid micelles that gradually differentiated into the eukaryotic endomembrane system. The eubacterial lipids eventually replaced the host's archaebacterial membrane lipids.

This model is topologically workable yet it has severe cell biological and evolutionary shortcomings. First, it is problematic why the eubacterial lipids formed vesicles de novo and didn't integrate into either the host's or the symbiont's plasma membrane. It is not clear what physicochemical barrier would have prevented the formation of membranes with mixed lipid composition. But even if one assumes that de novo vesicle formation had occurred, it is unclear why this had not been selected against (newly formed functionless lipid vesicles must have represented a severe disadvantage). Selection could either have prevented the transfer of the whole biosynthetic pathway or attenuated enzyme expression to a level that suffices to sustain the growth of the α-proteobacterial plasmamembrane.

Another serious problem concerns the origin from these lipid vesicles of a functional secretory compartment. Secretion and transmembrane protein synthesis cannot work without a system of transmembrane protein insertion and translocation (see the chapter by Thomas Schwartz). In most cellular membranes where secretion and transmembrane protein insertion occurres (including eubacteria, archaebacteria, and eukaryotes) the process is mediated by the universal protein conducting channel, the trimeric Sec61/SecY complex.[26] It is hard to conceive how and why this machinery got inserted into the newly formed endomembranes devoid of any protein component. A protein conducting channel can only be inserted into membranes where other protein conducting channels are already present to mediate such insertion. This severe constraint is a chief argument against de novo membrane generation that probably only happened once, at the origin of the first cellular membranes, before the divergence of all extant cellular life.

Autogenous Scenarios

None of the cell biological difficulties discussed above is encountered in autogenous models. Such models state that eukaryotic endomembranes evolved by the inward folding (invagination, tubulation, or vesiculation) of the host cell's plasma membrane.[19,27-32] The host acquired the proto-mitochondrium after it had evolved an endomembrane system sufficiently complex to support the internalization of entire cells. Assuming as host a cell descending from the immediate common ancestor of the sister groups archaebacteria and eukaryotes (or, less likely, a full-fledged archaebacterium) as well as extensive gene transfer from phagocytic prey[33] and the α-proteobacterial symbiont such models can account for the chimeric nature of the eukaryotic genome. These models also provide a cell biologically realistic framework for the understanding of the functioning and topology of the endomembrane system such as a continuous NE and secretory ER. They also provide selection pressure (increased efficiency of predation) and molecular mechanisms (e.g., coated vesicle origins and diversification) that could have operated during endomembrane evolution.

Different versions of autogenic models disagree about the mechanism and order of transformations that allowed more efficient predation and phagocytosis. For example, the development of secretory endomembranes, and not direct cell internalization, has recently been proposed to have initiated endomembrane evolution.[32] Regardless of these differences, the reader has to keep in mind that the selection pressure assumed to operate in all autogenic models is the same: to digest and eat other cells. As expressed by Stanier:[28] "the progressive evolution of the eukaryotic cell received its initial impetus from the acquisition of a novel cellular property, *the capacity to perform endocytosis*. The capacity for endocytosis would have conferred on its early possessors a new biological means for obtaining nutrients: predation on other cells."

Membrane Topogenesis in Autogenous Models

Autogenous models propose that eukaryotic endomembranes evolved from the plasma membrane by inward budding and subsequent topological separation. Different autogenic models disagree about the nature and function of the first endomembranes but agree about the major steps of membrane topogenesis. All autogenous models have the following cell biological constraints: i) No intracellular compartment could have segregated before the origin of transport between the topologically segregated membranes. Only this could have allowed balanced membrane growth and turnover. ii) As the secretory endomembrane system segregated topologically from the plasma membrane it had to contain the ribosome docking apparatus. The topological segregation and the redirection of cotranslational protein transport from the plasma membrane to these topologically segregated endomembranes was a key event during the origin of eukaryotes.

According to autogenous scenarios the nucleus evolved as vesicles and tubules of secretory endomembranes surrounded the cell's chromatin. The imperfect fusion of ER-derived vesicles and tubules around chromatin resulted in the double-membrane topology of the NE with continuity between the inner and outer nuclear membranes. The continuous membrane connections between the two NE leaflets are covered by NPCs from the cytoplasmic side. NPCs therefore cover the same side of the membrane in nuclear pores as the related vesicle coats in coated vesicles (see above). The topology of SecY/Sec61 channels at the rough ER and the outer NE also agrees with the topology of NPCs and coated vesicles.

The relatively late origin of the NE is also clear from simple topological considerations. 'Primary compartments' (ER, Golgi, endosomes) are surrounded by a single membrane and their lumenal content is developmentally continuous with the outside medium (ectoplasm). The nucleoplasm on the other hand is continuous with the cytoplasm (endoplasm).[27] The nucleus is a 'secondary compartment' that is generated by the imperfect fusion of membrane domains of primary compartments (ER vesicles). The intra-NE space is therefore developmentally continuous with the ectoplasm of the ER lumen and also the outside medium.

Nature of the First Endomembranes

The first detailed autogenous models proposed that the origin of nutrient uptake, either by endocytosis or phagocytosis (i.e., the complete internalization of other cells) was the initial step in the evolution of the endomembrane system.[28,34] The suggestion of the early origin of a true phagocytic compartment is very appealing and comes into ones mind first. However, phagocytosis, the engulfment and digestion of entire cells, requires the coordination of at least three processes: i) sensing and binding of prey, ii) membrane remodeling around the prey, iii) secretion of digestive enzymes and food uptake. The question arises in what order did these elementary steps evolve. Clearly, membrane remodeling is useless if the prey is not digested and absorbed. On the contrary, prey binding, digestion and food uptake can happen, even if not very efficiently, without the internalization of prey. Such considerations led to the idea that the elaboration of a membranous secretory system was the first step in the origin of eukaryotic endomembranes.[32] Christian de Duve expressed similar views about the early nature of eukaryotic endomembranes.[29,35] According to him endomembranes evolved by "infolding of the cell membrane, allowing the formation of internalized extracellular pockets into which captured food and secreted enzymes were trapped together". These infoldings "made larger membrane areas available for nutritive exchanges and allowed such exchanges to take place efficiently with deep-seated portions of the cell". A possible autogenous scenario via membrane tubulation and topological segregation is shown in Figure 3.

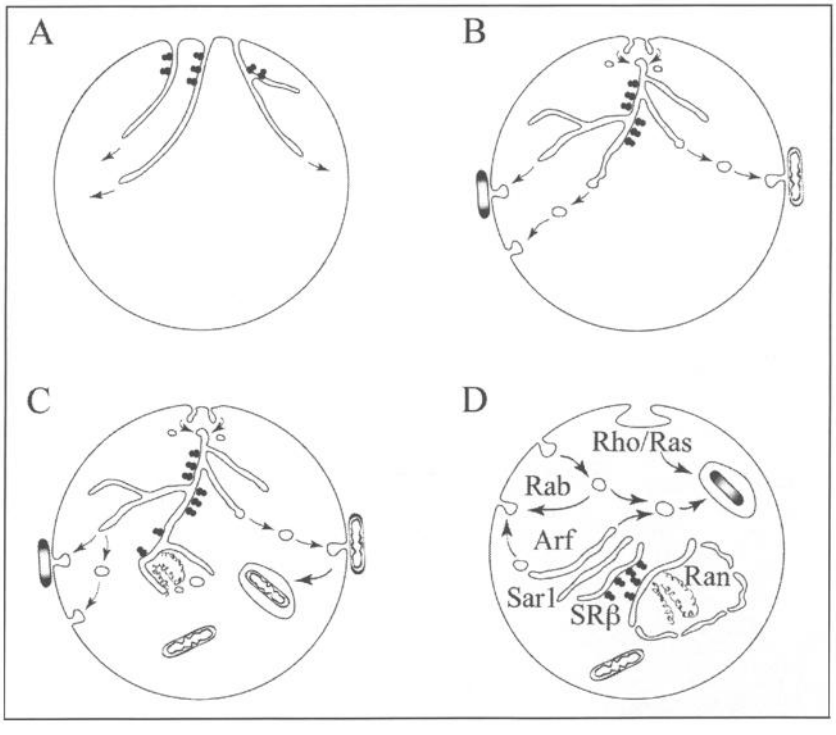


Figure 3. A possible autogenous scenario for the origin of eukaryotic endomembranes. A) Secretory membranes evolved first by membrane tubulations. B) Topological separation from the plasma membrane occurred by the fission of the tubules. Vesicular trafficking evolved to sort proteins and recycle lipids. C) Phagocytosis evolved later and led to the endosymbiotic origin of mitochondria. D) The nuclear membrane evolved when secretory membranes surrounded chromatin. Representative members of the small GTPase family are shown in proximity to the respective compartment where they act.

Figure 4, viewed on following page. Bayesian tree of eukaryotic small GTPases. The tree was rooted with the Gαs subunit of eukaryotic trimeric G-proteins. The major function carried out by the different subfamilies is shown on the right. Nodes marked with an arrowhead have a posterior probability of 1.

The early origin of secretory endomembranes is supported by the evolutionary history of eukaryotic small GTPases. These proteins are central regulators of endomembrane and cytoskeleton dynamics and diversified during early eukaryote evolution, in the eukaryotic stem lineage.[32] Given the poor taxon sampling in my original trees[32] I repeated the analysis with a much broader sampling using Bayesian estimation of phylogeny[36] (Fig. 4). For rooting I used the α subunit of trimeric G proteins that are more closely related to Ras-like small GTPases than eubacterial GTPases. The tree shows the early divergence of secretory small GTPases including Sar1, and Arfs. SRβ (not included here because of its derived sequence), a regulator of eukaryotic cotranslational ribosome targeting to the ER, also belongs to the Sar1/Arf clade.[32] Besides suggesting an early origin of secretory membranes the small GTPase tree also indicates the origin and diversification of major eukaryotic innovations,[32] including the nucleus and cilia (see also ref. 23).

Mechanisms of Endomembrane Generation

The development of discontinuous membrane compartments requires the generation of high membrane curvature and subsequent fission. These events are mediated by cytoplasmic protein machineries and by the chemical modification of membrane composition. For example amphipathic peptides that partially penetrate the lipid bilayer can directly deform membranes.[37,38] Eukaryotes evolved a spectacular capacity to physically curve and fuse lipid bilayers. The origin of these capacities was a necessary prerequisite for the evolution of endomembranes.

In vitro studies revealed that even single proteins can deform membranes. For example Epsin and endophilin, regulators of clathrin-mediated endocytosis, can directly bind and convert liposomes into tubules.[37,39] The small GTPase Sar1, in its active GTP-bound form can also transform liposomes into tubules.[40] This effect is dependent on the N-terminal amphipathic helix of Sar1. Membrane tubulation can also be induced by molecular motors bound to and pulling on membranes.[41,42] Interestingly, tubule formation also leads to lipid phase separation if a mixture of different lipids is used.[43] For a detailed discussion of the evolutionary significance of membrane tubulation see the chapter by Alexander Mironov et al.

Experiments with in vitro systems have immediate relevance for models of membrane evolution. Such studies first of all constrain models by showing what is physically possible. On the other hand they also indicate how minimal systems may have looked like and gradual changes occurred during evolutionary transitions.[32] Membrane tubule formation could in principle have evolved following cell wall loss and a few mutations in a single molecule (e.g., origin of an amphipathic helix). The phase separation of lipids in membrane tubules hints at how membrane domains may have differentiated during early endomembrane evolution.

Coevolution of Eukaryotic Endomembranes and Cytoskeleton

The autogenous origins of eukaryotic endomembranes cannot be understood without understanding the evolution of the cytoskeleton. The endomembrane system necessarily evolved in synergy with the cytoskeleton since membrane dynamics requires the action of molecular motors and the cytoskeleton scaffold.

The precursors of the microfilament and microtubule systems (actin, tubulin) can be found in prokaryotes. The prokaryotic FtsZ protein, a tubulin homologue, assembles into dynamic polymers and forms a contracting ring during cell division.[44-46] MreB and ParM resemble actin in their atomic structure and also form filaments.[47,48] MreB is important for the maintenance of cell shape whereas ParM drives movement of plasmids in the cell.[49] Membrane dynamics is strongly limited in prokaryotes because of the rigid exoskeleton. This changed radically after the loss of the

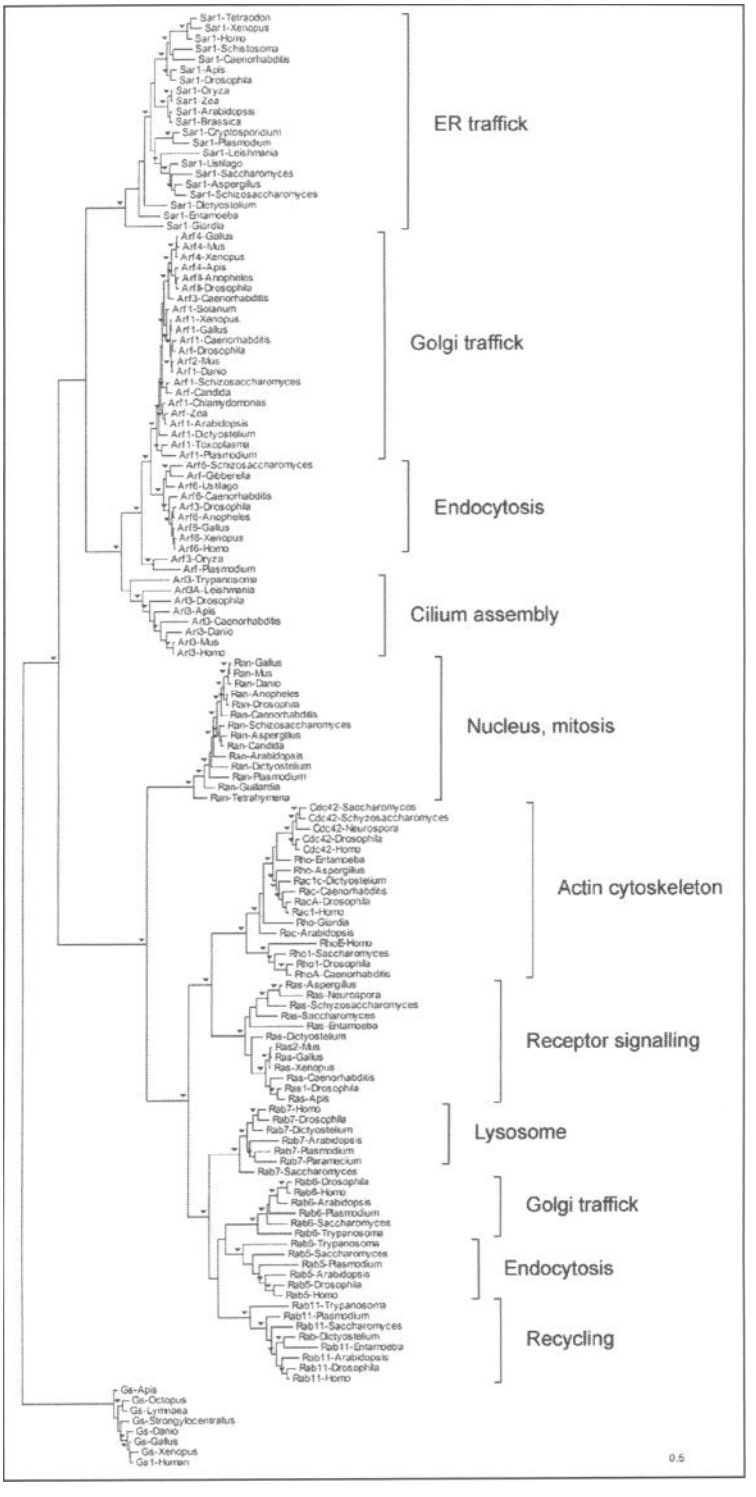


Figure 4. Please see legend on previous page.

cell wall and the elaboration of the endoskeleton and endomembrane system. As the physical constraints imposed on the plasma membrane by cell wall rigidity were released an almost unrestricted exploration of new avenues ensued. The plasma membrane could be tubulated, vesiculated, protruded, intracellular membrane dynamics and lamellipodial motility evolved. The novel physical properties of the plasma membrane provided a novel environment for the filament and motor systems. This could have triggered the rapid evolution of these systems as well, to internally stabilize the cell, and scaffold and move endomembranes. A relatively brief period of quantum evolution can explain the low similarity to prokaryotic homologs and the nonclock-like evolution of the eukaryotic filament system.

Understanding how the complex cytoskeleton of eukaryotes evolved also requires the understanding of functional minimal systems and possible gradual evolutionary steps. In vitro and in silico studies of cytoskeletal components can be very informative also in this case, e.g., to understand the evolution of microtubule asters and mitotic spindles[50,51] or beating cilia.[52]

Interspecific Interactions at the Origin of Eukaryotic Endomembranes

Several models lay great emphasis on interspecific interactions during eukaryote origins. The presumed interactions in putative ecological settings range from H_2 exchange-driven syntrophy to predation or parasitism. In the syntrophy hypothesis[8] methanogenic archaebacteria and sulfate-reducing δ-proteobacteria form permanent consortia driven by interspecific H_2 transfer. The close interspecific cell-cell interactions are thought to have led to extensive membrane development and the formation of a proto-nucleus. In the hydrogen hypothesis[12] interspecific association of an anaerobic, H_2-dependent archaebacterium and a H_2-producing eubacterium drives cellular merger. In phagotrophy models[28-32] the interspecific interaction that drives endomembrane development is predation, through capture, internalization and digestion of other cells. In the framework of these models the origin of mitochondria is a consequence of phagotrophic cell internalization and the evolution of intracellular enslavement or mutualism.

All of the above ecological models have parallels in extant microorganisms. For example H_2 transfer is known to drive the formation of symbiotic consortia in prokaryotes.[8,53] Phagotrophy is also widespread in eukaryotes and the cenancestor must have exercised it.[15] The problem with syntrophy models (besides topological ones pointed out above) therefore is that they have to presume the operation of two entirely different ecological driving principles - syntrophy and phagotrophy. Phagotrophy models are more elegant in this respect, as they explain eukaryote origins as a result of one ecological principle, one that is immensely important and widespread in extant eukaryotes (quite contrary to H_2 syntrophy).

The origin of an efficient endomembrane-based digestion and internalization system opened up an entirely new niche in the history of life and largely explains the origin of eukaryotes, including the origin of mitochondria. One may even look at it as an unavoidable evolutionary transition. In agreement with simple ecological principles, abundant microbial life led to the development of professional heterotrophic predators. Clearly the most efficient way of predation is to fully ingest and digest prey. The transition might not have been easy and straightforward but the selection pressure was constant and eventually the most efficient feeders won the day: the eukaryotes.

Acknowledgements

I am grateful to Balázs Papp for his comments and to István Molnár for extensive discussions and insightful comments on the manuscript. The author is supported by a FEBS Long Term fellowship.

References

1. Gray MW. The endosymbiont hypothesis revisited. Int Rev Cytol 1992; 141:233-357.
2. Hirt RP, Horner DS. Organelles, genomes and eukaryote phylogeny: An Evolutionary Synthesis in the Age of Genomics. London: Taylor and Francis, 2004.
3. Cavalier-Smith T. The membranome and membrane heredity in development and evolution. In: Hirt RP, Horner DS, eds. Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics. Vol 68. London: Taylor and Francis, 2004:335-351.
4. Mereschkowsky C. Über natur und ursprung der chromatophoren im pflanzenreiche. Biol Centralbl 1905; 25:593-604.
5. Sagan L. On the origin of mitosing cells. J Theor Biol 1967; 14(3):255-274.
6. Sogin ML. Early evolution and the origin of eukaryotes. Curr Opin Genet Dev 1991; 1(4):457-463.
7. Gupta RS, Golding GB. The origin of the eukaryotic cell. Trends Biochem Sci 1996; 21(5):166-171.
8. Moreira D, López-García P. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. J Mol Evol 1998; 47(5):517-530.
9. Wachtershauser G. From precells to Eukarya—a tale of two lipids. Mol Microbiol 2003; 47(1):13-22.
10. Horiike T, Hamada K, Kanaya S et al. Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. Nat Cell Biol 2001; 3(2):210-214.
11. von Dohlen CD, Kohler S, Alsop ST et al. Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. Nature 2001; 412(6845):433-436.
12. Martin W, Muller M. The hydrogen hypothesis for the first eukaryote. Nature 1998; 392(6671):37-41.
13. Gray MW, Burger G, Lang BF. Mitochondrial evolution. Science 1999; 283(5407):1476-1481.
14. Stechmann A, Cavalier-Smith T. Rooting the eukaryote tree by using a derived gene fusion. Science 2002; 297(5578):89-91.
15. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436(7054):1113-1118.
16. Martin W, Rotte C, Hoffmeister M et al. Early cell evolution, eukaryotes, anoxia, sulfide, oxygen, fungi first (?), and a tree of genomes revisited. IUBMB Life 2003; 55(4-5):193-204.
17. Thorsness PE, Fox TD. Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. Nature 1990; 346(6282):376-379.
18. Martin W, Russell MJ. On the origins of cells: A hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos Trans R Soc Lond B Biol Sci 2003; 358(1429):59-83, (discussion 83-55).
19. Devos D, Dokudovskaya S, Alber F et al. Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol 2004; 2(12):e380.
20. Mans BJ, Anantharaman V, Aravind L et al. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. Cell Cycle 2004; 3(12):1612-1637.
21. Jékely G. Glimpsing over the event horizon: Evolution of nuclear pores and envelope. Cell Cycle 2005; 4(2):297-299.
22. Avidor-Reiss T, Maer AM, Koundakjian E et al. Decoding cilia function: Defining specialized genes required for compartmentalized cilia biogenesis. Cell 2004; 117(4):527-539.
23. Jékely G, Arendt D. Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. Bioessays 2006; 28(2):191-8.
24. Margulis L, Dolan MF, Guerrero R. The chimeric eukaryote: Origin of the nucleus from the karyomastigont in amitochondriate protists. Proc Natl Acad Sci USA 2000; 97(13):6954-6959.
25. Martin W. A briefly argued case that mitochondria and plastids are descendants of endosymbionts, but that the nuclear compartment is not. Proc Roy Soc Lond B 1999; 266:1387-1395.
26. Osborne AR, Rapoport TA, van den Berg B. Protein Translocation by the Sec61/Secy Channel. Annu Rev Cell Dev Biol 2005; 21:529-550.
27. Blobel G. Intracellular protein topogenesis. Proc Natl Acad Sci USA 1980; 77(3):1496-1500.
28. Stanier RY. Some aspects of the biology of cells and their possible evolutionary signicance. Symp Soc Gen Microbiol 1970; 20:1-38.
29. De Duve C. Evolution of the peroxisome. Ann NY Acad Sci 1969; 168(2):369-381.
30. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 2002; 52(Pt 2):297-354.
31. Cavalier-Smith T. The origin of cells: A symbiosis between genes, catalysts, and membranes. Cold Spring Harb Symp Quant Biol 1987; 52:805-824.
32. Jékely G. Small GTPases and the evolution of the eukaryotic cell. Bioessays 2003; 25(11):1129-1138.
33. Doolittle WF. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 1998; 14(8):307-311.

34. Cavalier-Smith T. The origin of eukaryotic and archaebacterial cells. Ann NY Acad Sci 1987; 503:17-54.
35. De Duve C, Wattiaux R. Functions of lysosomes. Annu Rev Physiol 1966; 28:435-492.
36. Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003; 19:1572-1574.
37. Ford MG, Mills IG, Peter BJ et al. Curvature of clathrin-coated pits driven by epsin. Nature 2002; 419(6905):361-366.
38. McMahon HT, Gallop JL. Membrane curvature and mechanisms of dynamic cell membrane remodelling. Nature 2005; 438(7068):590-596.
39. Farsad K, Ringstad N, Takei K et al. Generation of high curvature membranes mediated by direct endophilin bilayer interactions. J Cell Biol 2001; 155(2):193-200.
40. Lee MC, Orci L, Hamamoto S et al. Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. Cell 2005; 122(4):605-617.
41. Roux A, Cappello G, Cartaud J et al. A minimal system allowing tubulation with molecular motors pulling on giant liposomes. Proc Natl Acad Sci USA 2002; 99(8):5394-5399.
42. Koster G, VanDuijn M, Hofs B et al. Membrane tube formation from giant vesicles by dynamic association of motor proteins. Proc Natl Acad Sci USA 2003; 100(26):15583-15588.
43. Roux A, Cuvelier D, Nassoy P et al. Role of curvature and phase transition in lipid sorting and fission of membrane tubules. Embo J 2005; 24(8):1537-1545.
44. Lowe J, Amos LA. Crystal structure of the bacterial cell-division protein FtsZ. Nature 1998; 391(6663):203-206.
45. Lowe J, Amos LA. Tubulin-like protofilaments in Ca2+-induced FtsZ sheets. Embo J 1999; 18(9):2364-2371.
46. Ben-Yehuda S, Losick R. Asymmetric cell division in B. subtilis involves a spiral-like intermediate of the cytokinetic protein FtsZ. Cell 2002; 109(2):257-266.
47. van den Ent F, Amos LA, Lowe J. Prokaryotic origin of the actin cytoskeleton. Nature 2001; 413(6851):39-44.
48. Moller-Jensen J, Jensen RB, Lowe J et al. Prokaryotic DNA segregation by an actin-like filament. Embo J 2002; 21(12):3119-3127.
49. Moller-Jensen J, Lowe J. Increasing complexity of the bacterial cytoskeleton. Curr Opin Cell Biol 2005; 17(1):75-81.
50. Surrey T, Nedelec F, Leibler S et al. Physical properties determining self-organization of motors and microtubules. Science 2001; 292(5519):1167-1171.
51. Nedelec F. Computer simulations reveal motor properties generating stable antiparallel microtubule interactions. J Cell Biol 2002; 158(6):1005-1015.
52. Camalet S, Jülicher F, Prost J. Self-organized beating and swimming of internally driven filaments. Physical Review Letters 1999; 82(7):1590-1593.
53. Bryant MP, Wolin EA, Wolin MJ et al. Methanobacillus omelianskii, a symbiotic association of two species of bacteria. Arch Mikrobiol 1967; 59(1):20-31.

CHAPTER 4

Origins and Evolution of Cotranslational Transport to the ER

Thomas U. Schwartz*

Abstract

All living organisms possess the ability to translocate proteins across biological membranes. This is a fundamental necessity since proteins function in different locations yet are synthesized in one compartment only, the cytosol. Even though different transport systems exist, the pathway that is dominantly used to translocate secretory and membrane proteins is known as the cotranslational transport pathway. It evolved only once and is in its core conserved throughout all kingdoms of life. The process is characterized by a well understood sequence of events: first, an N-terminal signal sequence of a nascent polypeptide is recognized on the ribosome by the signal recognition particle (SRP), then the SRP-ribosome complex is targeted to the membrane via the SRP receptor. Next, the nascent chain is transferred from SRP to the protein conducting channel, through which it is cotranslationally threaded. All the essential components of the system have been identified. Recent structural and biochemical studies have unveiled some of the intricate regulatory circuitry of the process. These studies also shed light on the accessory components unique to eukaryotes, pointing to early events in eukaryotic evolution.

Introduction

A system to integrate or translocate proteins into or across lipid bilayers is a fundamental requirement for all autonomous cellular life forms. Proteins destined for membrane translocation are synthesized with an N-terminal signal sequence composed of about 10 hydrophobic residues.[1] The signal sequence of a nascent protein is recognized as it emerges from the exit tunnel of the ribosome by the signal recognition particle (SRP), a large RNA-protein complex. Next, the SRP-bound ribosome-nascent-chain complex (SRP-RNC) is targeted to the membrane through the SRP receptor (SR). Once localized to the membrane, the nascent chain is transferred from the SRP to the protein conducting channel (PCC). Secretory proteins are threaded through the PCC and adopt their native structure on the other side of the membrane, whereas the PCC can open laterally to insert membrane proteins into the lipid bilayer. In cotranslational transport protein synthesis and translocation are temporally closely coupled. Such synchronization needs regulation, which is provided by guanine nucleotide binding proteins (G proteins) acting as molecular switches.

Cotranslational transport has been studied extensively over the past 30 years.[2] After identifying the essential molecules governing the process we have learned a great deal in recent years from x-ray crystallographic and cryo-electron microscopic studies. These structural

*Thomas U. Schwartz—Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, U.S.A. Email: tus@mit.edu

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

characterizations have provided snapshots of components of the system in various functional states and gradually unveil the mechanistic underpinning. The reader is referred to excellent recent reviews on the subject.[3-6] This review focuses on differences between the mediators of cotranslational transport in the three kingdoms of life. Even though universally conserved, cotranslational transport has become increasingly more sophisticated over time. Some features are only found in eukaryotes and it is likely that these features point to early events in eukaryogenesis.

Mediators of Cotranslational Transport

The Signal Recognition Particle

Signal recognition particles are conserved throughout all kingdoms of life[7] and were initially discovered in mammalian cells.[8] In mammals and other metazoans, SRP consists of a ~300nt (90kD) elongated and largely double-stranded RNA scaffold structure to which six proteins bind. These proteins are named according to their apparent molecular mass (in kD) SRP9, 14, 19, 54, 68 and 72. SRP is organized in two functionally distinct regions, the Alu- and the S-domain, that form two ends of the SRP RNA (Fig. 1). The S-domain, in addition to the architectural proteins SRP19, 68 and 72, also harbors the regulatory G protein SRP54. SRP54 recognizes the signal sequence of an emerging nascent chain at the exit site of the ribosome and also targets SRP-RNC to the membrane via interaction with SR. SRP54 is made of two functional domains and belongs to the small GTPase superfamily.[9] The N-terminus of SRP54 contains a composite domain made of a four-helix bundle (N domain) tightly connected to a G domain. The NG domain is universally conserved in SRP54, repeated in the SRs, but is not found in any other cellular context. The second functional domain of SRP54 is C-terminal, flexibly linked to the NG domain, and called the M domain for its methionine-rich composition. The M domain directly binds SRP RNA (Fig. 1A,B) as well as the signal sequence. The SRP RNA Alu domain binds heterodimeric SRP9/14 and is responsible for attenuating elongation by blocking the elongation-factor binding site upon SRP binding of RNC.[10]

Figure 1, viewed on next page. Components of the cotranslational transport pathway. A) Composite structure model of the eukaryotic signal recognition particle based on cryo-electron microscopic data (PDB code 1RY1). The SRP-RNA helices are labeled h3 through h8. SRP proteins for which crystal structures exist are included. SRP68 and SRP72 are missing and are expected to bind in the h5 region. SRP54 is the only universally conserved protein of SRP and its domains are labeled N, G, and M. In Gram-positive bacteria, like *Bacillus subtilis* SRP-RNA helices h6/h7 and SRP19 are missing (purple region) and SRP9/14 is functionally replaced by HBsu (not shown). In Gram-negative bacteria, like *Escherichia coli*, the SRP is solely composed of the SRP-RNA segment shown in red and SRP54. B) Crystal structure of SRP54 bound to its cognate SRP-RNA helix h8 (PDB code 1QZW). The M domain is responsible for RNA binding and signal peptide recognition. The composite NG domain is a regulatory GTPase. The insertion box domain (orange) and the N domain are the unique features of this GTPase family. C) Structure of the archaeal Ffh-FtsY NG domain complex (PDB codes 1RJ9 and 1OKK). This so-called engagement complex is the regulated interface between SRP and SR. Stable interaction of the two domains is GTP-dependent (red). Both GTP molecules are also necessary for mutual hydrolysis and subsequent separation of both domains. Ffh is the archaeal ortholog of SRP54, FtsY-NG the ortholog of SRα-NG. D) Structure of the eukaryotic SRP receptor (PDB code 1NRJ). SRβ is a G protein and unique to eukaryotes. SRα interacts with SRβ through its N-terminal SRX domain (yellow) in GTP-dependent fashion. SRβ is anchored to the ER membrane via an N-terminal TM helix not part of the crystal structure. E) Structure of the closed archaeal protein conducting channel (PDB code 1RH5) viewed from the cytosol. The α-subunit is the core of the structure and forms two half shells that are clamped together by the γ-subunit. The TM-helix of the β-subunit is peripherally attached to the α-subunit. A short helix of the α-subunit is positioned in the center of the structure and forms a plug that is proposed to move away during translocation of a peptide chain through the central pore.

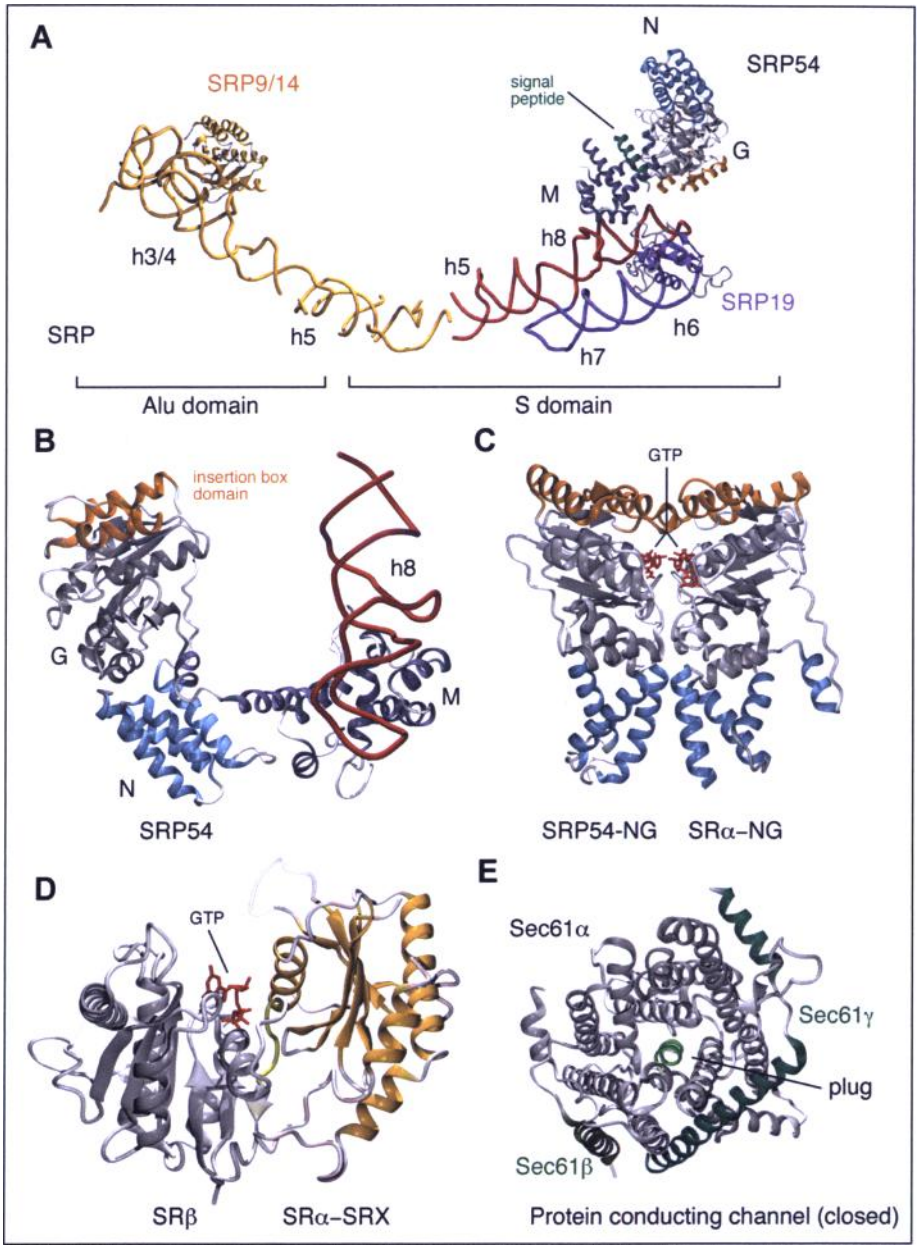


Figure 1. See figure legend on previous page.

The composition of SRP is largely conserved among all eukaryotes, however some irregularities occur in fungi and protozoans. In *Saccharomyces cerevisiae* SRP-RNA contains additional insertions in the Alu domain of yet unknown function.[11] In several fully sequenced eukaryotic parasites some of the architectural SRP proteins have not been identified. For example, in *Leishmania major, Trypanosoma cruzi,* and *Giardia lamblia* no SRP9 or SRP14 is found. *Encephalitozoon cuniculi* appears to lack SRP68 and SRP72.

The SRP repertoire of prokaryotes is much simpler then in eukaryotes. In archaea, only SRP19 and SRP54 are found.[12] Gram-negative bacteria, such as *Escherichia coli,* possess the simplest version of SRP. Here the particle contains a ~110 nt RNA and only the SRP54 ortholog Ffh. *E.coli* SRP can functionally replace its mammalian counterpart in vitro indicating the essential role of the G protein SRP54 for cotranslational transport.[13]

The SRP Receptor

SRs are an essential component of cotranslational transport pathways[14] and are thus phylogenetically well conserved. In bacteria and archaea, neither of which contain intracellular organelles, SR is a single-subunit protein termed FtsY. FtsY homologs share a C-terminal NG domain, which directly interacts with the NG domain of Ffh (the SRP54 ortholog) in the targeting reaction (Fig. 1C). The N-terminal region of FtsY is not strongly conserved. *E. coli* has a glutamate-rich region called A domain, which is able to reversibly attach to the plasma membrane.[15] Instead of an A domain homolog, some Gram-positive bacteria carry a transmembrane (TM) helix at the N-terminus of FtsY, thus permanently anchoring the receptor to the cytoplasmic membrane.[16] Replacing the A region of *E.coli* FtsY with a TM helix of an unrelated protein resulted in a fully functional receptor.[17]

In contrast to the single-subunit prokaryotic SR, eukaryotic SR is a heterodimer of an α-subunit (SRα) and a β-subunit (SRβ). SRα is homologous to FtsY in that it also contains a C-terminal NG domain. However, instead of the N-terminal membrane attachment domain, all eukaryotic SRαs contain a conserved domain SRX that directly interacts with SRβ (Fig. 1D).[18,19] SRβ is exclusively found in eukaryotes and is permanently anchored in the membrane of the endoplasmic reticulum (ER) via an N-terminal TM helix. SRβ is itself a G protein and it has been shown that stable interaction with SRα is GTP dependent,[19] suggesting a regulatory mechanism that involves reversible attachment of SRα to the membrane.

The Protein Conducting Channel

Passage of the nascent chain through the cytoplasmic membrane of prokaryotes or the ER membrane of eukaryotes is mediated by the highly conserved PCC.[20] PCC core components are the transmembrane proteins SecY and SecE in the cytoplasmic membrane of bacteria, and the orthologs Sec61α and Sec61γ in the eukaryotic ER membrane. In addition, a nonessential small integral membrane protein is part of the channel, SecG in bacteria or Sec61β in eukaryotes, which appears not to be well conserved.[21] The recently solved crystal structure of the archaeal PCC of *Methanococcus jannaschii* reveals its fundamental architecture[22] (Fig. 1E). 10 TM helices of Sec61α arrange as two symmetrical halve rings made of helices 1-5 and 6-10 that together form a cylinder. The two helices of Sec61γ form a clamp that holds both Sec61α halves together, and the single TM helix of Sec61β is at the periphery of the assembly with weak contacts to Sec61α. The center of the structure has an hourglass-like shape with a central constriction made of highly conserved residues. It is proposed that the central constriction seals the closed pore through which the nascent polypeptide would pass in an open conformation.[22] The side of the channel not clamped by the γ-subunit (between TM helices 2 and 7) could possibly be pried apart to open the pore allowing membrane proteins to laterally exit the channel into the lipid bilayer. Although the crystal structure suggests that only one Sec61 heterotrimer is sufficient for protein translocation, oligomerization of the PCC has been observed in vitro by single-particle reconstructions using cryo-electron microscopy and also crystallographic studies on 2-D lattices.[23-26] With the crystal structure of the disengaged channel at hand, it is now possible to design structure-based experiments to address the remaining mechanistic questions.

The Control of Cotranslational Targeting

Cotranslational protein targeting depends in all organisms on the synchronized interplay of at least two GTPases, namely SRP54 and SRα, that regulate the process and ensure directionality. Eukaryotes in addition use a third G protein, SRβ. The superfamily of G proteins is the largest and most important class of regulatory proteins. Fundamentally, G proteins

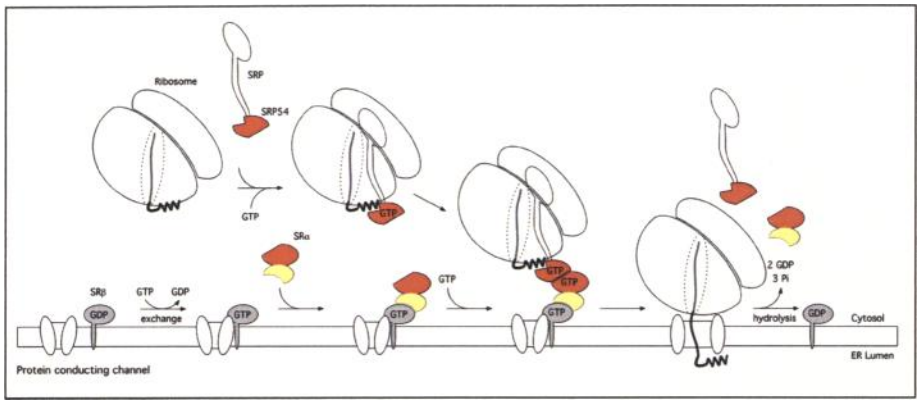


Figure 2. Model for the G protein controlled regulation of cotranslational protein transport in eukaryotes. SRP binds via SRP54 to the signal sequence of a nascent peptide chain, emerging from the ribosome exit site. Protein translation is concurrently arrested. The ribosome-nascent chain (RNC)-SRP complex is targeted to the ER membrane by SR. The NG domains of SRP54 and SRα have to be GTP-bound in order to form a stable complex. Further, SRβ has to be GTP-bound in order to stably interact with SRα through SRX. After or concomitantly with transfer of the nascent chain to the protein conducting channel in the ER membrane, SRP54, SRα and SRβ hydrolyze GTP and become available for a new targeting cycle.

cycle between GTP and GDP bound states, which result in specific conformational changes in the protein.[27] In the GTP bound state, G proteins bind to effector molecules to transmit a signal and are thus commonly referred to as 'switched on'. Hydrolysis of GTP is the 'off switch'. Frequently, accessory proteins are necessary to complete the cycle; GTPase activating proteins (GAPs) accelerate GTP hydrolysis whereas guanine nucleotide exchange factors (GEFs) reset the switch by facilitating the release of GDP and the rebinding of GTP. What do we know about the regulation of cotranslational targeting?

The initial step is the sampling of the emerging signal sequence at the exit tunnel of the ribosome by the M domain of SRP54[28,29] (Fig. 2). Conceptually, it is appealing that this interaction triggers a conformational change in the NG GTPase domain of SRP54, loading GTP as a result.[30] Despite a considerable repertoire of crystal structures of SRP54 this mechanism is not yet confirmed (see discussion in refs. 4,31). Concurrent to signal peptide binding, the Alu domain of SRP is positioned into the elongation factor binding site on the ribosome arresting translation.[10] GTP-binding of SRP54 enables interaction with the NG domain of SRα, which in turn binds GTP with increased affinity.[31] The interaction of SRP54 and SRα is exclusively mediated via a quasi-symmetrical interaction of the homologous NG domains. Crystal structures of this so-called engagement complex have been solved recently and reveal the fascinating atomic details of the GTPase mechanism.[32,33] In unprecedented fashion, both GTP molecules are part of the NG_{SRP54}-$NG_{SR\alpha}$ interface and form a composite nucleotide binding site.

After SRP54/SRα binding the SRP-RNC-SR complex is stalled. In prokaryotes, the complex now accumulates on the cytoplasmic membrane until a translocation-competent PCC is found.[34] Signal peptide transfer from SRP54 M domain to the PCC triggers mutual hydrolysis of GTP at the NG domains, disengaging SRP and SR and making them available for a new round of targeting.

In eukaryotes the situation is less clear since the role of the additional β-subunit of SR is yet only partially understood. With its N-terminal transmembrane helix SRβ is permanently anchored in the ER membrane. In order to stably interact with SRα, SRβ needs to be GTP bound. Consequently, GTP hydrolysis by SRβ should dissociate the complex. Regulated ER membrane recruitment of SRα might be an additional regulatory mechanism specific for

eukaryotes. One important step toward elucidating the potential regulatory role of SRβ is the identification of its GAP, since SRβ itself is catalytically inactive.[35] Noteworthy in any case is experimental evidence suggesting that possible SRβ-GTP hydrolysis does not occur prior to signal peptide transfer since in the absence of PCCs stalled SRP-RNC-SR complexes accumulate on the membrane just as is the case for prokaryotes.[36]

Evolutionary Considerations

Protein translocation across membranes is a basic necessity for all autonomous life forms. Besides the protein synthesizing ribosome the most important component of the translocation system is the PCC. Universal conservation of the essential PCC subunits Sec61α and -γ indicates that the channel evolved from a common ancestor. Coupling translation and translocation is a sophistication of the transport process that is not trivial. As outlined above, it requires a synchronized control mechanism. In fact, a simpler, post-translational transport pathway that converges on the PCC also exists. In this system the synthesized protein is kept in an unfolded, translocation-competent state by chaperones and delivered to the PCC. Post-translational transport is less efficient than cotranslational transport, since it requires additional energy in form of an ATP-driven motor that either pushes (SecA in *E.coli*) or pulls (BiP in *S.cerevisiae)* the chain through the PCC.[37] Cotranslational transport does not require energy in addition to that required for protein synthesis provided by the translation machinery.[38] Interestingly, post-translational transport is a frequently used pathway for secretory proteins in *E.coli* and *S.cerevisiae*, yet is only of marginal importance in higher eukaryotes. Unfortunately there is insufficient data for a classification of archaea in this respect,[39] but it is tempting to speculate that posttranslational evolved prior to cotranslational transport.

All cotranslational systems share the twin-GTPase mechanism involving the NG domains of SRP54 and SRα or their orthologs. Abundance of structural and genomic information as well as the relative ease with which G proteins can be detected, due to strongly conserved signature motifs, have revealed detailed insight into their phylogeny.[9] Based on structural and sequence conservation the NG domain constitutes a unique G protein family. Apart from its role in cotranslational transport the NG domain has not been detected in any other cellular context.

The functional implications of variations in the composition of SRP in different species are not straightforwardly explained. The absence of the Alu domain in *E. coli* SRP most likely explains why in this organism no translation arrest occurs during targeting. However, the apparent absence of individual architectural SRP proteins in certain parasitic eukaryotes and archaea could have several explanations. Interestingly, experimental evidence suggests that in *B. subtilis* the SRP9/14 heterodimer in the Alu domain is functional replaced by histone-like HBsu,[40] a protein with a substantially different structure. It is conceivable that replacement of SRP proteins by proteins with other functional annotations could extent beyond *B.subtilis* and SRP9/14. Another possibility is that the architectural SRP proteins do not have detectable sequence signatures and might therefore be very difficult to find by comparative genome analysis. This is not an uncommon phenomenon and has recently been observed many times in crystal structures that turn out to be well-known domain folds, despite having marginal sequence identity.[41,42]

Phylogenetic analysis of SR is less ambiguous. Apart from the universally conserved NG domain in SRα, a clear distinction can be drawn between the SR of prokaryotes and eukaryotes. Only eukaryotes contain the additional G protein SRβ. This statement can independently be tested by searching for SRαs that contain an SRX domain, which is necessary and sufficient for stable interaction with SRβ.[18,19] SRαs with an SRX domain are found exclusively in eukaryotes as well (unpublished data). SRβ belongs to the superfamily of small monomeric G proteins.[43] Members of this superfamily are widespread regulators of many cellular functions that characterize eukaryotic cells, i.e., vesicle trafficking, cytoskeleton remodeling and nuclear transport. Because of small G protein diversification especially in

eukaryotes, phylogenetic analysis of this protein class might provide clues on the evolutionary history of specific eukaryotic features.[44] SRβ is structurally and on the primary sequence level closely related to Arf/Sar-type G proteins, a family of vesicle transport regulators.[45] This is in close agreement with the phylogenetic analysis,[44] showing that Arf/Sar/SRβ branched off early from all other small G proteins. Despite their similarity, significant functional differences exist between Arf/Sar and SRβ. Arf/Sar are mechanistically characterized by reversible membrane attachment facilitated by an amphiphilic N-terminal helix that can be inserted into one leaflet of a lipid bilayer in nucleotide-controlled manner.[46] SRβ instead contains a bona fide TM helix at its N-terminus that permanently anchors the protein in the ER membrane. In addition, SRβ has a characteristic, highly conserved extension of helix 4 of the central G protein fold,[19] the function of which is enigmatic. These distinct features show that SRβ constitutes a separate G protein family.

Why did eukaryotes develop SRβ so early in their evolution? A plausible scenario is realized if one considers that the development of an endomembrane system may have been the earliest event during eukaryogenesis.[44] Before phagocytosis was developed, only small nutrients were transported through the membrane. Secretion of digestive enzymes was likely advantageous under such circumstances, but not necessarily when secreted into the free environment so that neighboring cells would profit equally. Recent evidence suggests that the common ancestor of Arf/Sar/SRβ might have been the protein that facilitated the initial engulfment of the cytoplasmic membrane. In a remarkable study it was shown that simple incubation of liposomes with Sar1-GTP led to deformation of the vesicles into long, tubular structures, strictly dependent on the presence of the exposed, N-terminal membrane-penetrating helix.[47] With GDP-bound Sar1, in which the N-terminal helix is retracted and folds into the protein core, no vesicle tubulation was observed. A mechanism to localize protein secretion to these engulfed membranes instead of the outer cytoplasmic membrane becomes necessary if the cell is to benefit from their presence. SRβ may have originated in eukaryotes as a means to target cotranslationally translocated proteins to these engulfed membranes through recruitment of the SRP-RNC complex. Eventually the engulfed membranes separated from the cytoplasmic membrane and formed the endomembrane system of modern eukaryotes. In the extant eukaryotes SRβ is functionally important to ensure targeting of SRP-RNC complexes to the ER, where PCCs are exclusively localized. In contrast to prokaryotes, eukaryotes modify proteins extensively through glycosylation and disulfide bond formation before they leave the cell, processes that only take place in the ER and the Golgi apparatus. Thus, the evolution of SRβ as a mechanism to target protein secretion to the endomembranes may have been a critical step for a defining process in eukaryogenesis.

References

1. von Heijne G. Signal sequences. The limits of variation. J Mol Biol 1985; 184:99-105.
2. Matlin KS. The strange case of the signal recognition particle. Nat Rev Mol Cell Biol 2002; 3:538-542.
3. Doudna JA, Batey RT. Structural insights into the signal recognition particle. Annu Rev Biochem 2004; 73:539-557.
4. Egea PF, Stroud RM, Walter P. Targeting proteins to membranes: Structure of the signal recognition particle. Curr Opin Struct Biol 2005; 15:213-220.
5. Halic M, Beckmann R. The signal recognition particle and its interactions during protein targeting. Curr Opin Struct Biol 2005; 15:116-125.
6. Keenan RJ, Freymann DM, Stroud RM et al. The signal recognition particle. Annu Rev Biochem 2001; 70:755-775.
7. Pool MR. Signal recognition particles in chloroplasts, bacteria, yeast and mammals (review). Mol Membr Biol 2005; 22:3-15.
8. Walter P, Blobel G. Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature 1982; 299:691-698.
9. Leipe DD, Wolf YI, Koonin EV et al. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 2002; 317:41-72.

10. Halic M, Becker T, Pool MR et al. Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature 2004; 427:808-814.
11. Van Nues RW, Brown JD. Saccharomyces SRP RNA secondary structures: A conserved S-domain and extended Alu-domain. RNA 2004; 10:75-89.
12. Zwieb C, Eichler J. Getting on target: The archaeal signal recognition particle. Archaea 2002; 1:27-34.
13. Powers T, Walter P. Cotranslational protein targeting catalyzed by the Escherichia coli signal recognition particle and its receptor. Embo J 1997; 16:4880-4886.
14. Gilmore R, Walter P, Blobel G. Protein translocation across the endoplasmic reticulum. II. Isolation and characterization of the signal recognition particle receptor. J Cell Biol 1982; 95:470-477.
15. de Leeuw E, Poland D, Mol O et al. Membrane association of FtsY, the E. coli SRP receptor. FEBS Lett 1997; 416:225-229.
16. Bibi E, Herskovits AA, Bochkareva ES et al. Putative integral membrane SRP receptors. Trends Biochem Sci 2001; 26:15-16.
17. Zelazny A, Seluanov A, Cooper A et al. The NG domain of the prokaryotic signal recognition particle receptor, FtsY, is fully functional when fused to an unrelated integral membrane polypeptide. Proc Natl Acad Sci USA 1997; 94:6025-6029.
18. Young JC, Ursini J, Legate KR et al. An amino-terminal domain containing hydrophobic and hydrophilic sequences binds the signal recognition particle receptor alpha subunit to the beta subunit on the endoplasmic reticulum membrane. J Biol Chem 1995; 270:15650-15657.
19. Schwartz T, Blobel G. Structural basis for the function of the beta subunit of the eukaryotic signal recognition particle receptor. Cell 2003; 112:793-803.
20. Simon SM, Blobel G. A protein-conducting channel in the endoplasmic reticulum. Cell 1991; 65:371-380.
21. Hartmann E, Sommer T, Prehn S et al. Evolutionary conservation of components of the protein translocation complex. Nature 1994; 367:654-657.
22. Van den Berg B, Clemons Jr WM, Collinson I et al. X-ray structure of a protein-conducting channel. Nature 2004; 427:36-44.
23. Beckmann R, Spahn CM, Eswar N et al. Architecture of the protein-conducting channel associated with the translating 80S ribosome. Cell 2001; 107:361-372.
24. Collinson I, Breyton C, Duong F et al. Projection structure and oligomeric properties of a bacterial core protein translocase. Embo J 2001; 20:2462-2471.
25. Menetret JF, Hegde RS, Heinrich SU et al. Architecture of the ribosome-channel complex derived from native membranes. J Mol Biol 2005; 348:445-457.
26. Morgan DG, Menetret JF, Neuhof A et al. Structure of the mammalian ribosome-channel complex at 17A resolution. J Mol Biol 2002; 324:871-886.
27. Vetter IR, Wittinghofer A. The guanine nucleotide-binding switch in three dimensions. Science 2001; 294:1299-1304.
28. Romisch K, Webb J, Lingelbach K et al. The 54-kD protein of signal recognition particle contains a methionine-rich RNA binding domain. J Cell Biol 1990; 111:1793-1802.
29. Zopf D, Bernstein HD, Johnson AE et al. The methionine-rich domain of the 54 kd protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence. Embo J 1990; 9:4511-4517.
30. Bacher G, Lutcke H, Jungnickel B et al. Regulation by the ribosome of the GTPase of the signal-recognition particle during protein targeting. Nature 1996; 381:248-251.
31. Powers T, Walter P. Reciprocal stimulation of GTP hydrolysis by two directly interacting GTPases. Science 1995; 269:1422-1424.
32. Egea PF, Shan SO, Napetschnig J et al. Substrate twinning activates the signal recognition particle and its receptor. Nature 2004; 427:215-221.
33. Focia PJ, Shepotinovskaya IV, Seidler JA et al. Heterodimeric GTPase core of the SRP targeting complex. Science 2004; 303:373-377.
34. Herskovits AA, Shimoni E, Minsky A et al. Accumulation of endoplasmic membranes and novel membrane-bound ribosome-signal recognition particle receptor complexes in Escherichia coli. J Cell Biol 2002; 159:403-410.
35. Legate KR, Andrews DW. The beta-subunit of the signal recognition particle receptor is a novel GTP-binding protein without intrinsic GTPase activity. J Biol Chem 2003; 278:27712-27720.
36. Song W, Raden D, Mandon EC et al. Role of Sec61alpha in the regulated transfer of the ribosome-nascent chain complex from the signal recognition particle to the translocation channel. Cell 2000; 100:333-343.

37. Stephenson K. Sec-dependent protein translocation across biological membranes: Evolutionary conservation of an essential protein transport pathway (review). Mol Membr Biol 2005; 22:17-28.
38. Gorlich D, Rapoport TA. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. Cell 1993; 75:615-630.
39. Bolhuis A. The archaeal Sec-dependent protein translocation pathway. Philos Trans R Soc Lond B Biol Sci 2004; 359:919-927.
40. Nakamura K, Yahagi S, Yamazaki T et al. Bacillus subtilis histone-like protein, HBsu, is an integral component of a SRP-like particle that can bind the Alu domain of small cytoplasmic RNA. J Biol Chem 1999; 274:13569-13576.
41. Caetano-Anolles G, Caetano-Anolles D. Universal sharing patterns in proteomes and evolution of protein fold architecture and life. J Mol Evol 2005; 60:484-498.
42. Chothia C, Gough J, Vogel C et al. Evolution of the protein repertoire. Science 2003; 300:1701-1703.
43. Sprang SR. G protein mechanisms: Insights from structural analysis. Annu Rev Biochem 1997; 66:639-678.
44. Jekely G. Small GTPases and the evolution of the eukaryotic cell. Bioessays 2003; 25:1129-1138.
45. Nie Z, Hirsch DS, Randazzo PA. Arf and its many interactors. Curr Opin Cell Biol 2003; 15:396-404.
46. Pasqualato S, Renault L, Cherfils J. Arf, Arl, Arp and Sar proteins: A family of GTP-binding proteins with a structural device for 'front-back' communication. EMBO Rep 2002; 3:1035-1041.
47. Lee MC, Orci L, Hamamoto S et al. Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. Cell 2005; 122:605-617.

CHAPTER 5

Evolution of the Endoplasmic Reticulum and the Golgi Complex

Alexander A. Mironov,* Victor V. Banin, Irina S. Sesorova, Viacheslav V. Dolgikh, Alberto Luini and Galina V. Beznoussenko

Abstract

By analyzing the morpho-physiological features of the Golgi complex, its relationship with the endoplasmic reticulum in different species, and the molecular machineries involved in intracellular transport, we conclude that; (1) all eukaryotic cells have either Golgi complexes or remnants thereof; (2) all eukaryotic cells have a large minimal set of proteins that are involved in intracellular transport; and (3) several indispensable molecular machines are always present in secreting eukaryotic cells. Using this information, our data about mechanisms of intra-Golgi transport and phylogenetic analysis of several molecular machines, we propose a model for the evolution of the Golgi complex and the endoplasmic reticulum.

Introduction

Understanding how eukaryotic cells transport newly synthesized proteins from the endoplasmic reticulum (ER) to their specific cellular destinations has been a central goal in cell biology since the beginning of this discipline several decades ago. Progress throughout this period has been uneven, however, and the traffic field has gone through various different stages and shifts of focus. These notable changes in the way we view intracellular traffic have been driven by work from several laboratories over the last few years, and they have been brought about by a reexamination of the morpho-functional organization of the transport pathways in vivo. The key feature of the most informative studies is the use of new morphological techniques that have been combined with a more precise synchronization of intra-Golgi transport.[1-4]

The Golgi complex is the central station along the secretory pathway. It receives newly synthesized proteins and lipids from the ER, and then distributes them to the plasma membrane and to the endosomal/lysosomal system. In polarized cells (e.g., epithelial cells and neurons), it sends proteins and lipids to the correct surface domains. The Golgi complex also serves as a "factory" for the posttranslational modification of proteins and lipids (mostly by glycosylation).

The Golgi complex (or its derivates) and the ER are present in all eukaryotic cells. The Golgi is indeed extremely complex, and varies significantly in cells from different species (see Table 1). Our current knowledge relating to the structure and function of the Golgi complex has been described in detail recently[5] along with an excellent introduction to the secretory pathway[6] and our own exhaustive overview of how we see the Golgi complex functioning.[7]

*Corresponding Author: Mironov Alexander—Department of Cell Biology and Oncology, Mario Negri Sud Institute, Via Nazionale 8, 66030 Santa Maria Imbaro (Chieti), Italy. Email: mironov@negrisud.it, http://www.negrisud.it/en/dcbo/Mironov/

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

Table 1. Features of the Golgi in cells from different species (based on refs. 32,33,67)

Feature	Presence of the feature								
	Mammals	Insects	Plants	Protozoa		Yeast		Others	
				Simple	Complex	S.c	P.p	M-sp.	Worms
Segregation from the ER	+	+	+	+	+	+	+	+	+
Network of smooth and varicose tubules	+	+	+	+	+	+	+	+	+
Presence of two compartments	+	+	+	+	+	+	+	+	+
Periodical continuity with the ER	+	+	+	+	+	+	+	+	+
Periodical continuity with the post-Golgi	+	+	+	+	+	+	+	+	+
H^+-ATP pump	+	+	+	+	+	+	+	+	+
Golgi glycosidases	+	+	+	+	+	+	+	+	+
Nucleotide transporters	+	+	+	+	+	+	+	-?	+
Matrix proteins	+	+	+	+	?	?	+	?	?
Movement by actin (fragmented Golgi)	±	±	+	+	±	±	±	–	±
Movement by microtubules	+ Central Golgi	±	–	–	–	–	–	–	±
Stacked disk-like cisternae	+	+	+	+	–	±	+	–	+
Sar1/COPII	+	+	+	+	+	+	+	Reduced	+
ARF/COPI	+	+	+	+	+	+	+	Reduced	+
COPI vesicles	+	+	?	?	?	+	+	–	?
AP/clathrin	+	+	+	+	+	+	+	No clathrin	+
Clathrin vesicles	+	+	+	+	?	?	+	–	?
SNAREs	+	+	+	+	+	+	+	Reduced	+

S.c. = *S. cereviseae*; P.p. = *P. pastoris*; M-sp. = microsporidia

We have demonstrated previously that large (at least bigger than coat protein, COPI-dependent vesicles) macromolecular aggregates of procollagen-I are transported through the Golgi complex without the participation of COPI vesicles.[8] Post-Golgi carriers are also not just small coat-dependent vesicles, but are instead much larger structures of irregular shape.[9] Comparing the transport of a small protein, the G protein of the vesicular stomatitis virus (VSVG), with that of the much larger aggregates of procollagen-I in the same cells, we demonstrated that VSVG and these aggregates move synchronously through the Golgi complex at rapid and indistinguishable rates. Both VSVG and these aggregates traverse a Golgi stack without leaving the cisternal lumen and without entering COPI vesicles in functionally relevant amounts.[2]

Furthermore, we have demonstrated that ER-to-Golgi carriers arise through cargo concentration and direct en bloc protrusion of specialized ER domains near COPII-coated exit sites, in a COPII-dependent manner and without the involvement of fusion of COPII-dependent vesicles. Fully protruded saccules then move centripetally.[3] Several independent lines of evidence have

also suggested that the Golgi enzymes are depleted in COPI-coated buds and in 50-60-nm COPI-dependent vesicles,[1] and we have recently shown that the arrival of cargo at the Golgi triggers the formation of tubules connecting the Golgi cisternae.[4] These connections are permeable to the Golgi enzymes, but not to procollagen-I,[4] as procollagen-I is obviously too big to fit inside such small tubules; VSVG tends to be mostly excluded because it forms aggregates in the plane of the cisterna[2,10] that will probably strongly limit its diffusion ability. Indeed, we have seen a rather strong exclusion of VSVG from tubular-elongated profiles surrounding the Golgi, as compared to the adjacent cisternae.[4]

As indicated, several different groups have demonstrated the depletion of the Golgi enzymes in peri-Golgi round profiles,[1,4,11-13] with only one report finding that they are not depleted in such peri-Golgi, COPI-coated, round profiles.[14] However, most of the 50-60 nm vesicles in the Golgi area are not coated.[15] A recent study by Cosson et al also clearly showed that Golgi enzymes are excluded from cisternal rims,[12] and we have confirmed the depletion of galactosyl transferase with nano-gold technology and antibodies against its cytosolic domain.[1]

Next, the in vitro isolated vesicle-like membranes that contained Golgi enzymes[16] are not able to fuse with Golgi cisternae.[17] Moreover, although this fraction is enriched in the Golgi enzymes, this appears to be because it also contains many small fragments of Golgi cisternae, such that the actual 50-60-nm vesicles in this fraction are indeed depleted of Golgi enzymes.[1] Finally, there is a recent report that a small (10% input) fraction of 60-nm vesicles enriched in CASP, giantin, mannosidase (Man)I and ManII can be seen after an incubation of isolated Golgi membranes with purified coatomer, recombinant myristoylated ARF1, GTP and an ATP regeneration system.[18] However, in the absence of ARF-GAP, this assay system is similar to in vitro systems based on the addition of GTPγS, and thus an artificial vesiculation of the perforated zones of the Golgi cisternae could have been induced.[19] Moreover, the same intra-Golgi transport rate is obtained in vitro in the presence of COPI- or ARF-depleted cytosol[20] indicating that COPI vesicles are not necessary for transport at least in cell free assay.

There are several additional unsolved questions that render less favorable the classical cisterna maturation/progression model of intra-Golgi transport that is based on the assumption that Golgi enzymes recycle by COPI vesicles: 1. All of the cisternae should have COPI-coated buds for Golgi enzyme recycling; however, there are only clathrin-coated buds on the last trans cisterna, where Golgi enzymes such as sialyltransferase and fucosyltransferase have been detected.[21] In this case, we should assume that clathrin-coated vesicles mediate the retrograde movement of Golgi enzymes; this is apparently not the case. 2. The role of COPI in intra-Golgi transport in vitro is also in doubt, since the depletion of COPI does not affect the rate of intra-Golgi transport.[20] 3. According to the classical cisterna maturation model,[22] the observed steady-state localization of the enzymes in the Golgi complex is dependent on COPI-vesicle-mediated retrograde transport. Recent studies have raised doubts regarding the role of COPI in this process. The presence of an activated, GTP-bound form of ARF, ARF-Q71L, did not affect the polarization of the Golgi stack after the washout of brefeldin A, an inhibitor of ARF activation.[23]

At present, most of the data available argue against a role for COPI and COPII vesicles as transport carriers, although COPI and COPII are extremely important for the transport of most of the cargoes. Thus, one of the main problems of the models of intra-Golgi transport based on COPI-vesicle-mediated transport of cargo or Golgi enzymes is the depletion of both cargo and Golgi enzymes in COPI-dependent vesicles. If COPI vesicles have no transport function, one still has to explain their abundance and the observations showing a block of transport when the functions of ARF/COPI are blocked.[7,24] The COPI machinery is essential for secretory traffic and COPI vesicles should have a significant role in intra-Golgi transport. The vesicles are abundant in ribbon stacks, where they represent up to 20% of the surface area of the Golgi,[21] although COPI vesicles are scarce in cells deprived of microtubules.[4] Two counteracting machineries, those of ARF/COPI and the SNAREs, contribute to the equilibrium between the formation and consumption of COPI-dependent vesicles: when the ARF/COP

machinery is blocked, the number of 52-nm vesicles decreases; if the SNARE machinery is inhibited, the number of COPI vesicles increases. Finally, when both machineries are suppressed, the shape of the Golgi becomes particularly stable.[1]

There is accumulating evidence that temporal inter-compartment membranous connections have a significant role in intracellular transport.[7,22] The cargo-dependent formation of intercisternal connections has been described recently, with these membrane connections being seen between the ER and the *cis*-Golgi at steady-state,[25] between different Golgi cisternae,[4,26] between the most trans cisterna and the TGN,[27] and between different domains of endosomes[28] after a 15 °C temperature block. These connections have an important role especially in the transport of soluble cargo proteins that undergo significant concentration within the Golgi complex (Trucco et al unpublished observations).

Given that connection-based models of intra-Golgi transport do not require the direct participation of COPI vesicles, the questions regarding their functional role remain open. One possible functional role would relate to their ability to concentrate certain SNARE molecules, and in this way to control fusion events occurring between different Golgi cisternae. Indeed, GOS28[29] and membrin[4,29] are enriched within round profiles in the Golgi, as compared to the cisternae. This suggests that COPI vesicles could regulate the ability of Golgi cisternae to fuse with each other.[30]

Schemes describing mechanisms of intracellular transport have significantly affected the modern theories of Golgi complex evolution. For instance, the phagotrophy theory of the origin of eukaryotes[31] clearly reflects the necessity for coat-dependent small vesicular carriers. On the other hand, no reasonable model of Golgi complex evolution can be formulated without taking into consideration our knowledge of the mechanisms of intra-Golgi transport.

The goal of this review is not to describe the peculiarities of the Golgi complex in different species. Instead, we use comparative analysis to understand the manner of Golgi complex and ER evolution in eukaryotes, and to build an evolutionary scenario that fits all of the constraints present in evolutionary trees and in current views of transport mechanisms. On one hand, we will use this knowledge to propose a model of Golgi complex evolution. On the other hand, our evolutionary model will also help us to check and correct models of intracellular transport.

Common and Peculiar Features of the Golgi Complex in Different Eukaryotic Cells

The appearance of Golgi stacks and the number of cisternae within a stack show considerable variation in different eukaryotes, with the simplest forms of the Golgi being a tubular network.[32,33] Even in microsporidia, which are obligate intracellular parasites that possess the smallest (only 2-3 MB) genome among eukaryotes,[34,35] the remnants of a Golgi complex can be seen. In the early stages of microsporidia development, the Golgi appears as tubular networks without any buds and coat-dependent vesicles (our unpublished observations), with cisternae appearing only in the late stage of their development.[36] The Golgi of the microsporidia *Paranosema grylli* appears as 300-nm networks of thin (25-40-nm diameter), branching and varicose tubules that display histochemical features of a Golgi, but that do not have buds and vesicles. Vesicles are not formed even if membrane fusion is inhibited. These tubular networks are connected to the ER, the plasma membrane and the forming polar tube. The sporewall and polar-tube proteins are transported from the ER to the target membranes through these tubular networks, within which they undergo concentration and glycosylation.[37]

One can divide the features of the Golgi into common and species-specific ones (Table 1). The common features include the segregation of the Golgi from the ER and from endosomes, the presence of two types of tubular networks, the first composed of smooth tubules and the second of varicose tubules.[27,38,39] Other common features include the presence of transient continuities with either the ER or the post-Golgi,[22] and the presence of the COPI and COPII machineries, glycosidases and matrix proteins (Table 1).

However, some features of the Golgi differ in different cells. The stacked, disk-like cisternae are absent in the simplest yeast, in microsporidia and in some other intracellular parasites. COPI-dependent vesicles have not been seen in microsporidia, but they are present in mammal, plant and insect cells, although it is still unclear whether these vesicles are present in yeast and other minimal intracellular parasites. The manner of Golgi movement is different in different species. For instance, in mammalian cells, worms and some insect cells, the Golgi undergoes centralization by microtubule-dependent dynein, whereas in yeast, plants and some insect cells, movement of the Golgi occurs with the help of an actin/myosin motor system. In microsporidia and other intracellular parasites there is no evidence of motor-dependent Golgi movement (Table 1). In contrast, the ER is similar in different species and in most cases; it is composed of rough (with attached ribosomes) and smooth ER.[40,41] Thus, the most simplified cells usually have a simpler Golgi, and in cells that are more complicated, the Golgi is more diversified.

The Minimal Set of Genes Involved in Intracellular Traffic

The reconstruction of the ancestral nature of intracellular trafficking depends on the topology and rooting of the eukaryotic tree.[42] Besides reconstructing ancestral states, another promising way to gain insights into the evolution of intracellular transport is to use minimal cellular systems. Comparative analyses with the genome of microsporidia, the minimal known eukaryote, could thus help us to determine what might be the minimal set of genes that are sufficient for intracellular traffic. Similarly, they could be very useful to show what compartments are essential for eukaryotic endo-membranes. However, we have to keep in mind that microsporidia do not represent the primitive condition, the ancestral state of eukaryotic ER and Golgi.

As indicated, microsporidia are the smallest eukaryotic intracellular parasites with a physical size of 2-3 μm in diameter, (depending on species) and a genome size of 2-3 Mb, on average. According to data from the genome project of the human microsporidian *Encephalitozoon cuniculi*[34] these cells possess, just as all other eukaryotes, all of the most important and well-characterized protein machineries that are involved in cotranslational translocation of polypeptide chains into the ER lumen, as well as in intracellular transport, although some of these machineries do apparently lack nonessential components. The presence of these (sometimes reduced) protein machines shows that in microsporidia the process of protein secretion should be similar to that in mammalian cells.

Analysis of the minimal genome of microsporidia has revealed that to perform intracellular transport, any eukaryotic cell has to possess a considerable number of genes that appear essential for intracellular transport.[34,43,44] Recently, we have demonstrated that microsporidia possess intracellular transport.[37] Microsporidia lack mitochondria and peroxisomes, and were first considered to be a deeply branching protist lineage that diverged before the endosymbiotic event that led to mitochondria. The discovery of a gene for a mitochondrial-type chaperone combined with molecular phylogenetic data later implied that microsporidia are atypical fungi that have actually lost their mitochondria during evolution.[34]

This most simplified representative of eukaryotes has a very limited number of proteins that are involved in intracellular transport,[34] which include those responsible for the incorporation of polypeptide chains into the ER lumen (two subunits of Sec61) and seven enzymes involved in glycosylation. Among these, five are glycosyltransferases. Among their six SNAREs, there are two R-SNAREs (SNC2 and synaptobrevin) and four Q-SNAREs (syntaxin 5, VAMP, Bos1 [2 different copies] and Vti1). It is also important to stress that these SNAREs in microsporidia are very similar to those in yeast. In turn, the SNAREs in yeast show a high degree of sequence conservation to the mammalian SNAREs, and they can be replaced by mammalian SNAREs without the generation of functional disorders. For example the human homolog can functionally replace the yeast v-SNARE Vti1p.[45]

The machinery responsible for the dismantling of SNARE complexes[46] has only one protein in microsporidia, Sec18 (homologous to mammalian NSF), although there are no proteins that are homologous to SNAPs.[34] This means that the SNARE machinery in microsporidia

might work slower than that in mammalian cells. The minimal set of Rab proteins (Ypt1, Rab1b, Rab5, Ytp6 and Rab10) and the Rab-GDP dissociation inhibitor are present.

Instead of the seven subunits of COPI that are typical for mammalian and plant cells, microsporidia have only six, with the missing one known to be ε-COPI. In mammalian cells, ε-COPI is important for the generation of COPI vesicles.[34] In cells containing a temperatureresensitive ε-COPI subunit (ldl F cells); after 6 h of incubation at the restrictive temperature there are no vesicles and the uncoating of COPI is inhibited.[47] The lack of ε-COPI from microsporidia and the requirement for this subunit to form COPI vesicles in mammalian cells is in agreement with the lack of Golgi vesicles from microsporidia. The ARF machinery in microsporidia is also limited with only two ARFs and one exchange factor for ARF. Another important gene missing, besides ε-COPI, is ARF-GAP. However, the genome does contain an ARF-like protein that can partially replace ARF-GAP.[48]

On the other hand, among four known subunits of the COPII machinery, only three, Sec13, Sec23 and Sec31, are found in microsporidia.[34] Although Sar1p is present, Sec12 that operates as a Sar1 exchange factor is absent. Finally, microsporidia lack clathrin. The reason for the absence of clathrin might be that microsporidia lack lysosomes (no lysosomal enzymes and no machinery responsible for such enzyme phosphorylation, and no clathrin have been found; however, H+-ATPases are present)[34] and endocytosis, and therefore they do not need to transport lysosomal enzymes or internalize proteins from the PM.

These findings, together with the limited effect of COPI mutations on transport and survival in yeast[49] suggest that the ancestral vesicle coating complex from which COPI, COPII and clathrin coats evolved was initially not able to generate coated vesicles, although its function as a protein complex that is able to bind the lipid bilayer through a small GTPase and participate in protein sorting could have been important for the proto-cell. For instance, the primary function for this proto-coatomer complex might have been the regulation of the lateral diffusion of proteins and their segregation along the lipid bilayer with the ability to bud off vesicles only evolving later.

Models of Evolution of the ER and the Golgi Complex

The Formation of the Nucleus and Eukaryotic Endo-Membranes

In most prokaryotes membrane-associated processes like the biosynthesis of membrane proteins, membrane lipids, and the attachment of the single chromosome to a membrane are associated with the plasma membrane. In contrast, in eukaryotes, these processes are associated with the nuclear/ER membrane system, while the plasma membrane serves other functions (e.g., exocytosis and endocytosis). Any hypothesis explaining the evolution of eukaryotes from a prokaryotic ancestor has to explain the origins of this segregation, i.e., the evolutionary origins of the endo-membrane system and the nuclear envelope. The development of the nuclear envelope of the proto-eukaryotic cell can be envisaged as an invagination of the plasma membrane that isolated the nuclear substance into the nuclear envelope by surrounding it with two lipid bilayers with a space between them. Indeed, it has also been hypothesized that the ER also originated as an infolding of the plasma membrane, which later became separated.[50,51] Removal of certain functions from the plasma membrane during evolution could have facilitated the adaptation of this membrane for other processes (e.g., exocytosis and endocytosis). For reasons of membrane topology, we can exclude such mechanisms for the formation of the nuclear envelope as the penetration of a proto-prokaryotic cell (or its spore) into another proto-prokaryotic cell (chronocyte, etc[52] see the chapter by G. Jekely).

The Appearance of the Golgi and the Small GTPases

The issue of Golgi development is a matter of debate. According to one model the formation and fusion of transport carrier vesicles was the main mechanism of Golgi evolution.[31] Our model differs significantly from this scheme. We propose that the development of tubular

invaginations and the evolution of a tubular network was the primary means of Golgi origin and vesiculation only evolved secondarily. The basis for our model is our knowledge about the function of the Golgi complex and a recent phylogenetic analysis of the Ras super family that led to the hypothesis that the emergence of eukaryotic membranes predated that of phagocytosis.[53,54] This was based on the early splitting off of secretory GTPases, Sar1, ARF and SRβ, from all other eukaryotic small GTPases. Members of the Ras, Rab and Rho families are requiring for phagocytosis, and thus phagocytosis may have appeared later than secretion.[54]

Sar1, a small GTPase, seems to be the most ancestral member of the small GTPase family. The attachment of Sar1 to a membrane induces the tubulation of the membrane due to an augmentation of membrane curvature.[55] This small GTPase cenancestor could have been responsible for the generation of tubular invaginations from specific domains of the plasma membrane, having a similar activity than that of modern Sar1.

The formation of plasma membrane invaginations could have happened by a process related to the recently established activity of Sar1. A Sar1-like activity could have been responsible for the production of primary internal curvature. These invaginations could sort the Sec translocation machinery into this domain, and the site of chromosome attachment could be segregated into these internal protrusions of the plasma membrane. Sar1 can also induce invaginations based on its interaction with microtubules, because it recruits kinesin to ER membranes.[56] ARF/COPI could have had a role in the constriction of tubules connecting different compartments.

Membrane tubulation was followed by the development of a proto-coatomer to provide a mechanism for membrane protein concentration and sorting. In addition, the proto-coatomer could have been responsible for the generation of ER processes from the original nuclear envelope by the introduction of membrane curvature.

We propose that continuity between the lumen of the nuclear envelope/ER and the environment could have been a disadvantage. This lead to the development of mechanisms to prevent the formation of a direct lumenal continuity between the nuclear envelope and the environment. This could have happened e.g., by the generation of immediate fission proximal to the fusion, as has been proposed in the framework of the carrier maturation model.[7] Thus, the next stage could be the isolation of the lumen of the ER and the Golgi complex from the environment. This was achieved by the development of the machineries that are able to constrict tubules connecting the plasma membrane and the ER via the Golgi complex. One of the possible candidates for such a machinery could be COPI and maybe TRAPP. In fact, analysis of the minimal protein machineries in microsporidia revealed that COPI is indispensable, and although reduced by elimination of ε-COP, it is still present in the genome of the minimal eukaryotes like microsporidia. One could envisage that COPI can form a rigid spherical cage, and that the assembly of these cages along membrane tubules could induce the formation of constriction sites and varicosities along the tubules. The ability of ARF/COPI to regulate the width of tubules connecting different organelles has made it possible to use physicochemical mechanisms for the concentration of cargo proteins within the Golgi complex.

Although we cannot deny the importance of COPI and COPII vesicles, their origin served, we would suggest, not for the production of transport carriers; these vesicles were rather the result of fission of varicose tubules connecting the proto-plasma membrane and the proto-ER. Thus, the main original function of the coat could have been to induce fission and then to pinch off membranous proto-tubules from the plasma membrane, thus dividing the ER and the plasma membrane.

It has been suggested that all membrane coats have a common ancestral origin, and reconstructions argue that COPI and COPII were present in the last common ancestor of eukaryotes.[31] Indeed, most of the components of COPI, COPII and clathrin/AP may share the same structural domains.[57] Phylogenetic studies have demonstrated an evolutionary link between the components of COPI and the adaptor protein complexes (APs) 1, 2 and 3.58.

However, the clathrin/AP coat is not based on any small GTPases. This indicates that this coat emerged later than other coats. As such, one would conclude that COPI is more ancient than clathrin, which is not vitally important in eukaryotes. At least, microsporidia have no clathrin genes at all. The secondary reduction of the secretory pathway in microsporidia[34] suggests that endocytosis and COPI and COPII vesicles were not essential for intracellular transport in primitive eukaryotic cells, whereas COPI, COPII and the SNARE machineries per se were necessary. The example of microsporidia also shows that clathrin is not essential for the minimal transport machinery, and thus it originated later than COPI and COPII.

Similarly, deletion of clathrin is not lethal in yeast and mammalian cells.[59,60] If we compare the development of the endocytic system in yeast and in some intracellular parasites with that in mammalian cells, it seems that most of the development in evolutionary terms occurred within the endocytic pathway. In contrast, the protein machineries involved in ER-to-Golgi transport are one of the most conserved systems. The elaboration of the endocytic system could have been linked to the formation of multicellular organisms (see below). As such, phagocytosis appeared later than intracellular secretion. Similar suggestions relating to the limited necessity for clathrin for transport can be derived following clathrin deletion.[60]

The next stage in the development of an endo-membrane system was the segregation of the Golgi from the network of membrane tubules. The selective advantage of its formation could have been to ensure that all of the secretory and membrane proteins are glycosylated by their obligatory passage through the Golgi. The generation of a compartment, situated along the membrane continuity between the ER and the plasma membrane, could have been the solution. This new compartment segregated most of glycosidases from the ER, and during evolution, these enzymes were selected from mostly type II proteins. Our comparative analysis of the structure of the modern Golgi complex (Table 1) indicates that membrane tubulation is probably an ancestral and essential (i.e., it cannot be lost even in parasites) mechanism and it was the first step in endo-membrane origin.

On the other hand, transformation of the plasma membrane into thick sterol-rich membranes could lead to the necessity to improve the uptake of substances from the outside and the emergence of phagocytosis, and then endocytosis.

The Evolution of the SNAREs

The next stage in endo-membrane development could have been the evolving of the SNARE machineries that are responsible for the restoration of membrane continuity broken by the activity of COPI or other constrictors. Recently, it has been shown that various syntaxin homologs are present in a wide range of eukaryotic lineages, prompting speculation of the presence of a primitive syntaxin in an early eukaryotic ancestor.[61] These findings suggest that the SNARE machinery involved in the acceleration of membrane fusion appeared in the early eukaryotic ancestor.

Examination of the microsporidian genome suggests that the SNAREs are indispensable for protein traffic since at least five different SNAREs are present and two of these are R-SNAREs.[34] Thus, in microsporidia there should be at least two fusion events. Indeed, in our study, we have revealed two states of the tubular network of Golgi remnants: one connecting with the ER, and the other with the plasma membrane.[37] The SNARE proteins were probably formed from the coiled-coil proteins remaining between varicosities, with the most plausible candidate being actin.

SNAREs probably evolved later than the ER/nuclear envelope and the Golgi complex. When the nucleus became massive and separated from the cytosol, its movement could lead to the breakdown of tubules, especially after their varicose transformation by ARF/COPI. However, the necessity for transport remained, and this was solved by the evolution of the SNARE machinery. To regulate SNAREs, the cell needed the Rab proteins that prevented the nonspecific reassembly of the SNARE complexes in the plane of the same membrane. The SNAREs become enriched in sites of fusion through the action of a coat. The COPI and COPII machineries can concentrate the SNAREs by binding to them.[62]

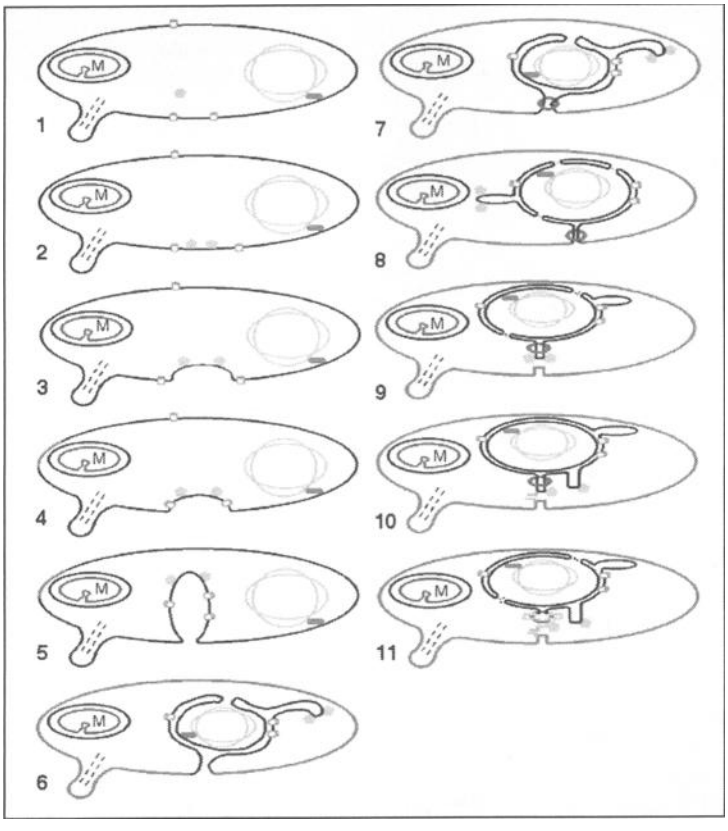


Figure 1. A scheme of the origins of the Golgi complex. The sequence of the events can be envisaged as the following: 1) Appearance of the small GTPase Sar1 (greenish pentagon in the cytoplasm), which is able to generate membrane tubules from the plasma membrane due to the insertion of curvature into the lipid bilayer. Mitochondria are indicated by M. 2) Attachment of Sar1 to the plasma membrane. 3) Generation of invaginations from the plasma membrane due to the activity of Sar1. The Sar1 attached to specific domains of the plasma membrane induces the formation of invaginations, seen as tubules protruding into the cytosol. 4) Generation of sterols by mitochondria and formation within the plasma membrane of domains with thicker membranes containing sterols (represented by a red line). 5) Activity of Sar1 lead to the sequestration of thin membrane domains. These domains accumulate the Sec translocation machineries and sites with chromosome binding. 6) Shift of DNA binding sites to inside the membrane protrusions (pro-nucleus) formed by the invaginations. Sequestration of the Sec translocation machinery to the outer sides of these membranous caps (invaginations connected with the plasma membrane). 7) Formation of ARF/COPI machinery (blue ring) for the constriction of membranous tubules connecting the pro-nucleus and the plasma membrane. 8) Formation of the ER from the nuclear envelope, separation of the nucleoplasm from the cytosol, and formation of nuclear pores based on the appearance of Ran. This induces the development of mitosis. 9) Separation of the lumen of the ER from the outside by constriction of the tubules connecting the nuclear envelope and the plasma membrane. This separation is achieved based on ARF/COPI: the activity of ARF/COPI leads to the formation of varicosities along membranous tubules and the sequestration of the Golgi enzymes nearby. 10) Breakdown of the varicose tubules induces the necessity for the generation of a fusion machinery for the transport of secretory proteins to the outside. Development of SNARE (light greenish bent lines) for the restoration of membrane continuity during the transport of carriers. This leads to the formation of the Rabs for the regulation of SNAREdependent membrane fusion. 11) Development of the Golgi (indicated by light blue lines) at the sites where the fusion machinery is concentrated by Sar1/COPII and ARF/COPI. A color version of this figure is available online at http://www.Eurekah.com.

The Diversification of the Golgi Complex

The Golgi enzymes show very low similarity to each other, and thus most likely emerged independently.[63] Once a Golgi-like compartment was initially formed, the various protist groups and their descendants optimized this basic system for their own requirements. In examining the evolution of the Golgi that followed, we should analyze trees of different proteins before coming to conclusions about the appearance of different features, as some of these features could be secondary.

The ancestrally tubular Golgi went through a transformation into stacked Golgi, as can be seen for example in *Pichia pastoris*. The stacked Golgi diverged into forms where its motility is based on the actin/myosin system, as in plants and some insects, or on microtubular motors, as in mammalian cells.[64]

During the evolution of multicellularity, cells with a uniform plasma membrane formed multicellular aggregates and polarized their membrane structures. The cells were divided into ectoderm and endoderm. The endoderm cells, inside the organism, acquired a basolateral plasma membrane. The apical membrane is probably older, although its appearance can be secondary in some cells. Once cellular aggregates were formed, the primitive endosomal membrane, which may have retained some of its plasma membrane properties, could have fused to the plasma membrane at the site of aggregation, forming a specialized plasma membrane domain (maybe derived from proto-endosomes) and a protected intercellular space. In yeast, Golgi-derived secretory vesicles fuse with the plasma membrane as separate entities.[65] Similarly, in animal cells, the secretory granules fuse with the apical plasma membrane as separate membranes.[66] This means that the plasma membrane of eukaryote ancestors was more similar to the yeast/plant plasma membrane and to the apical membrane of polarized animal cells, than to the basolateral membrane of polarized animal cells.

Conclusions

To summarize, we have compared the morpho-physiological features of the Golgi complex and the ER in different species, and the molecular machineries involved in intracellular transport. Our analysis allows us to conclude that all eukaryotic cells have either Golgi complexes or their remnants. All eukaryotic cells also have at least a minimal set of proteins that are involved in intracellular transport, and this set is indeed particularly large. Several indispensable molecular machineries are always present in secreting eukaryotic cells. Using this information, we have proposed a model of the evolution of the Golgi and the ER/nuclear envelope, where we have mostly analyzed the very early stages of Golgi complex evolution; this scheme of Golgi evolution is presented in Figure 1. The later steps in the evolution of the Golgi complex have to be further studied by analyzing the evolutionary history of different proteins involved in secretion.

References

1. Kweon HS, Beznoussenko GV, Micaroni M et al. Golgi enzymes are enriched in perforated zones of golgi cisternae but are depleted in COPI vesicles. Mol Biol Cell 2004; 15:4710-4724.
2. Mironov AA, Beznoussenko GV, Nicoziani P et al. Small cargo proteins and large aggregates can traverse the Golgi by a common mechanism without leaving the lumen of cisternae. J Cell Biol 2001; 155:1225-1238.
3. Mironov AA, Mironov Jr AA, Beznoussenko GV et al. ER-to-Golgi carriers arise through direct en bloc protrusion and multistage maturation of specialized ER exit domains. Dev Cell 2003; 5:583-594.
4. Trucco A, Polishchuk RS, Martella O et al. Secretory traffic triggers the formation of tubular continuities across Golgi sub-compartments. Nat Cell Biol 2004; 6:1071-1081.
5. Polishchuk RS, Mironov AA. Structural aspects of Golgi function. Cell Mol Life Sci 2004; 61:146-158.
6. Gurkan C, Koulov AV, Balch WE. An evolutionary perspective on eukaryotic membrane trafficking. This volume.
7. Mironov AA, Beznoussenko GV, Polishchuk RS et al. Intra-Golgi transport. A way to a new paradigm? BBA – Molecular Cell Research 1744:340-350.
8. Bonfanti L, Mironov Jr A, Martella O et al. Procollagen traverses the Golgi stack without leaving the lumen of cisternae: Evidence for cisternal maturation. Cell 1998; 95:993-1003.

9. Polishchuk RS, Polishchuk EV, Marra P et al. Correlative light-electron microscopy reveals the saccular-tubular ultrastructure of carriers operating between Golgi apparatus and plasma membrane. J Cell Biol 2000; 148:45-58.
10. Orci L, Glick BS, Rothman JE. A new type of coated vesicular carrier that appears not to contain clathrin: Its possible role in protein transport within the Golgi stack. Cell 1986; 46:171-184.
11. Cosson P, Amherdt M, Rothman JE et al. A resident Golgi protein is excluded from peri-Golgi vesicles in NRK cells. Proc Natl Acad Sci USA 2002; 99(20):12831-4.
12. Cosson P, Ravazzola M, Varlamov O et al. Dynamic transport of SNARE proteins in the Golgi apparatus. Proc Natl Acad Sci USA 2005; 102(41):14647-52.
13. Orci L, Amherdt M, Ravazzola M et al. Exclusion of golgi residents from transport vesicles budding from Golgi cisternae in intact cells. J Cell Biol 2000; 150(6):1263-70.
14. Martinez-Menarguez JA, Prekeris R, Oorschot VM et al. Peri-Golgi vesicles contain retrograde but not anterograde proteins consistent with the cisternal progression model of intra-Golgi transport. J Cell Biol 2001; 155(7):1213-24.
15. Marsh BJ, Mastronarde DN, Buttle KF et al. Organellar relationships in the Golgi region of the pancreatic beta cell line, HIT-T15, visualized by high resolution electron tomography. Proc Natl Acad Sci USA 2001; 98(5):2399-2406.
16. Lanoix J, Ouwendijk J, Lin CC et al. GTP hydrolysis by arf-1 mediates sorting and concentration of Golgi resident enzymes into functional COP I vesicles. EMBO J 1999; 18(18):4935-48.
17. Lin CC, Love HD, Gushue JN et al. ER/Golgi intermediates acquire Golgi enzymes by brefeldin A-sensitive retrograde transport in vitro. J Cell Biol 1999; 147(7):1457-72.
18. Malsam J, Satoh A, Pelletier L et al. Golgin tethers define subpopulations of COPI vesicles. Science. 2005; 307(5712):1095-8.
19. Weidman P, Roth R, Heuser J. Golgi membrane dynamics imaged by freeze-etch electron microscopy: Views of different membrane coatings involved in tubulation versus vesiculation. Cell 1993; 75(1):123-33.
20. Happe S, Cairns M, Roth R et al. Coatomer vesicles are not required for inhibition of Golgi transport by G-protein activators. Traffic 2000; 1(4):342-53.
21. Ladinsky MS, Mastronarde DN, McIntosh JR et al. Golgi structure in three dimensions: Functional insights from the normal rat kidney cell. J Cell Biol 1999; 144:1135-1149.
22. Mironov AA, Weidman P, Luini A. Variation on the intracellular transport theme: Maturing cisternae and trafficking tubules. J Cell Biol 1997; 138:481-484.
23. Bannykh SI, Plutner H, Matteson J et al. The Role of ARF1 and Rab GTPases in polarization of the Golgi stack. Traffic 2005; 6(9):803-19.
24. Rabouille C, Klumperman J. Opinion: The maturing role of COPI vesicles in intra-Golgi transport. Nat Rev Mol Cell Biol, 2005 (Epub ahead of print).
25. Krijnse-Locker J, Ericsson M, Rottier PJ et al. Characterization of the budding compartment of mouse hepatitis virus: Evidence that transport from the RER to the Golgi complex requires only one vesicular transport step. J Cell Biol 1994; 124(1-2):55-70.
26. Marsh BJ, Volkmann N, McIntosh JR et al. Direct continuities between cisternae at different levels of the Golgi complex in glucose-stimulated mouse islet beta cells. Proc Natl Acad Sci USA 2004; 101:5565-5570.
27. Ladinsky MS, Kremer JR, Furcinitti PS et al. HVEM tomography of the trans-Golgi network: Structural insights and identification of a lace-like vesicle coat. J Cell Biol 1994; 127(1):29-38.
28. Rink J, Ghigo E, Kalaidzidis Y et al. Rab conversion as a mechanism of progression from early to late endosomes. Cell 2005; 122(5):735-49.
29. Volchuk A, Ravazzola M, Perrelet A et al. Countercurrent distribution of two distinct SNARE complexes mediating transport within the Golgi stack. Mol Biol Cell 2004; 15:1506-1518.
30. Beznoussenko GV, Micaroni M, Trucco A et al. COPI vesicles regulate formation of intercisternal connections extracting Qb SNAREs from Golgi cisternae. J Cell Biol 2005b, (Submitted).
31. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 2002b; 52(Pt 2):297-354.
32. Becker B, Melkonian M. The secretory pathway of protists: Spatial and functional organization and evolution. Microbiological Reviews 1996; 60:697-721.
33. Hehl AB, Marti M. Secretory protein trafficking in Giardia intestinalis. Mol Microbiol 2004; 53:19-28.
34. Katinka MD, Duprat S, Cornillot E et al. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature 2001; 414:450-453.
35. Peyretaillade E, Biderre C, Peyret P et al. Microsporidian Encephalitozoon cuniculi, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. Nucleic Acids Res 1998; 26:3513-3520.
36. Vavra J, Larsson JIR. Structure of the microsporidia. In: Wittner M, Weiss LM, eds. Microsporidia and Microsporidiosis. Washington, DC: American Society for Microbiology, 1999:7-84.

37. Beznoussenko GV, Dolgikh VV, Morzhina EV et al. The microsporidia Golgi: Structure and mechanisms of function. Mol Biol Cell 2005a, (Submitted).
38. Ho HC, Tang CY, Suarez SS. Three-dimensional structure of the Golgi apparatus in mouse spermatids: A scanning electron microscopic study. Anat Rec 1999; 256(2):189-94.
39. Kepes F, Rambourg A, Satiat-Jeunemaitre B. Morphodynamics of the secretory pathway. Int Rev Cytol 2005; 242:55-120.
40. Hebert DN, Garman SC, Molinari M. The glycan code of the endoplasmic reticulum: Asparagine-linked carbohydrates as protein maturation and quality-control tags. Trends Cell Biol 2005; 15(7):364-70.
41. Puthenveedu MA, Linstedt AD. Subcompartmentalizing the Golgi apparatus. Curr Opin Cell Biol 2005; 17(4):369-75.
42. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436(7054):1113-8.
43. McPherson JD et al. A physical map of the human genome. Nature 2001; 409:934-941.
44. Tabata S, Kaneko T, Nakamura Y et al. Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. Nature 2000; 408:823-826.
45. von Mollard GF, Nothwehr SF, Stevens TH. The yeast v-SNARE Vti1p mediates two vesicle transport pathways through interactions with the t-SNAREs Sed5p and Pep12p. J Cell Biol 1997; 137:1511-24.
46. Hong W. SNAREs and traffic. Biochim Biophys Acta. 2005; 1744(3):493-517.
47. Guo Q, Vasile E, Krieger M. Disruptions in Golgi structure and membrane traffic in a conditional lethal mammalian cell mutant are corrected by epsilon-COP. J Cell Biol 1994; 125(6):1213-1224.
48. Lu L, Horstmann H, Ng C et al. Regulation of Golgi structure and function by ARF-like protein 1 (Arl1). J Cell Sci 2001; 114:4543-4555.
49. Salama NR, Schekman RW. The role of coat proteins in the biosynthesis of secretory proteins. Curr Opin Cell Biol 1995; 7:536-543.
50. de Duve C. Reflections on the origin and evolution of life. C R Seances Soc Biol Fil 1998; 192(5):893-901.
51. Rizzotti M. Early evolution. From the appearance of the first cell to the first modern organisms. Switzerland: Birkhäuser Verlag AG, Basel, 2000:180.
52. Hartman H, Fedorov A. The origin of the eukaryotic cell: A genomic investigation. Proc Natl Acad Sci USA 2002; 99:1420-1425.
53. Hartzell PL. Complementation of sporulation and motility defects in a prokaryote by a eukaryotic GTPase. Proc Natl Acad Sci USA 1997; 94(18):9881-6.
54. Jekely G. Small GTPases and the evolution of the eukaryotic cell. Bioessays 2003; 25(11):1129-38.
55. Lee MC, Orci L, Hamamoto S et al. Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. Cell 2005; 122(4):605-17.
56. Aridor M, Fish KN, Bannykh S et al. The Sar1 GTPase coordinates biosynthetic cargo selection with endoplasmic reticulum export site assembly. J Cell Biol 2001; 152(1):213-29.
57. Devos D, Dokudovskaya S, Alber F et al. Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol 2004; 2(12):e380.
58. Schledzewski K, Brinkmann H, Mendel RR. Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2, and 3 and the F subcomplex of the coatomer COPI. J Mol Evol 1999; 48(6):770-8.
59. Pishvaee B, Munn A, Payne GS. A novel structural model for regulation of clathrin function. EMBO J 1997; 16(9):2227-39.
60. Wang J, Virta VC, Riddelle-Spencer K et al. Compromise of clathrin function and membrane association by clathrin light chain deletion. Traffic 2003; 4(12):891-901.
61. Dacks JB, Doolittle WF. Novel syntaxin gene sequences from Giardia, Trypanosoma and algae: Implications for the ancient evolution of the eukaryotic endomembrane system. J Cell Sci 2002; 115:1635-1642.
62. Miller E, Antonny B, Hamamoto S et al. Cargo selection into COPII vesicles is driven by the Sec24p subunit. EMBO J 2002; 21(22):6105-13.
63. Opat AS, Houghton F, Gleeson PA. Steady-state localization of a medial-Golgi glycosyltransferase involves transit through the trans-Golgi network. Biochem J 2001; 358(Pt 1):33-40.
64. Mogelsvang S, Gomez-Ospina N, Soderholm J et al. Tomographic evidence for continuous turnover of Golgi cisternae in Pichia pastoris. Mol Biol Cell 2003; 14(6):2277-91.
65. Riezman H. Three clathrin-dependent budding steps and cell polarity. Trends in Cell Biology 1993; 3:330-332.
66. Mironov AA, Luini A, Buccione R. Constitutive transport between the trans-Golgi network and the plasma membrane according to the maturation model. A hypothesis FEBS Lett 1998; 440(1-2):99-102.
67. Beznoussenko GV, Mironov AA. Models of intracellular transport and evolution of the Golgi complex. Anat Rec 2002; 268:226-238.

CHAPTER 6

An Evolutionary Perspective on Eukaryotic Membrane Trafficking

Cemal Gurkan, Atanas V. Koulov and William E. Balch*

Abstract

The eukaryotic cell is defined by a complex set of sub-cellular compartments that include endomembrane systems making up the exocytic and endocytic trafficking pathways. Current evidence suggests that both the function and communication between these compartments are regulated by distinct families of proteins that direct membrane fission, targeting and fusion. These families include coat protein complexes (CPCs) involved in vesicle formation/fission, Rab GTPases involved in vesicle targeting, and soluble *N*-ethyl-maleimide-sensitive factor attachment protein receptors (SNAREs) involved in vesicle fusion. The origins of these gene families and their individual contributions to the evolutionary specialization of the membrane architectures of lower and higher eukaryotes are now better understood with the advent of powerful phylogenetic, structural and systems biology tools. Herein, we provide a perspective that suggests that while the core CPC and SNARE machineries have diversified modestly in the course of eukaryotic evolution, the Rab GTPase family expanded substantially to emerge as a key driving force in endomembrane specialization. The Rab GTPases appear to have provided the foundation for the intricate membrane architectures ranging from those requisite for the distinct amoebic life cycle stage of uni-cellular organisms such as the parasitic protozoa to the highly specialized tissue and cell type-specific endomembranes of multi-cellular eukaryotes. We propose that Rab-centric interaction networks orchestrate the divergent activities of fission and fusion through their capacity to control the sequential assembly of protein complexes that mediate endomembrane structure and communication.

Introduction

The presence of sub-cellular compartments is a truly eukaryotic feature that provides spatially distributed chemical micro-environments required for the normal cell and tissue function. These compartments are distinguished by the varying lipid composition of the encapsulating bilayer as well as unique sets of integral and peripheral membrane proteins. In addition to the sub-cellular compartments such as mitochondria and chloroplasts, elaborate endomembrane systems also define the exocytic and endocytic vesicular trafficking pathways. These include the endoplasmic reticulum (ER) and the contiguous nuclear envelope (NE), the Golgi apparatus, as well as post-Golgi compartments such as lysosomes, secretory granules, and early/late endosomes. While the endosymbiotic origin of mitochondria and chloroplasts is

*Corresponding Author: William E. Balch—Departments of Cell and Molecular Biology, and The Institute for Childhood and Neglected Disease, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, U.S.A. Email: webalch@scripps.edu

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

generally accepted,[1] hypotheses on the evolution of endomembranes from specialized invaginations of the plasma membrane largely remain controversial as they fail to account for specialized structures such as the nuclear pore complexes (NPCs).[2-4]

Endomembranes associated with sub-cellular trafficking are particularly dynamic structures that are in continuous communication afforded by the activity of highly-specialized protein complexes that harness and regulate the fundamental processes of membrane fission, tethering and fusion. For example, coat protein complexes (CPCs) mediate the biogenesis of cargo-bearing vesicles from the donor membranes while the Rab GTPases and soluble *N*-ethyl-maleimide-sensitive factor attachment protein receptors (SNAREs) mediate sub-cellular targeting and subsequent docking/fusion of these membrane-bound containers to the target membranes, respectively. In this respect, a hallmark of eukaryotic evolution has been the emergence of the Ras superfamily of GTPases that function as central regulators of membrane budding and trafficking, as well as cytoskeletal dynamics and the biogenesis of the nucleus.[4,5]

Notably, three out of at least seven major families comprising the Ras superfamily of small GTPases directly mediate specific aspects of endomembrane trafficking dynamics: the Sar1/Sara and Arf GTPase families (at least 2 and 6 members in mammalian cells, respectively) regulate membrane recruitment and stability of CPCs, while the Rab GTPases (almost 70 members in mammalian cells) play essential roles in all stages of vesicular trafficking.[5-8] Intriguingly, a recent phylogenetic analysis of the Ras superfamily led to the unprecedented hypothesis that the emergence of eukaryotic endomembranes might have predated that of phagocytosis, which is traditionally regarded as a prerequisite for the evolution of endomembrane systems, as well as that for the endosymbiotic organelles.[1,4] For instance, the membrane tubulating activity of small GTPases from the Ras superfamily, such as that of Sar1,[9,10] might have provided an alternative means for the first endomembrane biogenesis.[4]

Despite their respectively crucial roles in membrane trafficking, the molecular basis for the integrated function of the CPC, Rab, and SNARE machineries has so far been elusive. Traditionally, phylogenetic analyses of proteins in a gene family are used to identify potential functional relationships to other family members, such as in the case of Rabs and SNAREs.[8,11-13] Computational approaches that apply hierarchical clustering algorithms to systematic tissue mRNA expression profiling can be used to complement this phylogenetic annotation by providing further insights into the physiological activity of closely related and distant family members, and to different gene families in different cell types.[6,14] Indeed, recent results from our laboratory using one such systems biology approach now suggest that Rab-regulated activity hubs may constitute an integrated coding system, the membrome network, that orchestrates the dynamics of the specialized membrane architecture of differentiated cells.[6] In this chapter we will focus on the Rab-centric integration of the eukaryotic membrane trafficking machineries from an evolutionary perspective, and also briefly discuss the origins of the nuclear and exocytic/endocytic endomembrane systems.

Coat Protein Complexes: Cellular Machineries Driving Vesicle Formation/Fission

Biogenesis of transport containers that shuttle cargo between endomembranes and/or to and from the plasma membrane is mediated by CPCs. These include the coat protein complex II (COPII) that mediates ER-to-Golgi vesicular trafficking, coat protein complex I (COPI) that mediates intra-Golgi and Golgi-to-ER trafficking,[15] and the clathrin-based CPCs that are involved in endocytosis and trafficking between the Golgi, lysosomes and endosomes.[16] The core CPC machineries are highly conserved throughout eukaryotic evolution: *Saccharomyces cerevisiae* (6 members, 31 subunits), *Caenorhabditis elegans* (6 members, 29 subunits), *Drosophila melanogaster* (6 members, 29 subunits), and *Homo sapiens* (7 members, 53 subunits).[8] In other words, higher eukaryotes use the same basic modular system (i.e., core CPCs) where specificity demanded by the specialized membrane trafficking events and increased cargo complexity is achieved through the differential use and specialization of subunits (i.e., CPC components) that diversified significantly from yeast to human.

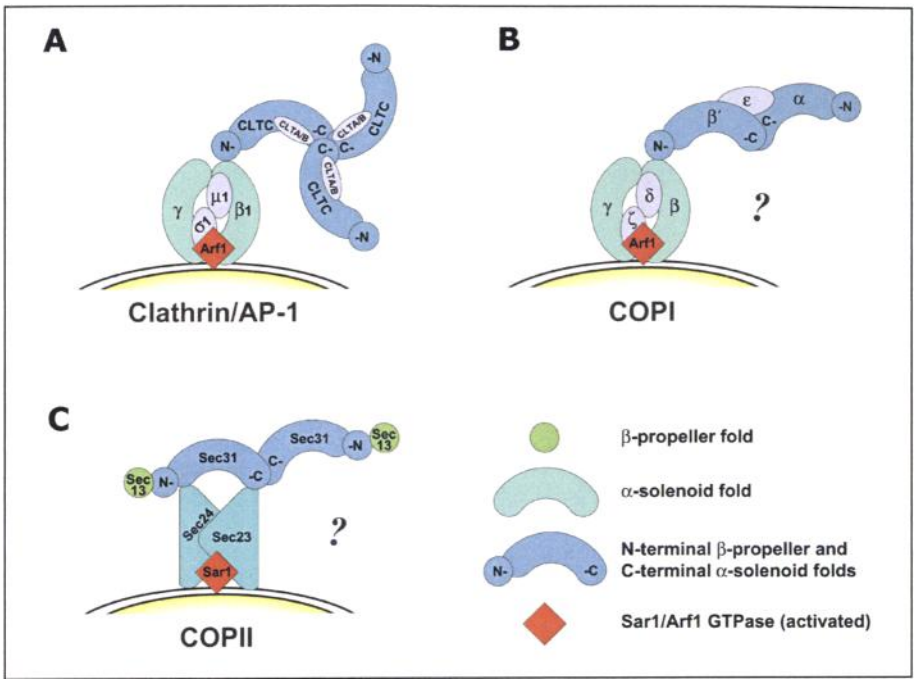


Figure 1. Schematic depiction of typical clathrin-based (A), COPI (B), and COPII (C) coat protein complexes (CPCs) emphasizing their structural and functional homologies emerging from biochemical, biophysical, and in silico prediction studies.[16,19,20,47]

Initial evidence towards the shared origins of highly-specialized CPCs emerged from phylogenetic studies that demonstrated an evolutionary link between the components of the COPI and Adaptor Protein complexes 1, 2 and 3 (AP-1, AP-2 and AP-3),[17] as well as from structural[18] and biochemical[19] comparisons of COPI and AP-2 or AP1/AP-3 subunits (Fig. 1A and B). In the case of COPI and AP-1/AP-3, a shared structural principle was already evident from the observation that the small Arf1 GTPase is involved in the membrane recruitment of both type of CPCs.[17,19] Moreover, a recent study using computational and biochemical methods predicted that most components of COPI, COPII and clathrin-based CPCs may share the same basic set of structural domains, namely β-propeller and α-solenoid folds that comprise of repeating WD-40 containing β-sheets, and α-helices, respectively (Fig. 1A-C).[20] While these structural folds are common protein domains, the distinctive domain arrangement of an N-terminal β-propeller followed by an α-solenoid observed in most CPC components is absent in eubacteria and archaebacteria.[20]

Based on the prediction that an essential step in the evolution of endomembranes must have been the emergence of coated vesicle budding, Cavalier-Smith[1] also previously hypothesized that COPI, COPII and clathrin-based CPCs are likely to have a common ancestral origin. Cavalier-Smith[1] further suggested the following sequence of events for the coated vesicle evolution: (1) an ancestral COPII evolved first to drive budding from primitive, rough-ER like endomembranes, (2) fusion of some of the COPII-driven vesicles with each other generated a smooth endomembrane compartment, a proto-Golgi/lysosome intermediate between the ER and the plasma membrane, (3) clathrin-based CPCs evolved allowing the separation of lysosomes from the Golgi, with a secondary role in endocytosis, and finally (4) COPI evolved, creating the trans-Golgi network as an intermediate compartment between Golgi cisternae and lysosomes. This hypothesis remains intriguing as it is consistent with the recent biochemical

and computational data (Fig. 1A-C),[20] and the earlier observations of closer phylogenetic, biochemical and structural links between the components of the clathrin-based and COPI CPCs (Fig. 1A-B).[17-19] Further support for this hypothesis is available from earlier phylogenetic analysis that suggested a more ancestral character for the COPII-mediating Sar1 GTPase among the members of the Ras superfamily.[4] This is also consistent with the fact that Sar1 has the unique capability to complement the sporulation defects of a *Myxococcus xanthus* strain deficient in MglA, a member of the prokaryotic GTPase family that is the closest relative of the eukaryotic Ras superfamily.[21] Finally, the argument that endocytosis might have emerged only as a secondary clathrin-based CPC function is suggested by the fact that the endocytic event mediated by clathrin/AP-2 is truly unique among CPCs in not requiring a small GTPase (e.g., Sar1 or Arf) for its membrane recruitment, but instead relies on direct interaction with a phospholipid, namely PtdIns(4,5)P_2.[16] This may reflect a more specialized process that is likely to have evolved along with a unique class of lipid effectors mediating cell surface signaling events.

A clearer picture is now emerging regarding the molecular basis for the overall membrane dynamics of coated vesicle formation and the precise role played by the individual CPC components. For example, the minimal COPII machinery that mediates ER cargo export comprises of the activated Sar1 GTPase (GTP-bound) and the Sec23/Sec24 (Sec23/24) and Sec13/Sec31 (Sec13/31) hetero-oligomers (Fig. 1C).[15,16,22] Here, Sec23 acts as a Sar1-specific GTPase activating protein (GAP) and Sec24 functions as a cargo adaptor specific for the ER exit signals, while Sec13/31 largely plays a structural role. The Sar1 GTPase may promote the initial membrane morphogenesis through its tubulating activity,[9,10] in addition to its well-established role in nucleating the COPII assembly.[22] This initial membrane curvature is then likely to be 'captured' by the electrostatic interactions between the concave surface of Sec23/24, which is enriched in basic residues and predicted to make an extensive contact with the underlying lipid bilayer,[9,23] and presumably further propagated by the polymerizing properties of the Sec13/31 hetero-oligomers.[16] During the biogenesis of clathrin coated vesicles (CCVs), the molding of the target membranes into regions of high curvature is also thought to be primarily promoted by the accessory proteins rather than the roughly spherical structure of the clathrin cage itself.[16,24] In general, cognate adaptors provide CPCs with a direct link to biosynthetic cargo destined for export and also coordinate cargo selection with vesicle and tubule formation.[16,25]

Intriguingly, the β-propeller and α-solenoid folds described above, as well as their unique N- to C-terminal domain arrangement predicted to constitute some of the CPC components (Fig. 1A-C), may also be shared with some of the NPC components.[2,20] This suggests unprecedented evolutionary links between the components of the NPCs and CPCs. The likelihood of these evolutionary links is strengthened by the observations that: (1) the ER and the nuclear envelope form a contiguous membrane where the latter is absorbed into the former during mitosis, and then subsequently reassembled from the former by membrane tubulation,[26] (2) the Sec13 subunit of the COPII coat complex and its larger isoform Sec13 like/Seh1, which has no hitherto known function in COPII vesicle biogenesis, play well-established structural roles in NPCs,[27,28] (3) assembly of nuclear membranes and NPCs is regulated by another member of the Ras superfamily, the Ran GTPase family,[5] and (4) unlike the conventional transmembrane channels, NPCs do not span the lipid bilayer but instead form aqueous channels that stabilize the sharp convex curvature formed by the contiguous inner and outer NE membranes, reminiscent of the basic CPC function during membrane vesiculation.[29] At the molecular level, one of the most fundamental events during the evolutionary leap from prokaryotes to eukaryotes must have involved a radical change in membrane topologies associated with the emergence of CPC-mediated budding and NPCs.[1] Based on recent computational and biochemical data, Devos et al[20] proposed that early eukaryotes might have had a 'protocoatomer' module that induced curvature of endomembranes, and which subsequently gave rise to variants with the specialized functions of CPCs and NPCs. In a sense, the primordial NPC could then be envisaged as a 'defective' vesicle budding complex enveloping and curving membranes linked to the chromatin.[20,30]

SNARE Proteins: Cellular Machineries Driving Membrane Docking/Fusion

The SNARE family consists of a cognate group of integral and peripheral membrane proteins that function in the final stages of vesicular transport. This step involves tethering/docking and subsequent fusion of the transport container with the target membrane.[13,31] The core structural feature of all SNAREs is an evolutionarily conserved SNARE motif of about 60 residues.[32] Based on their highly conserved structural features that contribute to the reversible assembly of quaternary docking-fusion complexes, SNAREs are classified into Q- and R-SNARE sub-families.[32] Here, each Q- and R-SNARE family member is believed to contribute differentially to docking and fusion by providing specific information that correctly directs the close juxtaposition of two membrane bilayers at specific steps of the exocytic and endocytic pathways. Finally, bilayer docking/fusion mediated by the SNARE complexes is highly regulated by a variety of pathway-specific effectors that either promote (matchmakers) or prevent (matchbreakers) SNARE assembly pathways.[33]

With the completion of major eukaryotic genome projects, it is now clearly evident that only a modest increase (~1.5-fold) has taken place in the number of SNARE family members with the expanding developmental complexity from yeast to human: *S. cerevisiae* (21 members), *C. elegans* (23 members), *D. melanogaster* (20 members), and *H. sapiens* (36 members).[8,13] At a first glance, this rather modest increase in SNARE diversification may suggest that multi-cellular organisms do not necessarily have an inherently more complex sub-cellular trafficking system, and that a core set of SNAREs is largely sufficient for requisite membrane fusion events.[8] This is reminiscent of the limited diversification observed for the core CPC machineries from yeast to human as discussed in the previous section. However, it is clearly evident that higher eukaryotes do indeed have more complex sub-cellular trafficking systems. To accommodate the needs of multi-cellular specialization, higher eukaryotes might have used the tissue-specific differential expression of an increased number of SNAREs and other additional regulatory membrane trafficking components.[6,8]

Rab GTPases: Key Regulators of Membrane Trafficking

Rab GTPases comprise the largest family of the Ras superfamily of small GTPases and function as molecular switches that regulate the dynamic assembly and disassembly of multi-protein scaffolds involved in vesicular traffic.[5,6,34,35] Most Rabs are post-translationally modified with geranylgeranyl hydrocarbon chains that enable their partitioning into membranes.[36] Moreover, the membrane association of Rabs is under the control of the Rab-GDI recycling system.[37] Fundamentally, Rab proteins can be viewed as simple molecular switches ('on' and 'off' being their GTP- and GDP-bound states, respectively) that are associated with membranes in their activated state. However, Rabs have weak intrinsic guanine nucleotide exchange and hydrolysis activities, therefore their interactions with downstream effectors are regulated by guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs) that promote the cyclical assembly and disassembly of Rab-mediated protein complexes.[38,39] Effector complexes formed in response to Rab activation perform diverse functions that include coupling of endomembranes to motors, and hence to the cytoskeleton, as well as long-range vesicle docking/tethering interactions that are likely to regulate the membrane fusion activity mediated by the SNARE machinery.[6,34,35]

The number of Rab family members closely correlates evolutionarily with increasing endomembrane complexity: *Schizosaccharomyces pombe* (7 members), *S. cerevisiae* (11 members), *C. elegans* (29 members), *D. melanogaster* (29 members), *Arabidopsis thaliana* (57 members), and *H. sapiens* (63 members).[8,11,12] This nearly 6-fold expansion in the number of Rab family members from yeast to human reflects the larger number of specialized trafficking pathways in the differentiated cell types forming organ systems of higher eukaryotes.[6] In protozoal parasites such as *Entamoeba histolytica* and *Trichomonas vaginalis*, the Rab diversity may be as large as or even higher than that observed in higher eukaryotes,[40,41] which could be due to expanding

membrane trafficking needs associated with their amoebic life cycle phase as in the case of *Dictyostelium discoidium* that may have up to 54 Rabs.[41]

Compared to the 6-fold increase in the number of Rab GTPases from yeast to human, as indicated previously, only a marginal and at most modest increase has taken place in the number of core CPC modules and SNARE family members, respectively. This raises the possibility that Rabs may function as tethering/targeting/fusion activity hub organizers that provide the primary diversification element for membrane trafficking pathways by altering the combinatorial potential for protein interactions through coupling their GTPase activity with effector interacting (switch) domains (Fig. 2).[6,38,42] While SNAREs direct late events leading to membrane docking/fusion, Rabs mediate vesicle targeting through the recruitment of membrane oriented tethering components that forge links with fusion factors to coordinate cargo transport with membrane flow. Moreover, the ability of Rab GTPases to couple transport containers to motor proteins can be conceptually viewed as a 'tethering' function that establishes the distribution of organelles within the cytoskeletal network. Such Rab-based hubs will rely heavily on the unique tissue distribution of Rab GEFs and GAPs as well as general modulators such as GDIs that facilitate Rab GTPase recycling.[6]

The above observations suggest that during the course of eukaryotic evolution and the accompanying increase in the developmental complexity (i.e., presence of an amoeboid life cycle phase, tissue differentiation and organogenesis), the Rab GTPases are likely to have emerged as the main regulatory system orchestrating the requisite membrane trafficking pathways. Given the assumption that the dramatic diversification of the Rab GTPases in higher eukaryotes reflects the membrane specialization, we have found that mRNA expression profiling provides a useful and unbiased bioinformatics approach to understanding Rab-centric organization of the membrane architecture of cells and tissues.[6,14] These Rabs-centric coding systems are likely to regulate specific membrane interactions and cargo flow between the sub-cellular compartments. We define this general system of Rab-regulated hubs of protein interactions in higher eukaryotes as the 'membrome' for a given cell type or transport activity.[6] In this view, Rab and SNARE machineries, which are possibly linked through the activity of tethers, constitute the minimal core components of the membrome and their activity is regulated by cohub components that include Rab/SNARE regulators and effectors, which directly or indirectly interact with the components of CPCs to define cargo trafficking pathways. In a given cell type (or specialized life cycle stage), the membrome varies substantially to reflect the unique expression profiles of its components, thereby dictating unique membrane architectures and their consequential functionalities.

How do Rab-based hubs function at the molecular level? Rab function is based on the simple chemical reaction of hydrolysis of a phospho-diester bond in the guanosinetriphosphate (GTP) molecule, which results in conformational changes in the GTPase effector/switch regions. Analysis of the primary protein sequences within the Rab family of proteins reveals strongly conserved motifs, namely RabF motifs, which contain the effector-interacting switch I and II regions (Fig. 2A and B).[11,12] In addition to the high degree of sequence conservation within the switch I and II regions, structural data demonstrate remarkable preservation of the three-dimensional fold.[43] However, despite this sequence and structural homology between the individual Rabs, a clearly divergent course of molecular evolution has taken place within the Rab family. Interestingly, phylogenetic analyses allow the arrangement of all Rabs into eight functional sub-groups, such as group V – endosomal, group III – secretory, etc.[12] It is now evident that the specificity of the interactions between the Rab GTPases and their corresponding effector proteins has been finely tuned during evolution through the introduction of subtle changes in the amino acid composition that map onto the surface of Rab effector regions. Interestingly, these changes have been limited to those preserving the chemical properties of the individual amino acid residues. For example, variations in the RabF1 motif of human Rabs are largely limited to a substitution of Isoleucine with a Valine, which are both alipathic amino acid residues, and/or an Aspartate with a Glutamate, which are both charged residues (Fig. 2B).

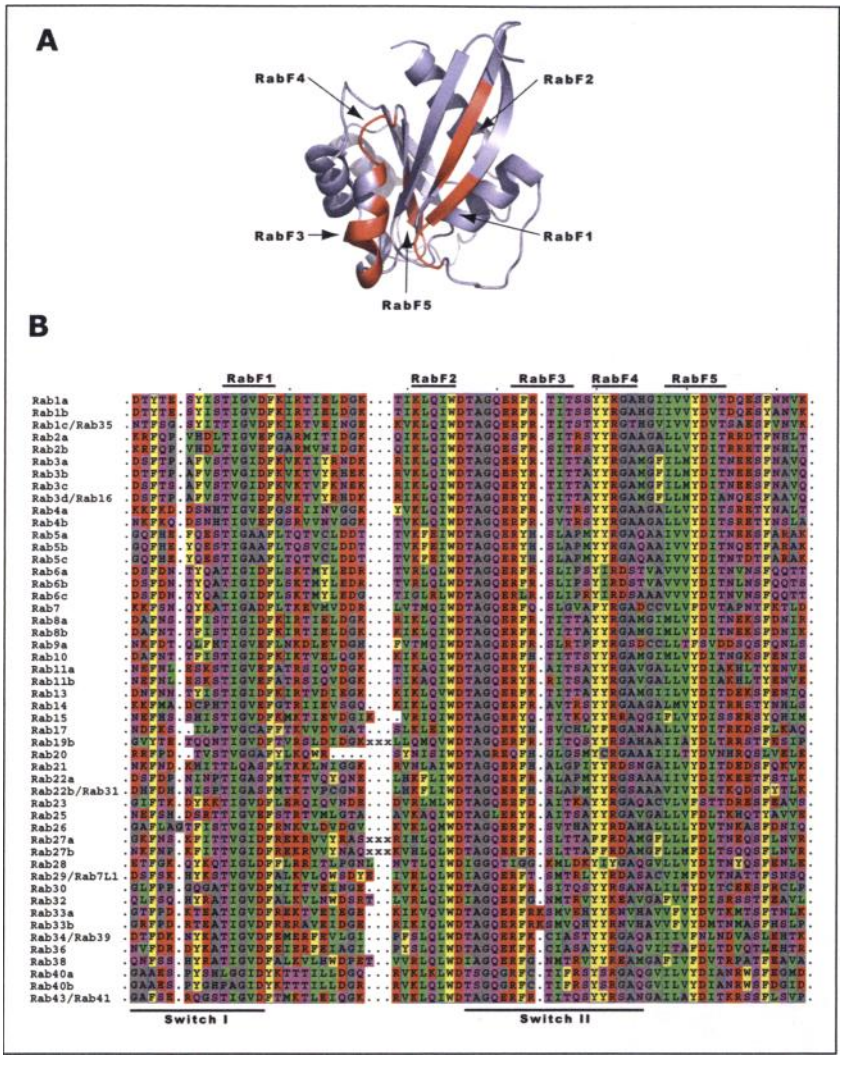


Figure 2. A) The three-dimensional structure of the rat Rab3A GTPase (PDB ID: 3RAB).[48] The conserved RabF motifs are highlighted in red. B) Primary amino acid sequence alignment of 50 human Rab GTPases [see Gurkan et al[6] for a complete list of GenBank® accession numbers for the human Rabs]. Only the conserved RabF motifs and the switch regions within (I and II) are shown. Amino acid residues are colored by their properties as follows: charged (KRDE) – *red*, polar (NQST) – *magenta*, aliphatic (ILMV) – *green*, aromatic (FYW) – *yellow*, others (APCGH) – *grey*. Short stretches of amino acid residues in Rab19b, Rab27a and b excluded for the clarity of presentation are shown as 'xxx'.

Concurrently, the Rab GTPases are promiscuous in their interactions with effectors. For example, Rab27a can interact with both melanophilin (Slac2-a) or MyRIP (Slac2-c), which are effectors that couple it to the molecular motors MyosinVa and Myosin VIIa, respectively.[44] MyosinVa is responsible for the localization of melanosomes in melanocytes,[45] while the same function in retinal pigment epithelial cells is performed by MyoVIIa.[46] Clearly, this is an evolutionary adaptation in response to the different motility needs of the same organelle in

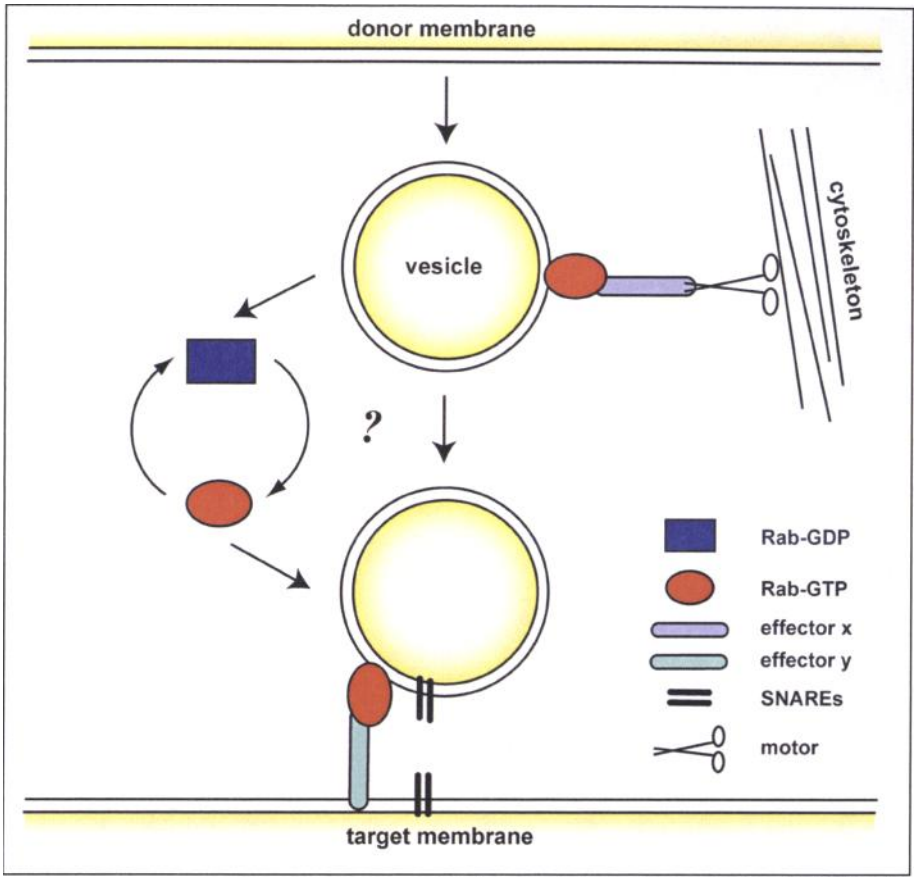


Figure 3. Schematic representation of the known Rab GTPase functions. Rabs operate as molecular switches in at least two different scenarios: tethering membrane transport vesicles to molecular motors (top) or effector molecules (bottom) that mediate the assembly of membrane fusion machinery. Both scenarios may be interchangeable depending on the availability of binding partners (local concentration, $[e_n]$, and binding affinity constant, *Kd*).

different cell types reflecting myosin specialization. Given that multiple specialized Rabs are present across a variety of different tissues,[6] it is very likely that this adaptation is a general feature within the Rab family. Thus, evolution has allowed the regulation of related trafficking steps using variations in the effector domain function coupled to the availability of different ensemble of effectors in each tissue.

In addition to the above levels of specificity, it is also possible that within the same cell type, Rabs may bind different effectors as a function of the spatial sub-cellular distribution of these partners. For example, the same Rab GTPase may be responsible for the tethering of an organelle to a motor and also for its interaction with the target membrane-bound tethering molecules as part of a sequential pathway of effector interactions (Fig. 3). If the function of Rab GTPases is to allow tethering of vesicle membranes to molecular motors (or other molecules), then a wide variety of biological activities arises from the combinatorial interactions of a given Rab with a wide array of downstream effectors and their respective binding partners. Such an array of interactions will depend on: (1) availability of each specific effector within the given tissue based on its expression levels and local protein concentration, and (2) the binding

constant for this interaction. In other words, the biological activity of a Rab GTPase can be defined as the following function *(f)*:

$$Biological\ activity = f(Kd_1,[e_1],Kd_2,[e_2]...Kd_m,[e_n])$$

where Kd_n is an affinity constant of a given Rab to an effector e_n, and $[e_n]$ is the local concentration of the given effector. The nature of this relationship remains to be tested as more accurate measurements become available and more effectors are identified.

The seemingly simple cycle of GTP-/GDP-bound ('on'/'off') state of Rabs that constitutes the basis for Rab activity as molecular switches, has also been augmented during evolution by multiple layers of regulation. In addition to the Rab-GDI recycling system that mediates the membrane recruitment of the Rab GTPases,[37] the specific activity of a given Rab is governed by the unique tissue/sub-cellular distribution of its requisite GEFs and GAPs that facilitate different Rab/effector exchange reactions. Thus, Nature has managed to respond kinetically to the various intracellular trafficking needs of complex multi-cellular organisms by utilizing these multiple regulatory mechanisms to generate a large variety of activities.[6]

Conclusions

The evolutionary leap from single to multi-cellular organisms did not require a substantial increase in the number of the core CPC and SNARE machineries that mediate the fundamental events of vesicle fission and fusion, respectively. Rather, they have a limited evolutionary course where the core CPC machineries (i.e., the cage-forming modules) and SNARE complexes (i.e., the membrane fusion modules) appear to have specialized with expanding cargo complexity through the diversification of the adaptor, regulatory factors and accessory proteins instead. In contrast, the drastic diversification of the Rab GTPases that regulate the protein interaction hubs, such as that orchestrating the tethering and fusion processes, suggests that Rabs are the principal evolutionarily element facilitating the expansion and specialization of membrane trafficking pathways. It now appears that specialization and functional divergence of membrane function has arisen through the ability of Rab GTPases to interact with a vast array of effectors in a combinatorial fashion. This is further fine-tuned by tissue-specific expression and the action of Rab GTPases and their respective regulators. A formidable challenge in the field of cell biology will be the elucidation of the molecular mechanisms by which Rab-regulated tethering-fusion activity hubs integrate with that of coated vesicle formation/fission to shape the unique architectures of eukaryotic cells.

Acknowledgements

This work is supported by grants from the National Institutes of Health (GM33301 and GM42336) to WEB. CG and AVK are Cystic Fibrosis Foundation Postdoctoral Research Fellowship recipients. This is TSRI Manuscript No.: 17795-CB.

References

1. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 2002; 52(Pt 2):297-354.
2. Mans BJ, Anantharaman V, Aravind L et al. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. Cell Cycle 2004; 3(12):1612-1637.
3. Blobel G. Intracellular protein topogenesis. Proc Natl Acad Sci USA 1980; 77(3):1496-1500.
4. Jekely G. Small GTPases and the evolution of the eukaryotic cell. Bioessays 2003; 25(11):1129-1138.
5. Colicelli J. Human RAS superfamily proteins and related GTPases. Sci STKE 2004; 2004(250):RE13.
6. Gurkan C, Lapp H, Alory C et al. Large-scale profiling of Rab GTPase trafficking networks: The membrome. Mol Biol Cell 2005; 16(8):3847-3864.
7. Donaldson JG, Honda A. Localization and function of Arf family GTPases. Biochem Soc Trans 2005; 33(Pt 4):639-642.
8. Bock JB, Matern HT, Peden AA et al. A genomic perspective on membrane compartment organization. Nature 2001; 409(6822):839-841.
9. Lee MC, Orci L, Hamamoto S et al. Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. Cell 2005; 122(4):605-617.

10. Aridor M, Fish KN, Bannykh S et al. The Sar1 GTPase coordinates biosynthetic cargo selection with endoplasmic reticulum export site assembly. J Cell Biol 2001; 152(1):213-229.
11. Pereira-Leal JB, Seabra MC. The mammalian Rab family of small GTPases: Definition of family and subfamily sequence motifs suggests a mechanism for functional specificity in the Ras superfamily. J Mol Biol 2000; 301(4):1077-1087.
12. Pereira-Leal JB, Seabra MC. Evolution of the Rab family of small GTP-binding proteins. J Mol Biol 2001; 313(4):889-901.
13. Hong W. SNAREs and traffic. Biochim Biophys Acta 2005; 1744(2):120-144.
14. Gurkan C, Lapp H, Hogenesch J et al. Exploring trafficking GTPase function by mRNA expression profiling: Use of the SymAtlas web-application and the membrome datasets. Methods Enzymol 2005; 403:1-10.
15. Lee MC, Miller EA, Goldberg J et al. Bi-directional protein transport between the ER and Golgi. Annu Rev Cell Dev Biol 2004; 20:87-123.
16. McMahon HT, Mills IG. COP and clathrin-coated vesicle budding: Different pathways, common approaches. Curr Opin Cell Biol 2004; 16(4):379-391.
17. Schledzewski K, Brinkmann H, Mendel RR. Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2, and 3 and the F subcomplex of the coatomer COPI. J Mol Evol 1999; 48(6):770-778.
18. Hoffman GR, Rahl PB, Collins RN et al. Conserved structural motifs in intracellular trafficking pathways: Structure of the gammaCOP appendage domain. Mol Cell 2003; 12(3):615-625.
19. Eugster A, Frigerio G, Dale M et al. COP I domains required for coatomer integrity, and novel interactions with ARF and ARF-GAP. Embo J 2000; 19(15):3905-3917.
20. Devos D, Dokudovskaya S, Alber F et al. Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol 2004; 2(12):e380.
21. Hartzell PL. Complementation of sporulation and motility defects in a prokaryote by a eukaryotic GTPase. Proc Natl Acad Sci USA 1997; 94(18):9881-9886.
22. Matsuoka K, Orci L, Amherdt M et al. COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes. Cell 1998; 93(2):263-275.
23. Bi X, Corpina RA, Goldberg J. Structure of the Sec23/24-Sar1 prebudding complex of the COPII vesicle coat. Nature 2002; 419(6904):271-277.
24. Owen DJ, Collins BM, Evans PR. Adaptors for clathrin coats: Structure and function. Annu Rev Cell Dev Biol 2004; 20:153-191.
25. Robinson MS. Adaptable adaptors for coated vesicles. Trends Cell Biol 2004; 14(4):167-174.
26. Ellenberg J, Siggia ED, Moreira JE et al. Nuclear membrane dynamics and reassembly in living cells: Targeting of an inner nuclear membrane protein in interphase and mitosis. J Cell Biol 1997; 138(6):1193-1206.
27. Siniossoglou S, Lutzmann M, Santos-Rosa H et al. Structure and assembly of the Nup84p complex. J Cell Biol 2000; 149(1):41-54.
28. Siniossoglou S, Wimmer C, Rieger M et al. A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. Cell 1996; 84(2):265-275.
29. Schwartz TU. Modularity within the architecture of the nuclear pore complex. Curr Opin Struct Biol 2005; 15(2):221-226.
30. Jekely G. Glimpsing over the event horizon: Evolution of nuclear pores and envelope. Cell Cycle 2005; 4(2):297-299.
31. Ungar D, Hughson FM. SNARE protein structure and function. Annu Rev Cell Dev Biol 2003; 19:493-517.
32. Fasshauer D, Sutton RB, Brunger AT et al. Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. Proc Natl Acad Sci USA 1998; 95(26):15781-15786.
33. Gerst JE. SNARE regulators: Matchmakers and matchbreakers. Biochim Biophys Acta 2003; 1641(2-3):99-110.
34. Deneka M, Neeft M, van der Sluijs P. Regulation of membrane transport by rab GTPases. Crit Rev Biochem Mol Biol 2003; 38(2):121-142.
35. Zerial M, McBride H. Rab proteins as membrane organizers. Nat Rev Mol Cell Biol 2001; 2(2):107-117.
36. Alory C, Balch WE. Molecular basis for Rab prenylation. J Cell Biol 2000; 150(1):89-103.
37. Alory C, Balch WE. Organization of the Rab-GDI/CHM superfamily: The functional basis for choroideremia disease. Traffic 2001; 2(8):532-543.
38. Spang A. Vesicle transport: A close collaboration of Rabs and effectors. Curr Biol 2004; 14(1):R33-34.

39. Bernards A. GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and Drosophila. Biochim Biophys Acta 2003; 1603(2):47-82.
40. Saito-Nakano Y, Loftus BJ, Hall N et al. The diversity of Rab GTPases in Entamoeba histolytica. Exp Parasitol 2005; 110(3):244-252.
41. Lal K, Field MC, Carlton JM et al. Identification of a very large Rab GTPase family in the parasitic protozoan Trichomonas vaginalis. Mol Biochem Parasitol 2005; 143(2):226-235.
42. Pfeffer SR. Rab GTPases: Specifying and deciphering organelle identity and function. Trends Cell Biol 2001; 11(12):487-491.
43. Pfeffer SR. Structural clues to Rab GTPase functional diversity. J Biol Chem 2005; 280(16):15485-15488.
44. Fukuda M. Versatile role of Rab27 in membrane trafficking: Focus on the Rab27 effector families. J Biochem (Tokyo) 2005; 137(1):9-16.
45. Futter CE, Ramalho JS, Jaissle GB et al. The role of Rab27a in the regulation of melanosome distribution within retinal pigment epithelial cells. Mol Biol Cell 2004; 15(5):2264-2275.
46. Gibbs D, Azarian SM, Lillo C et al. Role of myosin VIIa and Rab27a in the motility and localization of RPE melanosomes. J Cell Sci 2004; 117(Pt 26):6473-6483.
47. Matsuoka K, Schekman R, Orci L et al. Surface structure of the COPII-coated vesicle. Proc Natl Acad Sci USA 2001; 98(24):13705-13709.
48. Dumas JJ, Zhu Z, Connolly JL et al. Structural basis of activation and GTP hydrolysis in Rab proteins. Structure Fold Des 1999; 7(4):413-423.

CHAPTER 7

Reconstructing the Evolution of the Endocytic System:
Insights from Genomics and Molecular Cell Biology

Mark C. Field*, Carme Gabernet-Castello and Joel B. Dacks

Endocytosis is an essential process undertaken by most eukaryotic cells. At its most general, the term refers to the uptake of material from the cell milieu.[1] Cell biologists, however, have come to recognise a number of distinct modes of endocytic transport that are accompanied by differences in their underlying molecular mechanisms. Multiple modes can coexist in the same cell type and are frequently ongoing concurrently. Broadly, endocytic mechanisms can be subdivided based on the size of the ingested particle or cargo. Phagocytosis, or cell eating, is the uptake of large particles, including whole cells, and is accompanied by transport through large vesicular structures (>250nm in diameter). Pinocytosis, or cell drinking, involves uptake of rather smaller cargo, typically macromolecules and complexes. The study of endocytic pathways has, for very good technical reasons, focused on a small number of taxa, principally metazoa, yeast and a restricted number of protists. This has served well and has allowed the definition of a number of pathways in part by virtue of the molecules that are required for their operation.

A more complete understanding of eukaryotic diversity, both with respect to the evolutionary relationships between the major groups, and the availability of an increasingly representative taxonomic sampling of genomes now allows for a meaningful survey of the endocytic potential of eukaryotes and the reconstruction of major events in the evolution of the endocytic system. From a combination of single and multi-gene phylogenetic studies, along with morphological data, there are now recognized six major "super-groups" of eukaryotes,[2,3] as shown in Figure 1. The use of rare genetic characters, such as gene losses, innovations, and especially gene fusions, have also been key in establishing these groups and in rooting the eukaryotic tree. The presence of a derived gene fusion between dihydrofolate reductase and thymidilate synthase shared by the plantae, chromalveolata, excavata and rhizaria, and a unique myosin gene fusion in amoebae and opisthokonts have divided the eukaryotes into two groups; the bikonts on one side and the unikonts on the other.[4,5] Although this is still contentious, being based on only few characters and limited taxon sampling, ciliary root patterns and other diverse evidence[2] appear to support this root (see also the chapter by Brinkmann and Philippe).

These relationships are important as they allow us to make evolutionary deductions about traits within eukaryotes. If a trait (morphological or molecular) is found in representatives of most of the super-groups, and in particular in groups on both sides of the eukaryotic root (see Fig. 1), then the trait is ancient and any absences due to derived loss. If, on the other hand, a trait is found in a single group, then it is likely recently derived.

*Corresponding Author—Mark C. Field, The Molteno Building, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, U.K., Email: mcf34@cam.ac.uk

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

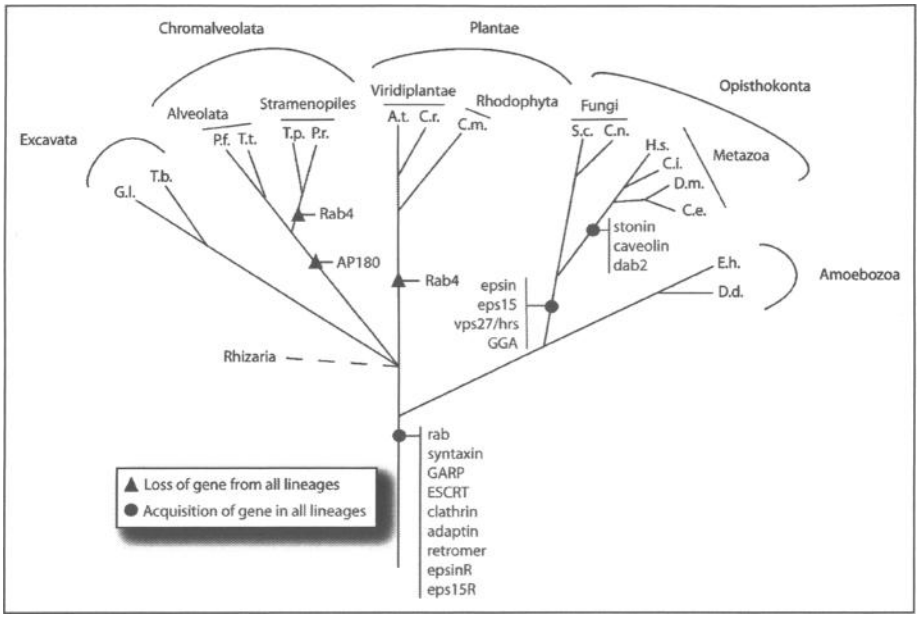


Figure 1. Model for major acquisition and loss of selected intracellular trafficking pathways during eukaryote evolution. The evolutionary relationships of various eukaryotic taxa are shown following the scheme of Simpson and Roger and references therein,[3] as well as evidence from Richards and Cavalier-Smith.[5] Positions of acquisitions and losses of gene families are indicated (green dots and red triangles respectively). Most lineages above a point of acquisition possess the gene, indicating that the last common ancestor of that clade would have had the respective gene, whilst all taxa above a red triangle lack this gene. A number of taxa-specific secondary losses are likely also present, but these have been omitted for simplicity (see Fig. 3). Taxa represented are H.s.; Homo sapiens, D.m.; *Drosophila melanogaster*, C.e.; *Caenorhabditis elegans*, C.i.; *Ciona intestinalis*, S.c; *Saccharomyces cerevisiae*, C.n; *Cryptococcus neoformans*, E.h.,; *Entamoeba histolytica*, D.d.; *Dictyostelium discoidium*, A.t.; *Arabidopsis thaliana*, C.r.; *Chlamydomonas reinhardtii*, C.m.; *Cyanidioschyzon merolae*, P.f.; *Plasmodium falciparum*, T.t.; *Tetrahymena thermophila*, G.l.; *Giardia lamblia*, T.p.; *Thalassiosira pseudonana*, P.r.; *Phytophthora ramorum* and T.b.; *Trypanosoma brucei*. A color version of this figure is available online at www.eurekah.com.

In this article we first consider the overall structures of the endocytic systems of selected model eukaryotes and then survey selected genomes for the presence of key factors involved in endocytosis and associated trafficking pathways. This latter aspect assumes that the mechanisms subtending various transport routes are conserved, and implies that the presence or absence of a particular factor mirrors the presence/absence of a pathway or process. This approach obviously falls short of 100% accuracy, but given high degrees of conservation we have observed in experimental studies of the deeply divergent trypanosomes[6] as well as detailed in silico analysis of restricted gene families across many taxa,[7] we consider this to be an informative strategy.

Defining Endocytosis

Phagocytosis has been observed in organisms on both sides of the eukaryotic root and can therefore be considered an ancient mechanism. Indeed some authors speculate that phagocytosis was the founding innovation and the driving evolutionary force for the evolution of the eukaryotic state.[8] In organisms such as *Paramecium* and amoebae (e.g., *Entamoeba* and *Dictyostellium*) the process serves to supply the cell with nutrients via the ingestion of bacteria

and other organic material. Phagocytic mechanisms are retained in multicellular eukaryotes, including metazoans, where the ability to ingest whole cells has become coopted for specialised functions, including defence against infectious agents as well as for management of programmed cell death or apoptosis; in both of these examples specialised phagocytes, macrophages, are responsible. At the molecular level, phagocytosis is characterised by a dependence on actin and also small GTPases of the Rho subfamily.[9] Beyond noting its obviously critical and ancient nature, however, we will not treat phagocytosis further here, in order to focus on the various and better characterized molecular components of pinocytosis.

Pinocytosis has been reported in the majority of eukaryote taxa where directly investigated and is also therefore an ancient mechanism. There are several types and multiple functions of pinocytotic endocytosis, which includes fluid-phase uptake of the media (which may include dissolved solutes) and receptor-mediated endocytosis (RME). Functions include nutrient uptake, environmental sampling, turnover of surface components and also cell signalling, whilst the various mechanisms can be divided into clathrin-dependant and -independent, reflecting a requirement for the conserved heterodimeric clathrin protein. Further, there are several modes of clathrin-independent endocytosis, at least one of these involves a cholesterol-binding protein caveolin, which can also be differentiated from clathrin-dependant endocytosis based on the morphology of endocytic structures associated with the cell surface membrane. Specifically caveolin is associated with caveolae, whilst clathrin-mediated mechanisms are associated with clathrin-coated pits and vesicles. Additional pathways that require neither clathrin nor caveolin are also present in some cells, but the lack of a marker molecule specific to these modes has precluded detailed investigation of these systems, and they will not be considered further here.

The General Structure and Morphological Evolution of Endocytic Systems

The basic architecture of the endocytic system is shown in Figure 2A. The principle features are (i) multiple routes from the surface, (ii) several functionally differentiated endosomal structures including the early endosome, the recycling endosome and late endosomes, (iii) integration with a degradative pathway, variously termed the lysosome, vacuole or reservosome in different systems, and (iv) close integration with the Golgi complex and the trans-Golgi network in particular, and hence exocytic traffic.[1] In general most systems retain these features, but there are examples of organisms where one or more aspect has been lost.

In some eukaryotic supergroups, endocytic specialization is absent,[10] whether due to a photosynthetic lifestyle and thus presumably reduced endocytic activity (Plantae) or due to the possible selective advantage of amoebic versatility in being able to phagocytose wherever prey may be available (Amoebae). Cercozoans also feed phagocytically by filose pseudopodia and lack a truly specialized endocytic region.[10] Numerous organisms however, demonstrate restriction in the spatial location of their endocytic apparatus. More extreme examples include *Paramecium*, where the phagocytic system is associated with a cytopharynx[11] and trypanosomatids where all endocytic activity is restricted to the flagellar pocket,[12] but restriction of cell surface sites where endocytic activity may occur is also observed in multicellular systems, including metazoa. The selective pressures that underpin such polarisation are likely multiple, and include segregation of function by control of membrane protein/lipid composition (polarised cells in metazoa), feeding efficiency in many excavates, and Chromalveolates, including the model ciliate *Paramecium* and immune evasion (the flagellar pocket in trypanosomes).

In Opisthokonts (i.e., many metazoan cells and in some fungi including *Saccharomyces cerevisiae*), there is limited differentiation of the plasma membrane, with endocytic activity being initiated from most regions of the membrane (Fig. 2D). Microdomains may exist whereby small areas of the membrane are marked for preferential endocytic activity, but specialised membranous structures are, by and large, absent. Polarised cells, for example epithelia and neuronal cell types, are an exception where specific endocytic activity is derived from distinct membrane microdomains.[13] In these examples, barriers to diffusion of surface proteins and

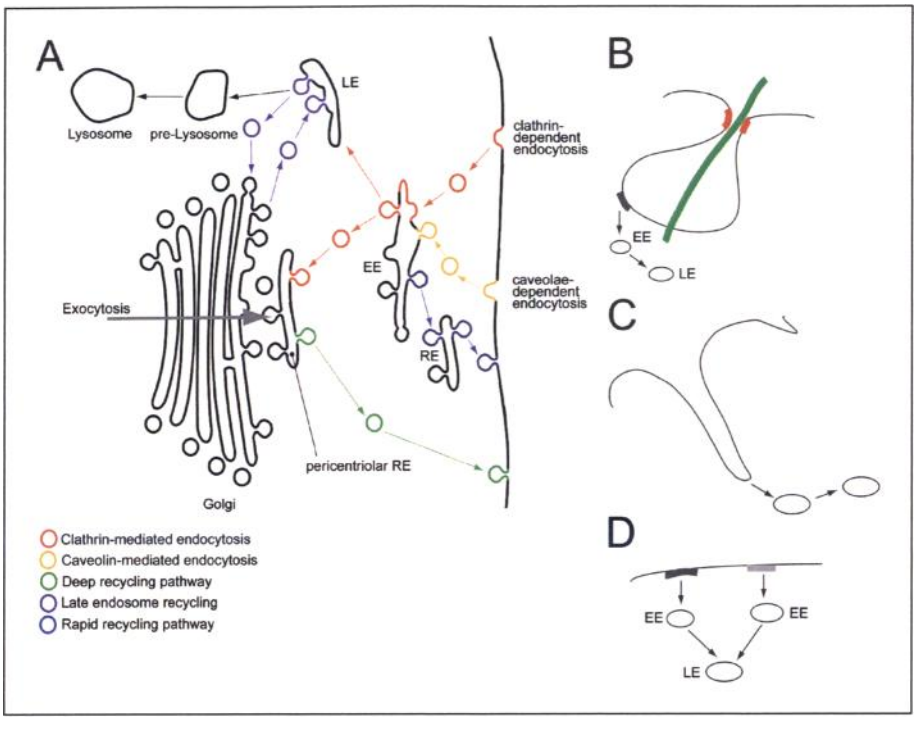


Figure 2. Schematics of arrangements of endocytic systems at the cell surface. A) general endocytic pathway showing the relationship between the Golgi, endosomes, cell surface and the lysosome. EE; early endosome, LE; late endosome, RE; recycling endosome. B) Strictly gated configuration. One (or more) endocytic mode originates from a sequestered area of the membrane. In this example the area of active endocytosis is shown as an invagination, with barriers to diffusion designated in red. Only clathrin mediated mechanisms are shown. The structure in green is a flagellum. Examples of this configuration can be found amongst the trypanosomes, although similar invaginations have been observed in many protists. C) Funnel configuration. A specialised invagination where endocytosis is restricted. In this example it is not clear if there is a strict barrier to diffusion of membrane proteins. A good example of this is the cytopharynx of *Paramecium*. D) Open access configuration. Multiple endocytic mechanisms, as denoted by the dark (clathrin) and light (caveolin) gray bars at the plasma membrane, originate from an essentially undifferentiated membrane. The presence of specific tags on membrane microdomains directing assembly of endocytic structures may serve to differentiate the membrane, but at the time of writing the presence of such tagged sites has yet to be rigorously established. The two distinct endocytic routes are shown converging later within the system, consistent with current knowledge from the mammalian system. Examples of this type of system are found in the metazoa.

lipids serve to maintain a functional division. These latter features are clearly restricted to multicellular organisms, and therefore likely represent specific specialisations that have arisen during evolution of given multicellular lineages.

In *Paramecium*, a well-studied representative of the Chromalveolata, the architecture of the cell is much more specialised. From the perspective of membrane transport, two organelles deserve specific comment; the cytopharynx and the cytoproct. The former is a specialised invagination in the plasma membrane that forms a funnel-like structure into which particles are guided by the beating of cell surface cilia; from the distal end of this organelle rapid phagocytosis takes place (Fig. 2C). Phagocytosis is followed by rapid acidification, presumably to aid both in digestion of prey and also in killing; acidification is achieved by an unusual process that

involves fusion with preexisting acidisomes derived from the Golgi complex.[14] This stands in stark contrast to the acidification of endosomes and phagosomes in the Opisthokonta which involves a bafilomycin-sensitive membrane H+ pump. Further, *Paramecium* phagosomes are cycled through the cell, and material is ultimately expelled from the cell via the cytoproct, an unusual example of exocytosis. In addition, the presence of a large number of trichocysts, essentially dense granule vesicles involved in rapid regulated exocytosis, also requires a rapid endocytic mechanism to recapture exocytosed membrane at the cell surface.[11] Subtending these surface events are several endosomal compartments, many of which appear to have rather unusual morphologies.[11]

In the kinetoplastida, which include free living and parasitic forms, the cell is highly polarised, with an invagination (the flagellar pocket) at one end of the cell where the flagellum enters the cell body and crosses the membrane (Fig. 2B). In some systems, e.g., *Trypanosoma brucei*, all endocytic and exocytic transport is directed towards the flagellar pocket membrane, which functions as a unique membrane domain. Further subdivision of the structure into flagellar pocket and cytostome is seen in *Trypanosoma cruzi*; and may serve to further differentiate the surface as the cytostome appears to be dedicated to endocytic activity.[15] It is speculated that the flagellar pocket plays a role in immune evasion in parasitic protozoa by sequestering immunogenic determinants.[6]

The sister group to kinetoplastids, the euglenids, also have a flagellar pocket-like structure. This, however, is used as an ingestion tubule for feeding, suggesting a more standard endocytic role for the flagellar pocket in the past. Strong molecular sequence data links the euglenozoa (euglenids plus kinetoplastids) to the heterolobosean amoeboflagellates,[16] which have a much more complex ventral feeding groove.[17] The presence of conserved homologous cytoskeletal features underlying the ventral groove links these three groups to other ventral groove possessing taxa, including *Malawimonas*, *Trimastix*, "core" jakobids, *Carpediemonas* and retortamonads. The scenario of a nongroove-possessing organism being linked to one possessing the suite of homologous ultrastructural characters is repeated several times for the excavate taxa.[17] Indeed, the molecular link within the groove containing taxa is the contentious point for this super-group, with no analysis uniting all ten proposed excavate groups into a single clade. However, the nodes separating the groups are generally poorly supported and rates of evolution for the genes used in the analyses vary quite strongly between excavate taxa, possibly explaining the failure to resolve them as a group.[2] Nonetheless, based on the molecular evidence linking kinetoplastids with feeding groove possessing taxa, as well as on the strong morphological evidence, it seems clear that the flagellar pocket is a highly diverged version of the ventral feeding groove. Overall, the specialization of endocytic machinery seems to be a wide-spread, if not always homologous, feature of eukaryotic cells being found in three of the six major eukaryotic super-groups.

Key Factors Involved in Endocytic Systems and their Evolutionary Distribution

Vesicle transport is controlled by a large number of proteins and lipids that interact dynamically to assemble and disassemble specific complexes in a time and location-dependent manner. Some of these factors are general, e.g., NEM-sensitive factor, an ATPase that plays a role in the majority of vesicle fusion events. However, many are specific to individual pathways, including Rab family GTPases, proteins of the SNARE superfamily and many others. Here we consider some of the key factors that are required for endocytosis and the associated recycling pathways, and by probing 17 genomes covering as wide a range of the eukaryotic tree of life as possible attempt to infer the evolutionary distribution of these factors (see results in Fig. 1).

Rabs and Syntaxins

The major endocytic and recycling pathways are modulated by one of four Rab proteins; Rab5 mediates endocytosis itself, Rab4 and Rab11 control rapid and deep recycling pathways respectively, whilst Rab7 regulates delivery to the lysosome/vacuole (Fig. 2A -see figure legend).

Figure 3, viewed on following page. Evolutionary distribution of selected major protein players in endocytosis. The presence or absence of endocytic factors, as assessed by BLAST are shown. Databases were searched in mid-2005 for the presence of a range of diagnostic protein factors with known major roles in a number of important endocytic pathways using translated BLAST. Over 550 independent BLAST analyses were performed, involving genome specific BLAST using authentic query sequences (typically from *H. sapiens* or fungi), followed by a reverse BLAST to the nr database. Whilst such analysis needs to be dealt with the appropriate caution, the presence of a factor is a good indicator that a specific pathway is present. Conversely, absence indicates that a pathway is likely to be absent. For most databases coverage of the relevant genome is of sufficient depth to make prediction reliable; however, several of the taxa discussed here have genome datasets that are not sufficient for confident prediction of absence, whilst even completed genome projects may lack ~1% of ORFs. Note that all databases are considered complete except *Tetrahymena thermophila, Giardia lamblia, Thalassiosira pseudonana, Chlamydomonas reinhardtii* and *Phytophthora ramorum*. Also *Entamoeba histolytica* and *Dictyostelium discoidium* genomes are nominally completed but depth of coverage may not be sufficient to ensure all ORFs are included in the current data sets. Filled symbols: protein:protein BLAST routines retrieve significant hit ($\sim 10^{-10}$), most relevant domains are present, predicted protein is of size consistent with orthology, and reverse BLAST is successful, open symbols: not returned from database searches, significant hit not obtained by BLAST, or inspection of sequence returned indicates nonorthology. We prefer the term "not returned" to "not found" as the possibility remains that an orthologue has been missed by our analysis. Large taxon groupings as are commonly recognised are subdivided into segments for each species within that grouping (if appropriate). Colours used are arbitrary and are for clarity only. Taxa represented are H.s.; *Homo sapiens*, D.m.; *Drosophila melanogaster*, C.e.; *Caenorhabditis elegans*, C.i.; *Ciona intestinalis*, S.c; *Saccharomyces cerevisiae*, C.n; *Cryptococcus neoformans*, E.h.,; *Entamoeba histolytica*, D.d.; *Dictyostelium discoidium*, A.t.; *Arabidopsis thaliana*, C.r.; *Chlamydomonas reinhardtii*, C.m.; *Cyanidioschyzon merolae*, P.f.; *Plasmodium falciparum*, T.t.; *Tetrahymena thermophila*, G.l.; *Giardia lamblia*, T.p.; *Thalassiosira pseudonana*, P.r.; *Phytophthora ramorum* and T.b.; *Trypanosoma brucei*. A color version of this figure is available online at www.eurekah.com.

Rab6 controls retrograde traffic through the Golgi complex. Rab proteins have several roles, including modulation of SNARE function and recruitment of effector proteins to specific vesicles, which in turn serve to propel transport forward - effectors include lipid kinases, components of the cytoskeleton and tethering factors. Syntaxins are coiled-coil transmembrane proteins, of the SNARE superfamily, that can form complexes in trans, i.e., between membranes. They are functionally involved in various stages of the fusion process.[18]

Extending conclusions from previous work,[19-21] we found that both the Rab and syntaxin families are deeply conserved (Fig. 3), and multiple members are present throughout evolution. Rab11 and Rab7 are universal, being found in all taxa sampled; whilst the conserved presence of a terminal endosomal compartment, i.e., lysosome/vacuole, was expected, the complete retention of Rab11, and hence a deep recycling system, was not so obvious. In addition, Rab6 was recovered from all taxa except *G. lamblia*; therefore Rab6, 7 and 11 represent core components of the endocytic system that are likely essential for eukaryotic life. Rab5 was also recovered from most taxa, with the only exceptions being *C. merolae* and again *G. lamblia*. *C. merolae* has a reduced genome and also lives at high acidity (pH1.5). It likely has minimal endocytic activity, albeit with retention of the Rab11/Rab7 pathways.[22] Although the *Giardia* genome database is incomplete and *Giardia* gene sequences are known to be divergent and thus vulnerable to mis-classification by BLAST, our failure to identify various *Giardia* components may be due to true absence. *Giardia* appears to have a minimal endomembrane system, with true early/late endosomes lacking, and a fused endosomal compartment (peripheral vacuole) instead.[23,24] It is clear from multigenome analysis and placement of the eukaryotic root,[2,25] that whilst this configuration has been suggested to represent a basal state,[23] it has likely arisen by secondary simplification. Rab4 is less well retained overall, being lost from representatives of the fungi, metazoa, chromalveolates and also all plantae. Hence the Rab4 pathway, whilst ancient, has been subjected to secondary loss from multiple taxa and the function of Rab4 in recycling is likely, in some circumstances, overshadowed by the Rab11 pathway.[26] The syntaxins

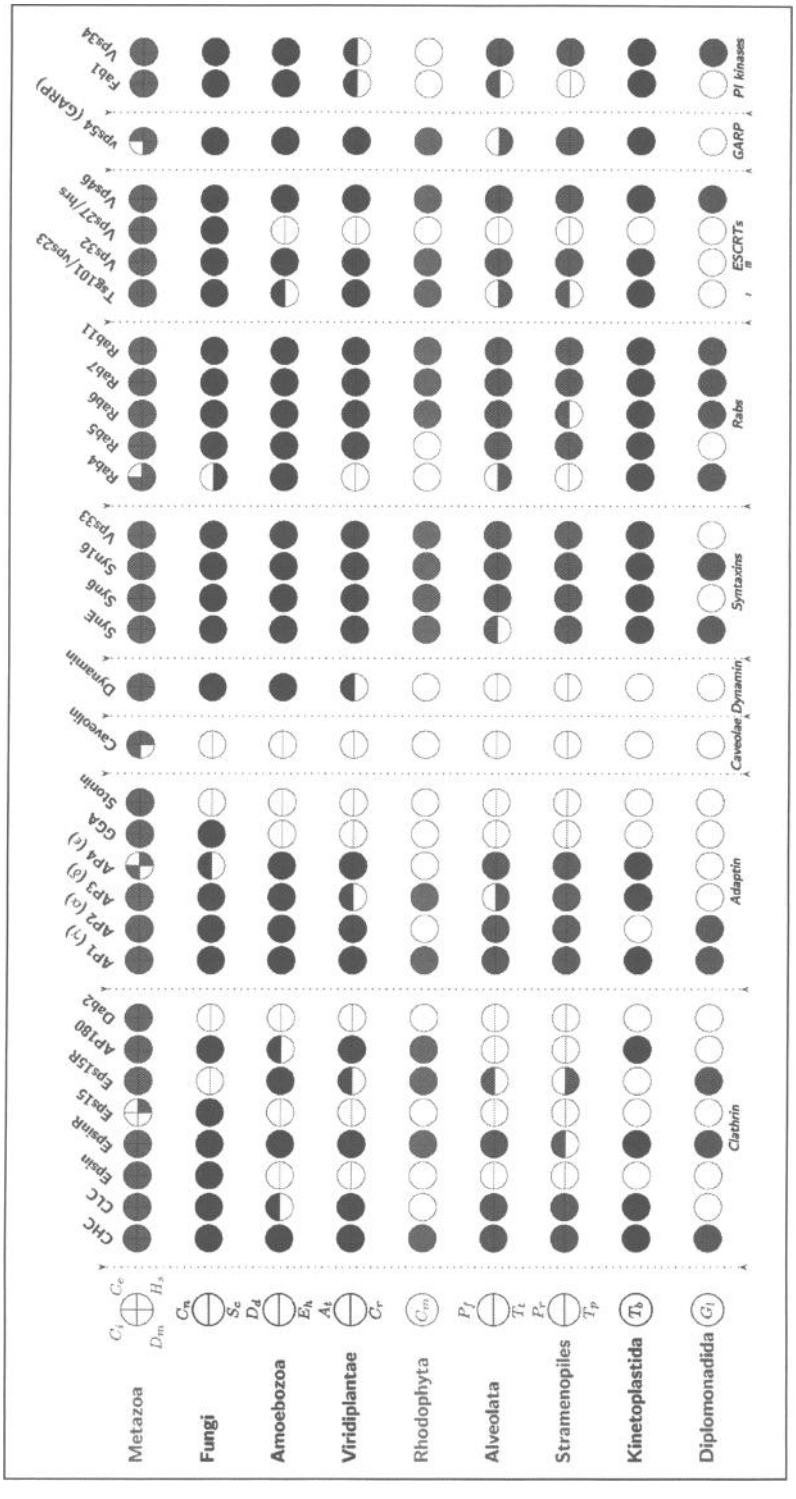


Figure 3. See figure legend on previous page.

are also highly conserved (Fig. 3), with the representatives of endosomal syntaxins (SynE), Syn6 and Syn16 being near universal. Absences of Syn6 in *G. lamblia* and SynE in *T. thermophila*, *C. reinhardtii* and *C. merolae* are explainable by artefact or secondary loss.

Two tethering factor complexes that interact with both Rab and SNARE proteins are worthy of consideration here. The Golgi-associated retrograde protein (GARP) complex is composed of four vps gene products (vps51, 52, 53 and 54) and acts as a tethering complex for Rab6. GARP also links with Syntaxin 10. Its major function appears to be mediating a retrograde pathway from endosomes to the Golgi complex.[27] A second recently described tethering complex, the HOPS complex, is involved in vacuole biogenesis, is an effector for Rab7 in yeast, and contains a vacuole SNARE protein.[28] Therefore, GARP and HOPS appear to share similar mechanisms in their function.

Both of these complexes are ancient (Fig. 1), but retention of GARP through evolution is less strong than for HOPS. Unusual aspects of the Golgi complex structure and function,[29,23] or the unusually high levels of sequence divergence in both *Plasmodium* and *Giardia* could equally account for their apparent lack of GARP. The absence of GARP from *C. intestinalis* was confirmed by unsuccessful searches for additional subunits of the complex; the functional implications of GARP loss in a metazoan genome are unclear. Vps33, a Syntaxin-binding protein and component of the HOPS complex, is universal, except in *Giardia*, where it could have been lost from the fused peripheral vacuole.

Endocytic Coats

The major endocytic coat proteins, clathrin and caveolin function in independent pathways but both operate mechanistically via a cooperative assembly process that subtends the plasma membrane and also includes the concentration of cargo in a local region or domain.[30] Searches for both the light and heavy chains showed that clathrin is a universal component of the eukaryotes; by contrast caveolin, which is primarily involved in endocytosis of lipid-anchored proteins, is only found in metazoa (Fig. 3). Therefore the latter pathway is a recent acquisition and indicates that the trafficking of lipid-anchored proteins is fundamentally different between divergent systems. Significantly this indicates that model systems may not always produce universally applicable generalizations regarding trafficking processes, in contrast to recent suggestions (e.g., 31).

Dynamin and Adaptins

Assembly of clathrin-coated pits at the plasma membrane results in the recruitment of a large number of cofactors, including the large GTPase dynamin and members of the adaptin complex family. The dynamin family is extensive, but not all members are components of membrane transport systems. Those involved in the endocytic system are responsible for vesicle scission from a donor membrane and appear to be defined by the presence of at least three domains, the GTPase, the "middle" region and the GTPase effector domain.[32] A ubiquitous requirement for dynamin in clathrin-dependent endocytosis has been challenged by the discovery of dynamin-independent endocytosis in *T. brucei*.[12] Further, we find the presence of dynamins, retaining the same domain configuration as those proteins documented to have a role in endocytosis in higher eukaryotes, being restricted to the metazoa, fungi, amoebozoa and higher plants and absent from the excavata and chromalveolata (Fig. 3). Based on the current understanding of relationships (Fig. 1), instead of invoking secondary loss, it is slightly more parsimonious to propose that dynamin is a later addition to the endosomal system, whilst a role in mitochondrial membrane biogenesis may be the more primitive function.

The adaptins are heterotetramers that in part are responsible for cargo recognition and retention in clathrin-coated membrane regions. There are four known adaptin complexes, with differing roles; AP-2 is the principal player in endocytosis, but AP-3 and 4 likely also have a role in the recycling and lysosome targeting portions of the endocytic system.[33] We found all adaptins to be widely distributed, indicating an ancient origin for all four complexes (Fig. 3).

Both AP1 and AP2 are universally represented, with the sole absence of AP2 from *T. brucei*, likely due to some novel aspects of endocytosis in that system as the complex is present in related kinetoplastids.[6] AP3 and AP4 are less well retained, and their absence is noted from a wide range of taxa. This is likely due to a dispensable function, dependant on the niche the organism occupies, with such conditions having been met on multiple occasions. Evidence from genetic studies in yeast and other systems support the idea that these adaptins are nonessential and not part of the mechanistic machinery.[34]

Distantly related to the adaptins are the GGAs; these modular proteins contain a domain related to the "ear' of γ-adaptin, function in a pathway modulated by ARF-like GTPases, and also recognise cargo as well as bind to components of the Rab5 system. A further adaptin-related coat complex are the stonins. In metazoa and yeast, the GGAs deliver post-Golgi cargo to the late endosome.[35] Our survey shows GGAs and stonins to be recent acquisitions, with GGAs found in Opisthokonts whilst the stonins are restricted to metazoan systems only.

Epsin and Associated Proteins

Four further proteins of special note participate in the early steps of endocytosis; epsin, eps15, dab2 and AP180. Epsin may be able to function to deform membrane in both the presence and absence of clathrin. It has a modular structure: comprising an epsin N-terminal homology domain, which binds phosphoinositides; ubiquitin-interaction motifs (UIMs); and a flexible region that includes binding sites for clathrin, AP-2 and epsin-homology domain-containing proteins such as Eps15, another UIM-containing endocytic factor. Epsin is implicated in mediating a clathrin-independent endocytic pathway as well as conducting interactions with multiple proteins and membrane lipids involved in the clathrin-dependent route.[36] Eps15 is another clathrin-binding protein that forms part of the major network of proteins that subtend the clathrin coat;[37] eps15 interacts with a host of other factors including the actin cytoskeleton and ubiquitin via UIMs, and data suggests is important for coordination of the endocytic system through interaction with the vesicle uncoating system. Both epsin and esp15 are restricted to metazoa and fungi, but the closely related epsinR (R for "related") and esp15R are widely distributed (Fig. 3). The major difference between epsin/eps15 and epsinR/eps15R is the presence of UIMs in epsin and eps15, but not in epsinR/eps15R, suggesting the recent acquisition of a ubiquitin-dependent aspect to endocytosis by the Opisthokonta.

A further component of this system may also be Dab2; this protein also likely functions as an adaptor, recognising cargo molecules via the presence of specific peptide signals and is present as part of the clathrin coat;[35] Dab2 is metazoa specific (Fig. 3). The final adaptor molecule, AP180, is also implicated in the recognition of membranes in clathrin-dependent transport processes.[38] In contrast to Dab2, we found AP180 to be an ancient component of the endocytic system, but which appears to have been lost from the Chromalveolate lineage (Fig. 1).

PI-Kinases

The control of the endocytic system is mediated in part by the Rab proteins, but is also integrated with the protein and lipid kinase system. Specifically, kinases act as effectors to transmit information through a pathway, frequently this information is derived ultimately from a GTPase such as a Rab protein. In the case of endosomal trafficking, considerable evidence indicates a specific role for phosphatidylinositol (PI) lipid kinases. The most important of these appear to be PI-3 and PI-5 kinases, including the vps34 and Fab1 gene products. Vps34 interacts with Rab7, likely controlling vacuolar/lysosomal targeting,[39] and is widely distributed, but with secondary losses within the plantae (Fig. 2). Fab1, also important for vacuolar transport, is clearly ancient but has been lost from lineages of the Plantae, Chromalveolates and the Excavata (Fig. 2), suggesting a more dispensable function. There is good evidence that, amongst other roles, formation of phosphatidylinositol phosphates generates specific binding sites for a number of other endocytic factors, including epsin and components of the ESCRT system.

ESCRTs

The ESCRT group of complexes mediate delivery of endocytic cargo to the multivesicular body and through to the lysosome. Three ESCRT complexes, each consisting of several orthologues of the yeast vps class E complementation group gene products, have been implicated in lysosomal delivery of receptors, viral proteins and other factors.[40] As shown in Figure 1, the three complexes are also ancient, with some secondary loss of the Tsg101/vps23 gene, a component of ESCRT I. Importantly, associated complexes are responsible for processing and delivery of endocytosed proteins that have been modified by ubiquitin to the multivesicular body (MVB), a late stage in the pathway to the lysosome, via ESCRT-dependant pathways. One factor important in MVB targeting, Vps27/Hrs, is in fact a recent development and is restricted to the fungi and metazoa (Fig. 2). This poses the important question of how proteins are targeted to MVBs in the majority of taxa.

Perspectives

There is ample evidence for the presence of conserved core vesicle fusion machinery throughout the eukaryotic lineage,[7] together with suggestions for retained coat complexes and other highly conserved factors.[6,19-21,41,42] However, many of these factors are involved in diverse pathways. Hence whilst it is no real surprise that these proteins are represented, their presence does serve to underscore the ancient origin of the majority of membrane-trafficking systems. An even more ancient origin of the overall system is suggested by the demonstration that Sec13p in *S. cerevisiae* is a component of both the COP II complex that mediates exit from the endoplasmic reticulum and also of the nuclear pore complex (NPC), implying the presence of a primitive membrane-binding and deforming complex that predates the endomembrane system.[43] The model was further supported by demonstration of considerable secondary structural similarity between several core proteins of the vesicle transport system (clathrin, adaptin, COP I and II) and members of the NPC Nup84p subcomplex.[44] The NPC, which also binds membrane and is responsible for increasing bilayer curvature in a manner akin to vesicle budding and membrane trafficking systems may have a common origin predating the emergence of a true eukaryotic cell. Hence establishment of the basic mechanisms of membrane trafficking is very ancient indeed, and the last common eukaryotic ancestor (LCEA) would have possessed the machinery to carry out such processes. The question then is not how ancient is the general system but, how elaborate had the endocytic system become in the LCEA?

The presence of multiple Rab and syntaxin family members throughout the eukaryota indicates that the LCEA likely possessed a differentiated endosomal system, which included at least two recycling systems depending on Rab4 and Rab11. Secondary losses are common, likely reflecting significant selective pressures on multiple lineages, and indeed Rab4 appears lost on at least two occasions. Hence the endocytic system displays a degree of flexibility and redundancy, but the core components are highly conserved, with the majority of organisms retaining nearly all of the factors we have analysed. Figure 2 illustrates the complex endocytic machinery present at the base of eukaryotes. In addition to the factors surveyed here, we have evidence for the ancient presence of the retrograde endosomal recycling coat complex, retromer (JBD et al, unpublished). The identification of individual components establishes the genomic presence of these genes early on. However, the fact that we can identify multiple members of complexes (e.g., The three ESCRTs, or the GARP plus syntaxin 6) suggests conserved functional interactions and pathways in diverse eukaryotes retained from the LCEA.

Nonetheless, a number of significant transitions are apparent, particularly associated with emergence of the Opisthokonta. For example, the appearance of the epsin/eps15 proteins may be explained by the transfer of the UIM into these proteins, facilitating exploitation of ubiquitination as a mechanism in endocytosis. As well as informing the evolution of the endocytic system, these traits also iterate back into understanding of the relationships amongst eukaryotic groups, representing novel examples of rare genetic characters. The Opisthokont innovations (Epsin, EPS15, GGA and Vps27/hrs) help to further cement the monophyly of this

supergroup, while the loss of AP180 (although weaker due to its being negative data) supports the, sometimes contentious, Chromalveolata.[45] Exploring the evolutionary details of these endocytic components, especially amongst unrepresented members of the unicellular relatives of Opisthokonts such as nuclearid amoebae[46] or other chromalveolate lines such as haptophytes and cryptophytes[45] may help to resolve relationships within the super-groups.

The great range of trafficking pathways in some of the protozoan systems perhaps suggests that there may be truly novel mechanisms at work for membrane transport in these systems. However, the comparatively poor experimental tractability of many of these organisms means that molecular studies are not very advanced and all evidence so far in fact points to a conservation of mechanism despite great novelty at the morphological level. The lack of acquisition of novel endocytic factors in the Excavata, Chromalveolata and Plantae could reflect a true paucity of innovation in these taxa, but an attractive alternative is that the poor current state of understanding of endocytosis in these taxa could explain this absence, and further experimental examination of tractable systems is clearly called for. These observations also impose limits on the validity of a model system, and highlight the need to sample endocytic systems across the eukaryota. For example, experimental data coupled with in silico analysis indicates that tryanosomes are indeed divergent from fungi and mammals and as such cannot be viewed as a model for processes within the Opisthokonta.[6,46] To understand the endocytic system in diverse taxa requires direct experimental work on representative organisms and cannot be inferred. The identification of truly novel factors in trypanosomes, *Giardia*, or any other system, will be an exciting and significant challenge.

Acknowledgements

We are especially grateful to Richard Coulson (EBI) for designing Figure 3. MCF's laboratory is supported by project and program grants from the Wellcome Trust. JD is a Wellcome Trust Postdoctoral Fellow. We are also grateful to the following for open access to genome sequence data that was used in our analysis; the Sanger Institute, The Institute for Genome Research, the Department of Energy (United States) and the *Cyanidioschyzon merolae* genome project.

References

1. Alberts B, Johnson A, Lewis J et al. Molecular biology of the cell. 4th ed. New York: Garland Publishing, 2002.
2. Simpson AG, Roger AJ. Excavata and the origin of amitochondriate eukaryotes. In: Hirt RP, Horner DS, eds. Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics. Boca Raton, Florida, USA: CRC Press, 2004.
3. Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. Curr Biol 2004; 14:R693-6.
4. Stechmann A, Cavalier-Smith T. Rooting the eukaryote tree by using a derived gene fusion. Science 2002; 297(5578):89-91.
5. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436:1113-1118.
6. Field MC, Carrington M. Intracellular membrane transport systems in Trypanosoma brucei. Traffic 2004; 5(12):905-13.
7. Dacks JB, Field MC. Eukaryotic cell evolution from a genomic perspective: The endomembrane system. In: Hirt RP, Horner DS, eds. Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics. London: CRC Press, 2004:309-334.
8. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 2002; 52(Pt 2):297-354.
9. Bokoch GM. Regulation of innate immunity by Rho GTPases. Trends Cell Biol 2005; 15:163-71.
10. In: Lee JJ, Leedale GF et al, eds. The Illustrated Guide to the Protozoa. Lawrence, Kansas: Society of Protozoologists, 2002.
11. Plattner H, Kissmehl R. Molecular aspects of membrane trafficking in paramecium. Int Rev Cytol 2003; 232:185-216.
12. Morgan GW, Hall BS, Denny PW et al. The kinetoplastida endocytic apparatus. Part I: A dynamic system for nutrition and evasion of host defences. Trends Parasitol 2002; 18(11):491-6.

13. Wang E, Pennington JG, Goldenring JR et al. Brefeldin A rapidly disrupts plasma membrane polarity by blocking polar sorting in common endosomes of MDCK cells. J Cell Sci 2001; 114:3309-21.
14. Allen RD, Ma L, Fok AK. Acidosomes: Recipients of multiple sources of membrane and cargo during development and maturation. J Cell Sci 1993; 106(Pt 1):411-22.
15. Porto-Carreiro I, Attias M, Miranda K et al. Trypanosoma cruzi epimastigote endocytic pathway: Cargo enters the cytostome and passes through an early endosomal network before storage in reservosomes. Eur J Cell Biol 2000; 79:858-69.
16. Baldauf SL, Roger AJ, Wenk-Siefert I et al. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 2000; 290(5493):972-7.
17. Simpson AG. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol 2003; 53(Pt 6):1759-77.
18. Teng FY, Wang Y, Tang B. The syntaxins. Genome Biology 2001-2, (reviews 3012.1-3012.7).
19. Dacks JB, Doolittle WF. Novel syntaxin gene sequences from Giardia, Trypanosoma and algae: Implications for the ancient evolution of the eukaryotic endomembrane system. J Cell Sci 2002; 115(Pt 8):1635-42.
20. Lal K, Field MC, Carlton J et al. Identification of a very large Rab GTPase family in the parasitic protozoan Trichomonas vaginalis. Mol Biochem Parasitol; 2005; 143:226-235.
21. Dacks JB, Doolittle WF. Molecular and phylogenetic characterization of syntaxin genes from parasitic protozoa. Mol Biochem Parasitol 2004; 136(2):123-36.
22. DeLuca P, Taddei R, Varano L. Cyanidioschyzon merolae: A new alga of thermal acidic environments. Webbia 1978; 33:37-44.
23. Lujan HD, Touz MC. Protein trafficking in Giardia lamblia. Cell Microbiol 2003; 5:427-34.
24. Hehl AB, Marti M. Secretory protein trafficking in Giardia intestinalis. Mol Microbiol 2004; 53:19-28.
25. Tovar J, Leon-Avila G, Sanchez LB et al. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature 2003; 426(6963):172-6.
26. Hall BS, Pal A, Goulding D et al. Rab4 is an essential regulator of lysosomal trafficking in trypanosomes. J Biol Chem 2004; 279:45047-56.
27. Conibear E, Cleck JN, Stevens TH. Vps51p mediates the association of the GARP (Vps52/53/54) complex with the late Golgi t-SNARE Tlg1p. Mol Biol Cell 2003; 14:1610-23.
28. Collins KM, Thorngren NL, Fratti RA et al. Sec17p and HOPS, in distinct SNARE complexes, mediate SNARE complex disruption or assembly for fusion. EMBO J 2005; 24:1775-1786.
29. Van Wye J, Ghori N, Webster P et al. Identification and localization of rab6, separation of rab6 from ERD2 and implications for an 'unstacked' Golgi, in Plasmodium falciparum. Mol Biochem Parasitol 1996; 83:107-20.
30. Pelkmans L, Burli T, Zerial M et al. Caveolin-stabilized membrane domains as multifunctional transport and sorting devices in endocytic membrane traffic. Cell 2004; 118:767-80.
31. Overath P, Engstler M. Endocytosis, membrane recycling and sorting of GPI-anchored proteins: Trypanosoma brucei as a model system. Mol Microbiol 2004; 53:735-44.
32. Praefcke GJ, McMahon HT. The dynamin superfamily: Universal membrane tubulation and fission molecules? Nat Rev Mol Cell Biol 2004; 5:133-47.
33. Robinson MS. Adaptable adaptors for coated vesicles. Trends Cell Biol 2004; 14(4):167-74.
34. Motley A, Bright NA, Seaman MN et al. Clathrin-mediated endocytosis in AP-2-depleted cells. J Cell Biol 2003; 162(5):909-18.
35. Bonifacino JS, Traub LM. Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu Rev Biochem 2003; 72:395-447.
36. Ford MG, Mills IG, Peter BJ et al. Curvature of clathrin-coated pits driven by epsin. Nature 2002; 419:361-6.
37. Dupre S, Urban-Grimal D, Haguenauer-Tsapis R. Ubiquitin and endocytic internalization in yeast and animal cells. Biochim Biophys Acta 2004; 1695:89-111.
38. Legendre-Guillemin V, Wasiak S, Hussain NK et al. ENTH/ANTH proteins and clathrin-mediated membrane budding. J Cell Sci 2004; 117:9-18.
39. Stein MP, Feng Y, Cooper KL et al. Human VPS34 and p150 are Rab7 interacting partners. Traffic 2003; 4:754-71.
40. Bowers K, Lottridge J, Helliwell SB et al. Protein-protein interactions of ESCRT complexes in the yeast Saccharomyces cerevisiae. Traffic 2004; 5:194-210.
41. Robibaro B, Hoppe HC, Yang M et al. Endocytosis in different lifestyles of protozoan parasitism: Role in nutrient uptake with special reference to Toxoplasma gondii. Int J Parasitol 2001; 31:1343-53.

42. Jurgens G. Membrane trafficking in plants. Annu Rev Cell Dev Biol 2004; 20:481-504.
43. Siniossoglou S, Lutzmann M, Santos-Rosa H et al. Structure and assembly of the Nup84p complex. J Cell Biol 2000; 149:41-54.
44. Devos D, Dokudovskaya S, Alber F et al. Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol 2004; 2:e380.
45. Harper JT, Waanders E, Keeling PJ. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int J Syst Evol Microbiol 2005; 55(Pt 1):487-96.
46. Berriman M, Ghedin E, Hertz-Fowler C et al. The genome of the African trypanosome Trypanosoma brucei. Science 2005; 309(5733):416-22.
47. Lang BF, O'Kelly C, Nerad T et al. The closest unicellular relatives of animals. Curr Biol 2002; 12(20):1773-8.

CHAPTER 8

Origins and Evolution of the Actin Cytoskeleton

Francisco Rivero* and Fatima Cvrčková

Introduction

The presence of a complex cytoskeletal system is a hallmark feature of eukaryotic cells, distinguishing them from their prokaryotic (bacterial or archaeal) "cousins". No extant prokaryote studied so far possesses obvious homologues of major cytoskeletal proteins shared universally among eukaryotes, such as e.g., actin or tubulin. However, several proteins exhibiting limited sequence similarity with certain cytoskeletal components, as well as the ability to form filaments, have been found.[1-3] These include, among others, relatives of actin and actin-associated proteins that will be discussed in detail below, the FtsZ family of bacterial and archaeal tubulin-related proteins participating in cell division[4] and an intermediate filament-like protein (crescentin) from *Caulobacter*.[5]

Evidence from large-scale genome sequencing indicates that eukaryotic nuclear genomes arose as a result of a "merger" of at least two ancestors.[6-8] There are even two subclasses of genomic DNA with different compositional characteristics; one of them, resembling recent Bacteria, has contributed e.g., the molecular apparatus of the energy metabolism, while the other, close to Archaea, brought, among other genes, components of the replication and proteosynthetic machinery.[9] The cytoskeleton has been also suggested as a candidate for archaeal heritage,[10] however, its origins remain mysterious. It has even been proposed that genes for cytoskeletal proteins come from a third ancestor of eukaryotes, a hypothetical "chronocyte", long extinct and surviving only through its eukaryotic descendants.[11] In any case, a cytoskeleton-like apparatus must have been present at least at the point of acquisition of endosymbionts that later gave rise to mitochondria.

Here we focus on a subset of cytoskeletal proteins, namely actin and a core of associated proteins participating in the control of actin dynamics, especially filament nucleation (reviewed in ref. 12). Identification of components shared by evolutionarily distant eukaryotic lineages (such as plants, yeast, Metazoa, slime molds and other selected protists), and in a few cases also the discovery of related proteins in prokaryotes, may provide the first step towards reconstructing the composition, and possibly also functional characteristics, of the initial set of "actin-associated modules" of the common ancestor of eukaryotes (Table 1).

We shall discuss selected parts of the actin-associated apparatus separately. First, we focus on the actin monomer itself, together with a class of evolutionarily conserved monomer-binding proteins that modulate the balance of monomeric and filamentous actin. Next, we will examine complexes that serve as "primers" nucleating new actin filaments. We shall then move to an assortment of actin-binding proteins that either regulate filament dynamics or mediate

*Corresponding Author: Francisco Rivero—Center for Biochemistry and Center for Molecular Medicine Cologne, Medical Faculty, University of Cologne, Joseph-Stelzmann-Str. 52, 50931 Köln, Germany. Email: francisco.rivero@uni-koeln.de

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

Table 1. Evolutionary distribution of ancient domains related to the eukaryotic microfilament system and their prokaryotic relatives

						Chromaveolata		Excavata			
Domain	**InterPro Number**	**Metazoa**	**Fungi**	**Amoebozoa**	**Plantae**	**Ciliophora**	**Apicomplexa**	**Kinetoplastida**	**Parabasalia**	**Diplomonadida**	**Prokaryotes**
Actin and ARPs	IPR004000	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y[a]
ARPC1 (Arc40)	IPR011046[b]	Y	Y	Y	Y	Y	Y	Y	N	(Y)	Y
ARPC2 (Arc35)	IPR007188	Y	Y	Y	Y	Y	N	Y	N	N	N
ARPC4 (Arc19)	IPR008384	Y	Y	Y	Y	Y	Y	Y	N	N	N
ARPC3 (Arc18)	IPR007204	Y	Y	Y	Y	Y	N	Y	N	N	N
ARPC5 (Arc15)	IPR006789	Y	Y	Y	Y	(N)	N	Y	(N)	N	N
FH2	IPR003104	Y	Y	Y	Y	Y	Y	Y	(N)	N	N
Profilin	IPR002097	Y	Y	Y	Y	Y	Y	Y	(Y)	N	Y[c]
ADF	IPR002108	Y	Y	Y	Y	Y	Y	Y	Y	N	N
GEL[d]	IPR007122	Y	Y	Y	Y	(N)	N	N	(N)	N	N
Capping protein	IPR001698	Y	Y	Y	Y	Y	Y	Y	Y	N	N
CH[e]	IPR001715	Y	Y	Y	Y	Y	N	N	Y	N	N
WH2	IPR003124	Y	Y	Y	Y	Y	N	N	(N)	N	Y[f]
VHP	IPR003128	Y	Y	Y	Y	Y	N	N	(N)	N	Y[f]
I/LWEQ	IPR002558	Y	Y	Y	N	(N)	N	N	(N)	N	N
MYO	IPR001609	Y	Y	Y	Y	Y	Y	Y	(Y)	N	Y[g]

Continued on next page

Table 1. Continued

The table is based on published data and on extensive searches of the GenBank non-redundant and EST databases, as well as specific databases for those species of Protists for which sequencing projects are available (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). Following species of Protists were considered: *Dictyostelium discoideum* and *Entamoeba histolytica* for the Amoebozoa; *Tetrahymena spp.* and *Paramecium spp.* for the Ciliophora; *Plasmodium spp.*, *Toxoplasma gondii* and *Cryptosporidium parvum* for the Apicomplexa; *Leishmania major* and *Trypanosoma spp.* for the Kinetoplastida; *Trichomonas vaginalis* for the Parabasalia and *Giardia lamblia* for the Diplomonadida. Note the highly simplified repertoire of microfilament system proteins of *G. lamblia*, caused presumably by secondary gene losses. Y, present in at least one species of the lineage; N, absent; (N), not found in the databases but presence cannot be ruled out because of unfinished sequencing projects or small and/or highly divergent domains that might have escaped identification while searching; (Y) marginal significance or short fragment only. Abbreviations: ADF, actin depolymerizing factor; ARP(C), actin related protein (complex); CH, calponin homology domain; FH2, formin homology domain 2; GEL, gelsolin repeat; I/LWEQ, the actin binding domain of talin and related proteins; VHP, villin headpiece; WH2, WASP homology domain 2; MYO, myosin motor domain.

Notes:

[a]Closest prokaryotic relatives of actin are proteins of the MreB/MRL (IPR004753) family (see text).

[b]ARPC1 proteins form a branch within the broader WD40-like protein family (IPR011046) that has also bacterial and archaeal members.

[c]Closest prokaryotic relatives are the MglB proteins, which are related to members of the Roadblock/LC7 (IPR004942) family (see text and ref. 37).

[d]Only the canonical gelsolin repeat has been considered. The more divergent gelsolin repeat of Sec23/24 is present in all eukaryotes, but there is no proof that this proteins interact with actin (see text for details).

[e]Only the CH domains of the type CH1-CH2 of the α-actinin family and of fimbrin have been considered. CH domains of other types are present in all eukaryotes. CH domains of the type CHc characteristic of choline/carnitine-O-acyltransferases (IPR000542) are found in bacteria of the genera *Nocardia, Mycoplasma* and *Photobacterium*, which may have acquired the protein by horizontal transfer from their vertebrate hosts. Also in *Leishmania* and *Trypanosoma* the protein may have originated from their metazoan hosts or vectors.

[f]The Interpro resource lists a few bacterial proteins with either VHP (*Methanosarcin, Mycoplasma, Legionella*) or WH2 domains (*E. coli, Vibrio, Rickettsia, Wolbachia, Parachlamydia*), and even some viral proteins with WH2 domains, but in the absence of functional data these results must be interpreted very cautiously.

[g] Myosin is a member of the P-loop ATPase/GTPases (IPR001687) superfamily that has also numerous prokaryotic members.

association of other cellular structures with actin filaments, including also actin-dependent motors. Finally, we shall summarize the potential evolutionarily conserved aspects of the regulatory mechanisms controlling the structure and function of the actin network.

The Actin Cytoskeleton in the Cellular Context

While much of actin's fame derives from studies of metazoan muscle actin/myosin complex, nonmuscle actin participates in a range of essential processes of eukaryotic cell morphogenesis, in motility of (nonmuscle) metazoan and amoeboid cells and in intracellular transport. Actin fibers contribute to intracellular movement either by providing direct locomotive force through filament assembly, or serving as "tracks" along which structures travel, driven by molecular motors such as myosins. The polymerization-based mechanisms are believed to be evolutionarily older than those involving molecular motors,[13] thus justifying our focus on the actin nucleation machinery.

Actin filament assembly is believed to participate in the "exploratory behavior" of soft-bodied cells (typical metazoan cells, such as fibroblasts or neurons, or amoeboid cells), i.e., to the formation of filopodia and lamellipodia, as well as membrane ruffles.[14,15] Actin "comets" can also propel organelles and intracellular parasites across the cytoplasm, utilizing filament assembly forces.[16,17] Even in wall-encased cells of plants or fungi filament assembly contributes to cell shape development, as documented for yeast buds,[18] plant trichomes,[19] tip-growing root hairs[20,21] and pollen tubes.[22] However, the resulting networks of actin filaments are believed to serve mainly as tracks for motor-driven delivery of exocytotic vesicles to the expanding regions of the cell surface (e.g., see ref. 23). Perhaps with the exception of trichomes, even these cases can be considered examples of "exploratory behavior", as nonmotile cells indeed can explore the environment only by expanding (growing) into it.

Actin is also indispensable for essential processes of the cell cycle. Cytokinesis usually involves exocytosis, depending on vesicle delivery along actin tracks, and at least in metazoan and amoeboid cells also constriction of a subcortical actomyosin ring.[24,25] Very recently, actin has been also implicated in chromosome congression during oocyte mitosis.[26] Thus, actin is central to at least two basic functions of life, namely "tactile" interaction with the environment and cell multiplication.

Actin- and Monomer-Binding Proteins

Actin exists in the cytoplasm of eukaryotic cells either as soluble monomers (G-actin) or as filaments (F-actin). The in vivo balance between F- and G-actin, as well as filament turnover, is to a large extent controlled by proteins that regulate the availability and nucleotide-bound state of monomers. This, in turn, affects the rate of filament polymerization (predominantly at plus or barbed ends) and depolymerization (preferentially at minus or pointed ends), resulting in net growth, shrinkage or treadmilling. While some modulators of actin dynamics are restricted only to certain eukaryotic lineages, such as e.g., the actin-sequestering proteins β-thymosin of Metazoa (see discussion of WH2 domains below) or toxofilin of *Toxoplasma*, others are shared by most eukaryotes and therefore obviously ancient, in particular profilin and proteins of the actin depolymerizing factor (ADF)/cofilin family, which together act as major determinants of cytoplasmic actin dynamics.[27,28]

Actin

Actin is a remarkably conserved protein, with overall identity about 85% between the most divergent family members. However, most species possess multiple actin isoforms exhibiting both structural and functional differences. The number of actin genes per genome ranges from 1 (in yeasts) to almost 100 in some plants (reviewed in refs. 12,29). The closest eukaryotic relatives of actin comprise the family of actin-related proteins (ARPs),[30] some of which will be discussed below. The evolutionary separation between actin *sensu stricto* and the

ARPs apparently occurred already in the ancestral eukaryote. The actin/ARP gene family can thus be viewed as a single unit if we are considering the early steps of eukaryotic evolution.

Distant but undisputable prokaryotic homologues of actin have been found. The MreB protein forms filaments both inside cells of rod-like bacteria and in vitro.[1,31,32] Both actin and MreB are considered members of an ancient superfamily including also the bacterial cell division protein FtsA, the ATPase domain of Hsp70 and even the enzyme hexokinase. The split between the MreB and actin lineages and the rest of the family apparently occurred very early in evolution,[1,2] presumably before the establishment of molecular mechanisms guaranteeing a relatively low mutation rate, and possibly already in the hypothetical era of the RNA-based life.[11] Most prokaryotic members of the MreB family are bacterial, with just a handful of homologues identified in Archaea (ref. 10 and our database searches). Although this could reflect limited availability of data from Archaea, the possibility that actin's ancestors originated in Bacteria and/or arrived into the proposed archaeal ancestor by horizontal gene transfer also cannot be excluded.

Profilin

Profilin, an abundant small protein that may aid "charging" of G-actin by ATP, promoting thereby filament assembly, is ubiquitous in eukaryotes, with the exception of greatly reduced parasites such as *Giardia*. Like actin, it exists in multiple isoforms in many organisms.[33,34] However, it is somewhat less conserved: profilins within a single organism (such as *Dictyostelium*), may share as little as 55% identity.[35] This also complicates searches for possible prokaryotic relatives.

Profilin is related to members of an ancient family including the Roadblock/LC7-related proteins, as well as prokaryotic (both bacterial and archeal) MglB proteins.[36,37] This family contains a number of proteins implicated in ATPase or GTPase regulation, including dynein light chains (i.e., a subunit of a tubulin-dependent motor complex). MglB is implicated in gliding motility of microbial cells. Curiously, in some bacteria it resides in the MglAB operon that includes also MglA, a small GTPase,[36] suggesting a very ancient relationship between G-proteins and ancestors of the actin module.

ADF/Cofilin and the Gelsolin Repeat

Members of the actin depolymerization factor (ADF)/cofilin family preferentially bind ADP-associated G- and/or F-actin and increase the G-actin level via filament severing and depolymerization. The ADF/cofilin family, defined by the presence of the structurally conserved ADF homology domain, is ubiquitous or at least widespread in eukaryotes,[38] suggesting that this domain and the associated severing and depolymerizing activity was present in the common ancestor of eukaryotes. However, members of the ADF family are relatively poorly conserved, and no candidate prokaryotic relatives have been found so far. Bikonts possess only proteins with a single ADF domain, usually without extensions, but the domain appears duplicated or with characteristic extensions in unikonts (Fig. 1).

Interestingly, the ADF domain shares structural similarity with the gelsolin repeat (including the actin binding site), despite lack of sequence similarity.[39,40] The ADF and the gelsolin repeat might thus be descendants of an ancient actin-severing protein (see Fig. 1 for a detailed description). The gelsolin repeat is found as a tandem of three or more copies in diverse proteins. The prototype of this family, gelsolin, contains six repeats and acts as a calcium-regulated protein that caps the barbed ends of actin filaments, promotes nucleation and severs existing filaments.[41] Typical gelsolin repeat proteins are present in bikonts and in plants (Table 1), indicating that a gelsolin-like protein must have been present in the common ancestor of eukaryotes. Their absence in other lineages is thus apparently due to secondary loss. Moreover, proteins of the Sec23 family, conserved in all eukaryotes, contain a single diverged gelsolin repeat at their C-terminus, supporting the ancient status of the ADF/gelsolin domain. Sec23 is a component of COPII coated vesicles involved in recruiting and formation of prebudding complexes whose interaction with actin has not been investigated so far.[42]

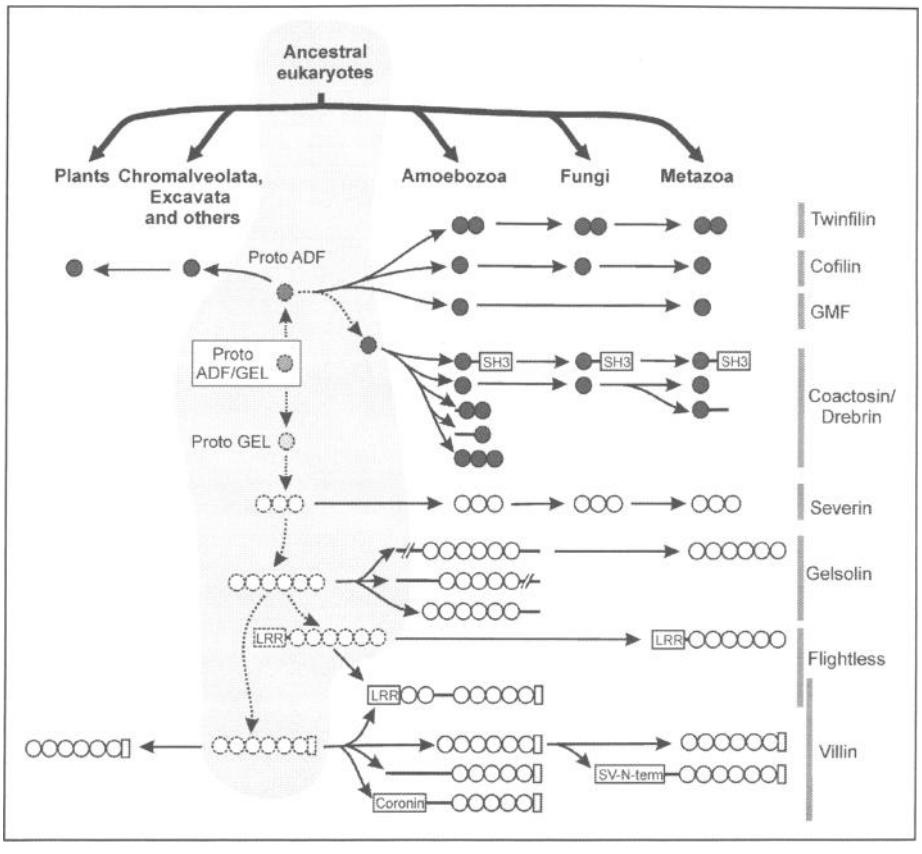


Figure 1. Reconstruction of the evolutionary history of proteins of the ADF/cofilin and gelsolin repeat families. The diagram is based on protein repertoires of selected representatives of each lineage: *A. thaliana, P. falciparum, D. discoideum, S. cerevisiae* (plus a severin-like protein of filamentous fungi absent in yeast) and *H. sapiens*. The eukaryotic genealogy is based on ref. 56. Proteins occurring in multiple isoforms are depicted only once. Based on structural similarity, the ADF domain and the gelsolin repeat might be descendants of an ancient actin-severing protein (proto ADF/GEL). Subsequent divergence, duplication and shuffling of additional domains and extensions gave rise to the diversity of ADF and GEL proteins of extant organisms, particularly unikonts. The ADF domain is duplicated in twinfilin, a protein that binds only G-actin and does not promote actin filament depolymerization; the duplication occurred apparently early after branching of eukaryotes into unikonts and bikonts. Around the same time, the coactosin/drebrin and the GMF (glia maturation factor) subfamilies arose. Its members bind only to F-actin and sometimes possess a characteristic C-terminal extension, such as the SH3 domain of Abp1. A very ancient multiplication of the ancestral GEL repeat gave rise to proteins with three gelsolin repeats such as those of *D. discoideum* (severin), filamentous fungi and Metazoa (CapG). It is intriguing that no proteins with a single (with the exception of Sec23/24 not depicted here; see text) or double repeat have subsisted. A further duplication generated proteins with 6 repeats (such as gelsolin), which then acquired extensions and additional domains, like leucine-rich repeats (LRR) (flightless of Metazoa) or a C-terminal villin headpiece (white rectangle) that conferred actin-bundling properties to some of them (villin and related proteins). In Amoebozoa a flightless-related protein resulted from the fusion of part of a flightess gene to a villin gene, whereas villidin resulted from the fusion of a coronin gene to a villin gene. In Amoebozoa also, frequently the first GEL domain of the tandem of six has been lost, and other domains have eroded so that they are difficult to recognize.

Actin Nucleation Complexes

Actin can form filaments in vitro even in the absence of other proteins. However, elongation of preexisting filaments at extendable ends (i.e., usually free barbed ends) is more efficient and requires lower actin concentration than establishment of new filaments out of G-actin. Such ends can arise either by fragmentation of existing filaments (mediated e.g., by proteins of the ADF/cofilin family), by removal of proteins capping a preexisting end, or by de novo nucleation aided by a specific protein complex. Nucleation of new filaments in vivo may present a unique regulatory node where multiple signaling pathways converge, resulting in precise control of the actin network structure. So far, two independent nucleation mechanisms have been studied in detail – Arp2/3 mediated nucleation and nucleation mediated by formins (for a review see refs. 12,43).

The Arp2/3 Complex

The Arp2/3-dependent nucleation complex consists of seven subunits (Arp2, Arp3, and ARPC1 to ARPC5). Arp2 and Arp3 are members of the ARP family mentioned above. All subunits are well-conserved throughout major eukaryotic lineages, although some losses apparently occurred, in particular in Parabasalia and Diplomonadida[44,45] (Table1). ARPC1 is a member of the WD40-like protein family that has also bacterial and archaeal members but no prokaryotic relatives have been found for the remaining four subunits.

Curiously, at least in some organisms, loss of certain subunits is compatible with survival. In *Arabidopsis*, homozygous mutants lacking single genes Arp2, Arp3, ARPC5 or one out of two isoforms of ARPC2, are viable and fertile, although they exhibit a distinct mutant phenotype of distorted trichomes and malformed epidermal cells.[19,46] For the ARPs themselves this may reflect a partial complementation of the mutant defect by other members of the family or even actin; however, the nonessentiality of ARPC5, together with its absence in several eukaryotic lineages suggests that this subunit may be a later addition in the Arp2/3 nucleation complex. Alternatively, most of the tasks requiring actin nucleation may have been taken over by formins in plants (see below). However, deletion of genes encoding Arp2, Arp3, ARPC1 and ARPC5 is lethal in the budding yeast, while loss of ARPC 2, 3 and 4 results in growth defects of varying severity[47] (see also www.yeastgenome.org); perhaps the ARPC5 subunit might have become indispensable in the specific context of the yeast cell, since budding heavily depends on establishment of Arp2/3-dependent actin structures.[18]

Formins

Formins are defined by the presence of the approximately 400 residues long, predominantly α-helical FH2 domain, capable to form a ring-shaped flexible dimer that caps the barbed end and allows processive elongation of the actin filament. The FH2 domain can be found in most eukaryotes (Table 1) and is usually preceded by the proline-rich FH1 domain that interacts with profilin-actin and funnels actin monomers to the nucleation site. The FH1-FH2 combination probably constitutes the minimal core fully functional in terms of actin nucleation and elongation activity; diverse formins differ in their comparative capping vs. nucleating activities, as well as in their requirements for cofactors such as profilin.[48-51] A plant formin has been recently found to possess an unique ability to bundle actin filaments.[52]

The FH1-FH2 core is usually accompanied by additional domains. Diverse formin classes differ mostly in their N-terminal regions, which generally have regulatory and targeting roles.[53-55] A common architecture characterized by the presence of an N-terminal GTPase binding motif (GBD/FH3) and a C-terminal autoinhibition domain can be found among formins of Amoebozoa, Fungi and Metazoa.[54] This domain combination, allowing regulation of formin activity by activated Rho GTPases, and establishing thereby a direct link between the nucleation machinery and regulatory signaling pathways, apparently arose only within the unikont lineage,[56] since it can be found neither in the formins of Apicomplexa, Kinetoplastida and Ciliophora, nor in those of plants, which acquired either N-terminal membrane anchors or a variant Pten-like domain, possibly also allowing association to membranes.[53]

Other Actin-Binding Proteins

In addition to the predominantly or exclusively G-actin binding proteins, there is a growing list of proteins that bind to actin filaments and exert actions as diverse as capping, severing, crosslinking, attachment to other cellular structures and force transmission. A number of domains with bona fide actin-binding properties that are shared by many proteins throughout various lineages has been identified; duplications and domain shuffling apparently produced much of the present diversity of ABPs. We shall briefly discuss several representatives: the heterodimeric capping protein, the calponin homology (CH) domain and the small VHP and WH2 domains.

The Heterodimeric Capping Protein

Capping proteins bind tightly to the barbed end of actin filaments and prevent the addition or loss of actin subunits. They are composed of two subunits, an α subunit of 32-36 kDa and a β subunit of 28-32 kDa. Interestingly, the α and β subunits of chicken skeletal muscle capping protein have a strikingly similar folding, despite lacking sequence similarity, so that the entire molecule has a pseudo 2-fold rotational symmetry.[57] We suggest that the capping protein was initially homodimeric, but a gene duplication followed by substantial divergence resulted in a heterodimeric protein. The heterodimer must have brought a significant selective advantage, because it has undergone relatively little change since the initial diversification of the two subunits. This event must have taken place very early, because the heterodimeric capping protein can be found in all lineages studied, except in the greatly simplified parasitic diplomonads (Table 1).

The Calponin Homology Domain

The CH domain is a module of about 100 residues with a globular α-helical fold, present in a large family of proteins that can be subdivided into subfamilies according to their domain composition.[58] Not all CH domains bind to actin. A typical CH-containing actin-binding motif consists of a tandem pair of CH1 and CH2 domains, although a sole CH1 domain also can bind to actin. CH1-CH2 proteins are numerous in unikonts, but there are a few examples in bikonts (Table 1). Variant domains such as CH3, CHe and CHc apparently do not bind actin: the CH3 domain is found in many signaling and a few cytoskeleton proteins of many, predominantly multicellular, lineages; CHe binds microtubules and is ubiquitous in eukaryotes;[59] the CHc domain is found in choline/carnitine-O-acyltransferases, enzymes involved in fatty acid metabolism and transport found in Metazoa, Fungi and Kinetoplastida.[60]

The distribution of CH domains suggests that they were present already in the ancestral eukaryote. We postulate that the proto-CH domain did not bind actin and diversified already very early. An ancient duplication gave rise to the tandem CH1-CH2. The actin binding properties could have appeared either before or after this duplication, and the tandem arrangement was clearly advantageous in terms of interaction with F-actin because few proteins have either CH1 or CH2 only; these could have originated by subsequent loss of one of the domains. A further duplication of the CH1-CH2 tandem produced the actin-bundling protein fimbrin, documented in lineages as diverse as unikonts, plants (where they constitute the only CH1-CH2 proteins) and the chromalveolate *Tetrahymena thermophila*.[61] In unikonts the CH1-CH2 family expanded and diversified considerably by domain shuffling, leading to acquisition of tail regions composed of spectrin repeats, filamin repeats and other domains.[58]

The VHP and WH2 Domains

Actin-binding modules such as the villin headpiece (VHP) and the WASP homology domain 2 (WH2) are well documented mainly in multicellular eukaryotes (Table1). However, their small size (about 35 residues) hampers reliable database searches, therefore their presence in other eukaryotes cannot be ruled out. Also the alleged bacterial VHP or WH2 domains, as well as viral proteins with WH2 domains (listed in the InterPro resource[62]) have to be interpreted very cautiously in the absence of functional data.

The VHP appears at the extreme C-terminus of diverse proteins, alone or in combination with other domains, particularly the gelsolin repeat. The VHP binds to F-actin (although there are exceptions) and confers actin-bundling properties to villin and related proteins.[63] The WH2 domain binds preferentially ATP-associated G-actin. It can be found alone as a single domain (β-thymosins) or as a tandem of two (actobindin), three (ciboulot) or four (spire) copies. Isolated WH2 domains such as in β-thymosins sequester G-actin and maintain it in a nonpolymerizable form. In contrast, two or three WH2 domains in tandem as in actobindin or ciboulot promote elongation of barbed filament ends similar to profilin,[64] while the four domains of spire together promote nucleation of new filaments, independent of "classical" nucleation complexes.[65] The WH2 domain may also associate with other domains, as in WASP, verprolin-related proteins and cyclase-associated protein.[66]

A similarity between the actin-binding regions of the VHP and the WH2 domain has been proposed.[39] Both domains might have evolved from a short domain or loop that diverged before the unikont/bikont split to accommodate the different properties, F-actin binding vs. G-actin binding, of the VHP and the WH2 domain, respectively. Nevertheless, it cannot be ruled out that the apparent sequence similarity results from evolutionary convergence.

Other Domains and Proteins

The number of proteins able to interact with actin is very large, and we cannot make a comprehensive account on the evolutionary history of all of them. We will not discuss domains whose presence in a protein does not automatically correlate with actin-binding properties, although they are shared by numerous actin-binding proteins (such as the WD repeat, the kelch repeat and the LIM domain), proteins apparently specific for a limited number of lineages, or those lacking reliably recognizable domains. Future structural studies may reveal relationships among proteins that have passed unnoticed because of apparent absence of sequence similarity. This already happened e.g., in case of the β-trefoil fold, initially described in fascin but later discovered in the *Dictyostelium* membrane-associated protein hisactophilin on the basis of structural data. It has been proposed that the β-trefoil fold arose by duplication of an ancestral gene encoding a homotrimeric single-repeat protein.[67] Similarly, the I/LWEQ domain, named after the conserved initial residues in each of four repeated blocks, might have originated by duplications of an ancestral single-repeat protein. This domain is characteristic of two classes of proteins involved in actin organization, namely the focal adhesion protein talin (Amoebozoa, Metazoa) and the polarisome protein Sla2p/HIP-1 (Amoebozoa, Fungi and Metazoa), and apparently originated within the unikont lineage, prior to the branching between Amoebozoa and Opisthokonta.[68]

Effectors and Regulators of the Actin Cytoskeleton

The actin-binding proteins discussed in the last section present only a limited selection of molecules that serve as "interfaces" ensuring the integration of the actin network into the cellular web of interactions. Thorough analysis of all proteins that could be considered "effectors" of the actin cytoskeleton, as well as of its regulatory inputs, would exceed the scope of this review. Instead, we shall introduce selected examples that can provide additional insights into the early stages of cytoskeleton evolution. First, we shall focus on the major actin-dependent motor protein, myosin, which is responsible for the movement of a variety of cargoes along the actin network; later we shall discuss the evolutionarily conserved aspects of regulatory pathways controlling the structure and dynamics of the actin cytoskeleton.

Myosin, the Prototype Motor

In terms of sequence, myosins can be recognized by the presence of an ancient, well-conserved domain, the myosin head, an ATPase capable of converting the chemical energy from ATP hydrolysis into mechanical movement along an actin filament. The myosin head has been so far found in nearly all eukaryotes studied (Table1), indicating that, similar to actin itself, it apparently arose no later than in the common eukaryotic ancestor. Indeed, it has

been proposed that the myosin motor domain might have originated from a common ancestor with the microtubule-dependent motor kinesin, as they share a similar 3D structure of the core.[69] Although no readily identifiable homologue of the myosin head can be found in prokaryotes, both myosin and kinesin motor domains are related to proteins of the P-loop NTPase superfamily. This superfamily includes both ATPases and GTPases and has also prokaryotic members, suggesting the possible evolutionary root of both motor domains.[70,71]

Myosins have blossomed into an abundant and diverse protein family during eukaryote evolution. At least 18 myosin classes have been established on the basis of both myosin head sequence and overall domain composition;[72,73] however, a recent detailed analysis of 23 genomes covering the whole eukaryotic kingdom distinguishes already 37 myosin classes, with representatives of up to 13 classes found in a single species.[74]

Evolutionary events documented in myosin evolution include mutations, domain shuffling, domain fusions, partial deletions, duplications, and losses, which makes evolutionary studies extremely difficult. Nevertheless, a thorough phylogenetic analysis, based on domain structure rather than on sequence analysis only, not only provided supportive evidence for the unikont/bikont model of early eukaryote evolution, but also allowed identification of three supposed ancestral myosin families.[74] One of them corresponds to myosin I, previously suggested to be one of the oldest myosin classes whose members may have originally functioned as generalists, while more recently evolved families, limited to particular lineages, may have been optimized for specialized functions.[72]

Regulatory Inputs Controlling the Actin Cytoskeleton

While a number of cellular components, including the constituents of the actin cytoskeleton discussed above, is well conserved throughout evolution, many others are not. In particular, this is often the case of components of the regulatory circuits controlling the function and mutual interactions of conserved "core" molecular modules. Such variable regulatory connections—or "protocols" *sensu* Csete and Doyle[75]—between well-conserved molecular assemblies may provide means for generating the great diversity of form and function from a relatively small set of molecular building blocks, as observed in present cells (see also ref. 12). Pathways controlling the structure and function of the actin cytoskeleton provide a good example, since they may seem to be almost entirely lineage-specific on the first glance. We shall now focus on the control of the actin nucleation machinery to illustrate this point; however, a similar argument could be constructed also for other actin-related regulatory pathways.

The Arp2/3 complex, which alone is inactive, can be activated by a variety of cofactors.[76] These include e.g., fungal myosin I and Abp1p, metazoan-specific cortactin, the multidomain protein CARMIL, found in Amoebozoa and Metazoa, and several proteins of broader distribution, such as e.g., coronin (see ref. 12). Prominent among Arp2/3 regulators is the large family of conserved WH2 domain-containing WAVE (WASP family verprolin homology) proteins that form a core of a large multiprotein regulatory complex. This complex has been long considered specific to Metazoa, Fungi and Amoebozoa; however, homologues have been recently found also in plants,[19] indicating that the WAVE-associated complex is ancient. It remains unclear whether an analogous system can be found also in the remaining deep-branching eukaryotic lineages; sequence searches for homologues are hampered by low sequence complexity of this extremely proline-rich protein. Also formins are embedded in complex signal networks.[12,77]

Despite the evolutionary plasticity of molecular mechanisms controlling the actin nucleation, some common motifs emerge if we follow the regulatory pathways backward from their actin targets to the upstream inputs. Remarkably, many regulators of actin nucleation, including WAVE proteins and at least some formins, are themselves controlled by Rho GTPases, a class of regulatory proteins with multiple outputs, initially described as major regulators of actin remodeling but later found to participate also in microtubule dynamics, endocytosis, vesicle trafficking, gene transcription, the response to oxidative stress, cytokinesis, cell cycle progression and apoptosis.[78-80] Small GTPases of the Rho family are present in all eukaryotes, although the "classical" subfamilies of Rho proper, Rac and Cdc42 are probably specific to Metazoa and Fungi.

Detailed discussion of Rho GTPases and their cofactors would exceed the scope of this review; a thorough evolutionary analysis of small GTPases has been published by others[81] (see also chapter by Balch). However, we should at least mention the fact that although the pathways whereby Rho GTPases control their downstream effectors vary greatly across eukaryotic lineages (e.g., see ref. 82), the common motif of actin control by a small GTPase appears to be almost invariant, despite of the varying molecular implementations (see ref. 12 for more detailed discussion). Remarkably, roots of this arrangement could be traced to very early stages of evolution, since MglB, the prokaryotic relative of profilin, often resides within the same operon with a small GTPase (see above and ref. 36).

Conclusions

We hope that careful evaluation of the above outlined data may allow an attempt at reconstructing the microfilament system of the ancestral eukaryote. We can assume that actin-binding domains shared by most eukaryotes (in particular those found on both sides of the unikont-bikont divide[74]) were present in the common ancestor of the eukaryotes. This means that any protein found both in a representative of the Fungi/Metazoa/Amoebozoa group (the unikonts) and in any of the remaining eukaryotic lineages examined (plants or nonamoeboid unicellular eukaryotes) is likely to be of ancestral origin. Quite many such proteins can be found (Table1); however, interpretation of their phyletic distribution is not always straightforward. Only actin itself is indeed found in all lineages. However, if we disregard the greatly reduced and incompletely characterized genome of the diplomonad *Giardia*, and allow for occasional gene loss and/or divergence beyond recognition in rapidly evolving lineages (in particular the unicellular ones), we have to realize that the common eukaryotic ancestor must have had a fairly elaborate cytoskeletal apparatus. This could have been expected, since the ancestor must have been already able to internalize the prokaryotes that later became endosymbionts and gave rise to mitochondria; therefore it must have been capable of some form of phagocytosis.

The ancestral eukaryote thus already possessed not only actin and myosin, but all relevant basic activities required for remodeling of the microfilament system. A profilin-like protein sequestered actin monomers and promoted nucleotide exchange, rendering the monomers ready for polymerization. Nucleation was achieved both by the Arp2/3 complex and by the dimeric FH2 domains. A dimeric capping protein capped the fast growing end of the filaments. This protein was initially homodimeric, but a gene duplication followed by substantial divergence resulted in the current heterodimeric protein. An ADF/gelsolin-related protein was responsible for severing and depolymerization of the filaments. Also proteins responsible for other activities, in particular bundling, crosslinking or membrane association of filaments, were probably already present; however, it has to be said that the evolutionary fate of these activities is more difficult to reconstruct, because in general they cannot be attributed to a single or very few well defined domains. These activities apparently evolved independently several times, frequently by lineage-specific combination and "fine tuning" of preexisting ancient domains (such as the CH domain, the VHP, the WH2 domain), or utilizing novel domains evolved in a single lineage (such as the I/LWEQ domain, which is restricted to the unikonts). Another universal, and most likely ancestral, feature of the actin cytoskeleton is its regulation by means of small GTPases.

Existence of prokaryotic relatives of several components of the actin cytoskeleton may provide clues towards reconstruction of its origin. The simplest predecessor of the above-described machinery might have consisted of a "core" of just a few proteins. The polymerization dynamics of MreB-related "protoactin" filaments could have been controlled by an MglB-like "protoprofilin", which might have already had some regulatory connection with a small GTPase. Nucleation of novel filaments might have first occurred spontaneously, perhaps on broken filament ends, later a specific conformation of two protoactin monomers supported by a WD40-like protein provided a more efficient nucleation core. Gene duplication supplied material for evolution of "nucleation-optimized" protoactin variants, nowadays known as Arp2 and Arp3, the specific WD40-like protein evolved into ARPC1, and the nucleation complex

later acquired additional subunits. All the other components, including the alternative nucleation complex based on dimeric FH2 domains, could have appeared later (although still prior to the acquisition of mitochondria) either de novo or by recruitment of preexisting domains (possible in case of the VHP and WH2 domains), perhaps with exception of a motor protein that could have interacted already with the protoactin filaments. Such a scenario can in principle be only speculative; however, we hope that at least some of the issues can be resolved by future work in the field of molecular phylogenetics, as well as by functional characterization of prokaryotic relatives of the recent cytoskeletal proteins.

Acknowledgements

The authors would like to thank Marek Eliás and Angelika Noegel for helpful discussion. Work in authors' laboratories has been supported by GACR grant 204/05/0268 (FC) and by grants of the Deutsche Forschungsgemeinschaft, the Center for Molecular Medicine Cologne and the Köln Fortune Program of the Medical Faculty of the University of Cologne (FR).

References

1. Doolittle RF, York AL. Bacterial actins? An evolutionary perspective. BioEssays 2002; 24:293-296.
2. Egelman EH. Actin's prokaryotic homologs. Curr Opin Struct Biol 2003; 13:244-248.
3. Margolin W. Bacterial shape: Concave coiled coils curve Caulobacter. Curr Biol 2004; 14:R242-R244.
4. Bramhill D. Bacterial cell division. Annu Rev Cell Dev Biol 1997; 13:395-424.
5. Ausmees N, Kuhn JR, Jacobs-Wagner C. The bacterial cytoskeleton: An intermediate filament-like function in cell shape. Cell 2003; 115:705-713.
6. Gupta RS, Golding GB. The origin of the eukaryotic cell. Trends Biochem Sci 1996; 21:166-171.
7. Horiike T, Hamada K, Shinozawa T. Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria supported by the newly clarified origin of functional genes. Genes Genet Syst 2002; 77:369-376.
8. Rivera MC, Lake JA. The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature 2004; 431:152-155.
9. Rivera MC, Jain R, Moore JE et al. Genomic evidence for two functionally distinct gene classes. Proc Natl Acad Sci USA 1998; 95:6239-6244.
10. Li JY, Wu CF. Perspectives on the origin of microfilaments, microtubules, the relevant chaperonin system and cytoskeletal motors - a commentary on the spirochaete origin of flagella. Cell Res 2003; 13:219-227.
11. Hartman H, Fedorov A. The origin of the eukaryotic cell: A genomic investigation. Proc Natl Acad Sci USA 2002; 99:16128-16133.
12. Cvrcková F, Bavlnka B, Rivero F. Evolutionarily conserved modules in actin nucleation: Lessons from Dictyostelium discoideum and plants. Protoplasma 2004; 224:15-31.
13. Mitchison TJ. Evolution of a dynamic cytoskeleton. Philos Trans R Soc Lond B Biol Sci 1995; 349:299-304.
14. Small JV, Stradal T, Vignal E et al. The lamellipodium: Where motility begins. Trends Cell Biol 2002; 12:112-120.
15. Bader MF, Doussau F, Chasserot-Golaz S et al. Coupling actin and membrane dynamics during calcium-regulated exocytosis: A role for Rho and ARF GTPases. Biochim Biophys Acta 2004; 1742:37-49.
16. Goldberg MB. Actin-based motility of intracellular microbial pathogens. Microbiol Mol Biol Rev 2001; 65:595-626.
17. Pollard TD, Borisy GG. Cellular motility driven by assembly and disassembly of actin filaments. Cell 2003; 112:453-465.
18. Pruyne D, Legesse-Miller A, Gao L et al. Mechanisms of polarized growth and organelle segregation in yeast. Annu Rev Cell Dev Biol 2004; 20:559-591.
19. Szymanski DB. Breaking the WAVE complex: The point of Arabidopsis trichomes. Curr Opin Plant Biol 2005; 8:103-112.
20. Baluska F, Salaj J, Mathur J et al. Root hair formation: F-actin-dependent tip growth is initiated by local assembly of profilin-supported F-actin meshworks accumulated within expansin-enriched bulges. Dev Biol 2000; 227:618-632.
21. Ringli C, Baumberger N, Diet A et al. ACTIN2 is essential for bulge site selection and tip growth during root hair development of Arabidopsis. Plant Physiol 2002; 129:1464-1472.

22. Vidali L, McKenna ST, Hepler PK. Actin polymerization is essential for pollen tube growth. Mol Biol Cell 2001; 12:2534-2545.
23. Mathur J, Hulskamp M. Microtubules and microfilaments in cell morphogenesis in higher plants. Curr Biol 2002; 12:R669-R676.
24. Field C, Li R, Oegema K. Cytokinesis in eukaryotes: A mechanistic comparison. Curr Opin Cell Biol 1999; 11:68-90.
25. Hales KG, Bi E, Wu JQ et al. Cytokinesis: An emerging unified theory for eukaryotes? Curr Opin Cell Biol 1999; 11:717-725.
26. Lénárt P, Bacher CP, Daigle N et al. A contractile nuclear actin network drives chromosome congression in oocytes. Nature 2005; 436:812-818.
27. Bamburg JR, Drubin DG. Actin depolymerizing factor (ADF)/cofilin. In: Kreis T, Vale R, eds. Guidebook to the Cytoskeletal and Motor Proteins. Oxford: Oxford University Press, 1999:19-23.
28. Paavilainen VO, Bertling E, Falck S et al. Regulation of cytoskeletal dynamics by actin-monomer-binding proteins. Trends Cell Biol 2004; 14:386-394.
29. McCurdy DW, Kovar DR, Staiger CJ. Actin and actin-binding proteins in higher plants. Protoplasma 2001; 215:89-104.
30. Frankel S. Arps, divergent members. In: Kreis T, Vale R, eds. Guidebook to the Cytoskeletal and Motor Proteins. Oxford: Oxford University Press, 1999:49-52.
31. van den Ent F, Amos LA, Löwe J. Prokaryotic origin of the actin cytoskeleton. Nature 2001; 413:39-44.
32. Jones LJ, Carballido-Lopez R, Errington J. Control of cell shape in bacteria: Helical, actin-like filaments in Bacillus subtilis. Cell 2001; 104:913-922.
33. Pollard TD. Profilins. In: Kreis T, Vale R, eds. Guidebook to the Cytoskeletal and Motor Proteins. Oxford: Oxford University Press, 1999:117-120.
34. dos Remedios GC, Chhabra D, Kekic M et al. Actin binding proteins: Regulation of cytoskeletal microfilaments. Physiol Rev 2003; 83:433-473.
35. Rivero F, Eichinger L. The microfilament system of Dictysotelium discoideum. In: Loomis WF, Kuspa A, eds. Dictyostelium Genomics. Norfolk: Horizon Bioscience, 2005:125-171.
36. Koonin EV, Aravind L. Dynein light chains of the Roadblock/LC7 group belong to an ancient protein superfamily implicated in NTPase regulation. Curr Biol 2000; 10:R774-R776.
37. Kurzbauer R, Teis D, de Araujo ME et al. Crystal structure of the p14/MP1 scaffolding complex: How a twin couple attaches mitogen-activated protein kinase signaling to late endosomes. Proc Natl Acad Sci USA 2004; 101:10984-10989.
38. Maciver SK, Hussey P. The ADF/cofilin family: Actin-remodeling proteins. Genome Biol 2002; 3:R3007.
39. Van Troys M, Vanderkerckhove J, Ampe C. Structural modules in actin-binding proteins: Towards a new clasification. Biochim Biophys Acta 1999; 1448:323-348.
40. Hatanaka H, Ogura K, Moriyama K et al. Tertiary structure of destrin and structural similartity between two actin-regulating protein families. Cell 1996; 85:1047-1055.
41. Kwiatkowski DJ. Functions of gelsolin: Motility, signaling, apoptosis, cancer. Curr Opin Cell Biol 1999; 11:103-108.
42. Bi X, Corpina RA, Goldberg J. Structure of the Sec23/24-Sar1 prebudding complex of the COPII vesicle coat. Nature 2002; 419:271-277.
43. Higgs HN, Pollard TD. Regulation of actin filament network formation through Arp2/3 complex: Activation by a diverse array of proteins. Annu Rev Biochem 2001; 70:649-676.
44. Machesky LM, Gould KL. The Arp2/3 complex: A multifunctional actin organizer. Curr Opin Cell Biol 1999; 11:117-121.
45. Beltzner CC, Pollard TD. Identification of functionally important residues of Arp2/3 complex by analysis of homology models from diverse species. J Mol Biol 2004; 336:551-565.
46. Mathur J. The ARP2/3 complex: Giving plant cells a leading edge. BioEssays 2005; 27:377-387.
47. Giaever G, Chu AM, Ni L et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature 2002; 418:387-391.
48. Evangelista M, Pruyne D, Amberg DC et al. Formins direct Arp2/3-independent actin filament assembly to polarize cell growth in yeast. Nature Cell Biol 2002; 4:32-41.
49. Li F, Higgs HN. The mouse formin mDia1 is a potent actin nucleation factor regulated by autoinhibition. Curr Biol 2003; 13:1335-1340.
50. Kovar DR, Kuhn JR, Tichy AL et al. The fission yeast cytokinesis formin Cdc12p is a barbed end actin filament capping protein gated by profilin. J Cell Biol 2003; 161:885-887.
51. Otomo T, Tomchick DR, Otomo C et al. Structural basis of actin filament nucleation and processive capping by a formin homology 2 domain. Nature 2005; 433:488-494.

52. Michelot A, Guerin C, Huang S et al. The Formin Homology 1 domain modulates the actin nucleation and bundling activity of Arabidopsis FORMIN1. Plant Cell 2005; 17:2296-2313.
53. Cvrcková F, Novotny M, Pícková D et al. Formin homology 2 domains occur in multiple contexts in angiosperms. BMC Genomics 2004; 5:44.
54. Rivero F, Muramoto T, Meyer A-K et al. A comparative sequence analysis reveals a common GBD/FH3-FH1-FH2-DAD architecture in formins from Dictyostelium, fungi and metazoa. BMC Genomics 2005; 6:28.
55. Higgs HN, Peterson KJ. Phylogenetic analysis of the formin homology 2 domain. Mol Biol Cell 2005; 16:1-13.
56. Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. Curr Biol 2004; 14:R693-R696.
57. Wear MA, Cooper JA. Capping protein: New insights into mechanism and regulation. Trends Biochem Sci 2004; 29:418-428.
58. Korenbaum E, Rivero F. Calponin homology domains at a glance. J Cell Sci 2002; 115:3543-3545.
59. Tirnauer JS, Bierer BE. EB1 proteins regulate microtubule dynamics, cell polarity, and chromosome stability. J Cell Biol 2000; 149:761-766.
60. van der Leij FR, Huijkman NC, Boomsma C et al. Genomics of the human carnitine acyltransferase genes. Mol Genet Metab 2000; 71:139-153.
61. Watanabe A, Yonemura I, Gonda K et al. Cloning and sequencing of the gene for a Tetrahymena fimbrin-like protein. J Biochem 2000; 127:85-94.
62. Mulder NJ, Apweiler R, Attwood TK et al. The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res 2003; 31:315-318.
63. Vardar D, Chishti AH, Frank BS et al. Villin-type headpiece domains show a wide range of F-actin-binding affinities. Cell Motil Cytoskeleton 2002; 52:9-21.
64. Hertzog M, van Heijenoort C, Didry D et al. The beta-thymosin/WH2 domain: Structural basis for the switch from inhibition to promotion of actin assembly. Cell 2004; 117:611-623.
65. Quinlan ME, Heuser JE, Kerkhoff E et al. Drosophila Spire is an actin nucleation factor. Nature 2005; 433:382-388.
66. Paunola E, Mattila PK, Lappalainen P. WH2 domain: A small versatile adapter for actin monomers. FEBS Lett 2002; 513:92-97.
67. Ponting CP, Russell JB. Identification of distant homologues of fibroblast growth factors suggests a common ancestor for all beta-trefoil proteins. J Mol Biol 2000; 302:1041-1047.
68. McCann RO, Craig SW. The I/LWEQ module: A conserved sequence that signifies F-actin binding in functionally diverse proteins from yeast to mammals. Proc Natl Acad Sci USA 1997; 94:5679-5684.
69. Kull FJ, Sablin EP, Lau R et al. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. Nature 1996; 380:550-555.
70. Kull FJ, Vale RD, Fletterick RJ. The case for a common ancestor: Kinesin and myosin motor proteins and G proteins. J Musc Res Cell Motil 1998; 19:877-886.
71. Leipe DD, Wolf YI, Koonin EV et al. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 2002; 317:41-72.
72. Thompson RF, Langford GM. Myosin superfamily evolutionary history. Anat Rec 2002; 268:276-289.
73. Hodge T, Cope MJ. A myosin family tree. J Cell Sci 2000; 113:3353-3354.
74. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436:1113-1118.
75. Csete ME, Doyle JC. Reverse engineering of biological complexity. Science 2002; 295:1664-1668.
76. Weaver AM, Young ME, Lee WL et al. Integration of signals to the Arp2/3 complex. Curr Opin Cell Biol 2003; 15:23-30.
77. Deeks MJ, Hussey P, Davies B. Formins: Intermediates in signal transduction cascades that affect cytoskeletal reorganization. Trends Plant Sci 2002; 7:492-498.
78. Johnson DI. Cdc42: An essential Rho-type GTPase controlling eukaryotic cell polarity. Microbiol Mol Biol Rev 1999; 63:54-105.
79. Etienne-Manneville S, Hall A. Rho GTPases in cell biology. Nature 2002; 420:629-635.
80. Sorokina EM, Chernoff J. Rho-GTPases: New members, new pathways. J Cell Biochem 2005; 94(2):225-231.
81. Jekely G. Small GTPases and the evolution of the eukaryotic cell. BioEssays 2003; 25(11):1129-1138.
82. Cotteret S, Chernoff J. The evolutionary history of effectors downstream of Cdc42 and Rac. Genome Biol 2002; 3:R0002.

CHAPTER 9

Origin and Evolution of Self-Consumption: Autophagy

Timothy Hughes and Tor Erik Rusten*

Summary

While misfolded and short-lived proteins are degraded in proteasomes located in the nucleus and cytoplasm, the degradation of organelles and long-lived proteins in the lysosome occurs by the process of autophagy. Central and necessary to the autophagic process are two conserved ubiquitin-like conjugation machineries. These conjugation machineries appear to be specific for autophagy and can together with genetic and morphological data be used to trace the natural history of autophagy. Here we discuss the origin and evolution of autophagy.

Introduction

The term autophagy, meaning eating *(phagy)* oneself *(auto)*, refers to the transport of cytoplasmic components to the degradative organelle called the vacuole in yeast and plants, and lysosome in other eukaryotes. For simplicity we will use the term lysosome in this chapter. The transport of cytoplasmic cargo can be membrane-driven or carried out by selective transport of certain proteins over the lysosomal membrane (Fig. 1). The latter process, called chaperone-mediated autophagy, is unrelated to other autophagy pathways and will not be further discussed. Autophagosomes en route to the lysosome can readily be detected and recognised by electron microscopy and have been observed mis-regulated in a number of pathological conditions like cancer, myopathies, neurodegeneration and bacterial infections.[1] The relevance of autophagy in these disease conditions has been unclear until recently since no good markers or means of manipulating the activity of the process were known. The ability to study autophagy was greatly increased by the identification of the central autophagy machinery and the identification of specific subcellular markers for autophagy in yeast. Autophagy is currently being intensively studied in a number of model organisms.[2]

To gain a better foundation for understanding the mechanism and relevance of autophagy to extant eukaryotes, it is useful to look at the evolution and presence of autophagy across different groups. In this chapter, we first review the different modes of autophagy and the specific molecular machinery involved. To assign the presence or absence of the process in different groups of organisms, we rely on positive identification of autophagy based on morphological and genetic data as well as presence of the autophagy-specific molecular machinery within taxa. We then infer the likely origin and subsequent loss of autophagic capacity within eukaryotic lineages based on the current understanding of eukaryotic phylogeny.[3,4] Finally, we discuss the relationship of autophagy to that of other degradation pathways used for degradation of transmembrane proteins (endocytosis) and short-lived proteins (proteasomes).

*Corresponding Author: Tor Erik Rusten—Department of Biochemistry, The Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway. Email: torr@klinmed.uio.no

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

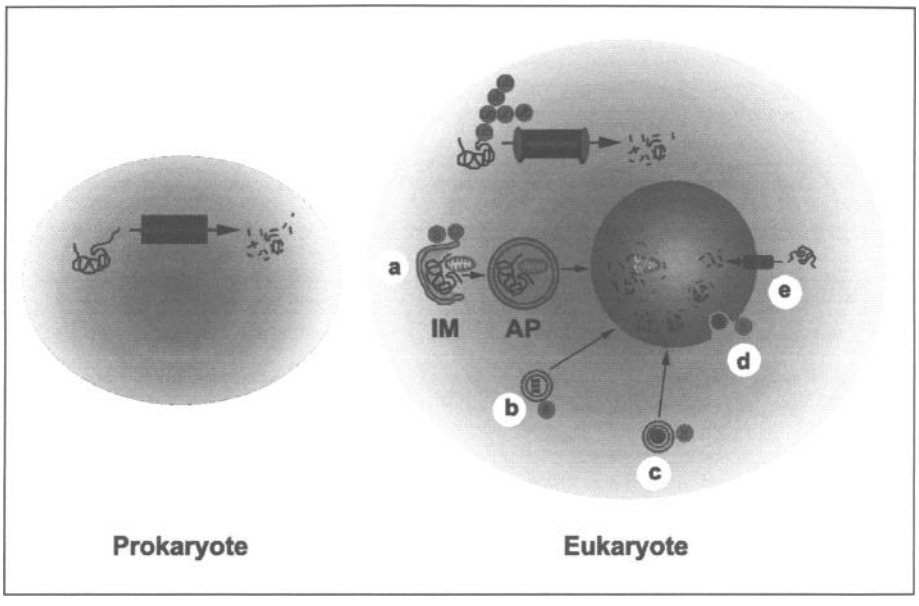


Figure 1. Schematic model of proteasomal and autophagic degradation pathways in pro- and eukaryotes. Proteins are degraded in prokaryotes by chambered proteases termed the 19S proteasome upon recognition of primary amino acid target sequences (red). The more complex 26S proteasome in eukaryotes is able to recognise and degrade polyubiquitinated (blue) proteins. a) Macroautophagy occurs by enveloping a part of the cytoplasm containing proteins and organelles in a cup shaped double isolation membrane (IM). The IM then closes resulting in the autophagosome (AP). The limiting membrane of the AP then fuses with the lysosome releasing the inner vesicle and its cargo for degradation. The two specific macroautophagic pathways, b) cytoplasm to vacuolar targeting (CVT) and c) macropexophagy occur in a similar manner. d) Microautophagy occurs by enveloping cytoplasm directly into the lysosome. Sometimes, direct uptake of organelles can occur such as during micropexophagy (d). While macroautophagy needs both Atg12 and Atg8 conjugation machineries, microautophagy appears to need only Atg8. e) The unrelated process of chaperone-mediated autophagy can transport proteins over the lysosomal membrane in an ATP-dependent manner. Ub-ubiquitin, 8-Atg8, 12-Atg12.

Degradation of Proteins in Chambered Proteases in Pro- and Eukaryotes

Degradation of intracellular proteins and organelles is important for cell homeostasis, as it regulates enzymatic activity, removes toxic or misfolded proteins and produces free amino acids. Both bacterial and eukaryotic cells degrade proteins by the use of proteases.[5] A set of threonine proteases work in concert in a superstructure called the proteasome, with a catalytically active lumen. By controlling entry of proteins into the proteasomal lumen, general cytoplasmic proteins are shielded from its activity. Whereas proteins in bacteria destined for degradation carry degradation signals embedded in their primary sequence, eukaryotes have evolved a protein conjugation process which tags proteins with polyubiquitin allowing specific recognition and degradation by the proteasome.

Different Modes of Autophagy Degrade Long-Lived Proteins and Organelles in Eukaryotes

Eukaryotic cells differ from prokaryotes by the presence of membrane bound organelles and a nucleus. Turnover and degradation of these structures cannot be performed by the

proteasome and instead occurs by the process of autophagy. Autophagy occurs by the transport of cytoplasmic components into the lysosome for degradation. Two modes of autophagy exist, macroautophagy and microautophagy (Fig. 1, reviewed in refs. 2,6). Microautophagy is the direct engulfment of cytoplasm and organelles by the lysosomal membrane. Macroautophagy, on the other hand, involves the isolation of a part of the cytoplasm by a cup-shaped double membrane structure and its subsequent degradation in the lysosome. Upon closure of the isolation membrane, the resulting autophagosome fuses with the lysosomal membrane releasing the inner vesicle with the cargo that is ultimately degraded. The capture of cytoplasmic components is unspecific and generally consists of endoplasmic reticulum, cytosol and organelles like mitochondria. Two variants of macroautophagy called cytoplasm-to-vacuole targeting (Cvt) and pexophagy differ from macroautophagy by being specific in cargo selection and by forming smaller autophagic vesicles. The Cvt pathway in the yeast *Saccharomyces cerevisiae* is a constitutive biosynthetic transport pathway for the vacuolar-resident enzyme aminopeptidase I. To date no evidence for this pathway has been found in other organisms than *S. cerevisiae*.[2] Pexophagy is, as the name suggests, specific autophagy of peroxisomes and has been studied extensively in the yeasts *Pichia pastoris*, *Hanensula polymorpha*, and *Pichia methanolica*.[7] Pexophagy in yeast, is induced by the rapid adaptation from a feeding substrate requiring large amounts of peroxisomes to one that does not. Based on biochemical data, selective pexophagy has also been suggested to occur in rat hepatocytes.[8]

The Autophagy-Specific Molecular Machinery Involves Two Ubiquitin-Like Conjugation Pathways

Genetic studies in yeast for genes necessary for macroautophagy, Cvt and pexophagy have revealed overlapping genetic requirements for all the pathways (reviewed in refs. 6,9). Many of the genes involved in completion of autophagy are clearly involved in other cellular processes and thus serve as poor indicators for the presence of autophagic capacity. Others, here termed the central autophagy machinery, appear specific for autophagy. Yeast mutants for genes of the central machinery, and thus autophagy deficient, grow normally under nonstarvation conditions. Moreover, this apparent nonrequirement in other developmental processes extends to metazoa since mouse mutants not capable of performing autophagy develop virtually normally until birth, upon which they appear to die from starvation.[10] The central autophagy machinery necessary for the membrane dynamics of autophagosome formation, consists of two ubiquitin-related conjugation pathways (Fig. 2, reviewed in refs. 11,12). The ubiquitin related molecules (Ubl) Atg12 and Atg8 both share limited sequence identity with ubiquitin but engage in protein modification systems strikingly similar to that of ubiquitin. The carboxy-terminal glycine of Atg12 is covalently attached to a lysine in the centre of Atg5 (Fig. 2).[13] Like in the ubiquitin conjugation pathway the C-terminal glycine of Atg12 is first activated by an E1 enzyme, Atg7, creating a high energy thioester bond. Subsequently Atg12 is transferred to a E2-like enzyme called Atg10 forming a second thioester bond. The conjugation of Atg12 to the target Atg5 is necessary for the progression of the second conjugation reaction of Atg8 to its substrate phosphatidylethanolamine (PE). A cysteine protease, Atg4, is necessary for activation of Atg8 revealing the C-terminal glycine. As for Atg12, the E1-like enzyme transferring Atg8 to a E2 like enzyme is Atg7 making it the only known E1 like enzyme to be conjugating two separate Ubl's. The E2 enzyme of Atg8, Atg3 transfers the Atg8 to phosphatidylethanolamine. Finally, like the case for ubiquitin, Atg8 is deconjugated by a protease and can potentially be used for another round of conjugation. The protease responsible is again Atg4.

The genetic requirements for microautophagy have mostly been studied in the context of micropexophagy in *Pichia pastoris*, and overlap with that of macroautophagy in requiring the Atg8 conjugation machinery.[14,15] Indeed, Atg8 gets recruited to the site of microautophagy on the lysosomal membrane and its activation by Atg4 is needed for micropexophagy.[14] It is less clear whether the Atg12 conjugation machinery is needed. To date, neither Atg5, Atg10, nor Atg12 has been shown to be required for micropexophagy or the Cvt pathway.[9] This suggests

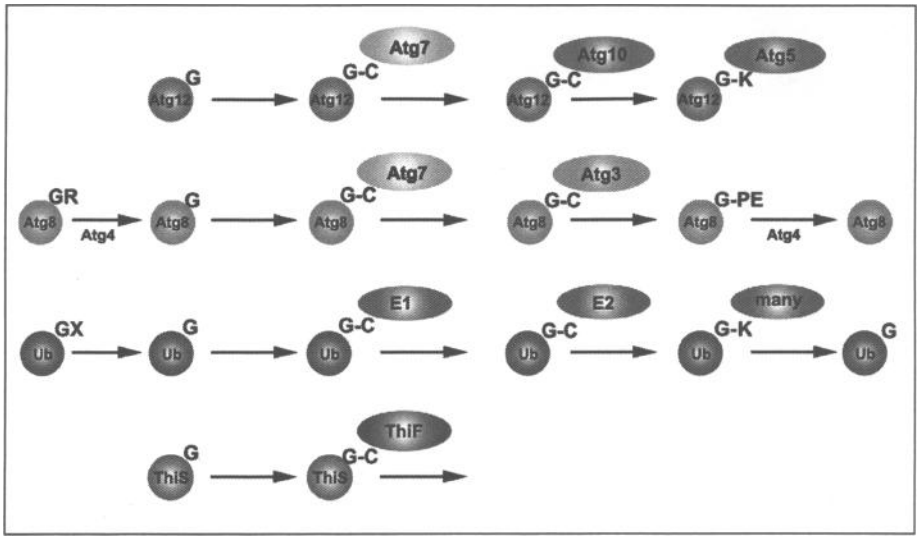


Figure 2. Comparison of conjugation pathways of selected ubiquitin superfamily members and the biosynthetic pathway of thiamine. The ubiquitin-like proteins, Atg12 and Atg8 necessary for autophagy undergo conjugation reactions closely resembling that of bacterial ThiS and ubiquitin itself. Atg7 is analogous to E1 and ThiF enzymes (See text for details, ref. 33).

that the Atg12 conjugation machinery is not strictly required for Atg8 conjugation and insertion into autophagosomal membranes in all types of autophagy.

In summary, it appears reasonable to assume that the presence of both the Atg12 and Atg8 machineries reflects the capacity for both macro and microautophagy and the presence of the Atg8 machinery alone reflects a capacity to conduct only microautophagy.

Is Autophagy a Pan-Eukaryotic Process?

Since prokaryotes neither have internal membranes nor lysosomes, autophagy necessarily has originated at a later point of evolution. The question arises whether autophagy is a pan-eukaryotic process and would be present in the last common eukaryotic ancestor (LCEA). Autophagy can be positively identified on morphological grounds based on electron microscopy, presence of autophagy-specific genes and genetic evidence for their necessity for autophagy in the respective organisms. The identification of the central autophagy machinery in yeast spurred the molecular genetic study of autophagy in other model systems.[7] Homologues of many of the central autophagy encoding genes like Atg5, Atg7, Atg10, Atg12 and Atg8 are present in all eukaryotic model organisms investigated to date (reviewed in ref. 2). Genetic, morphological and sequence-based evidence for autophagy exists from the major eukaryotic groups of plants (*Arabidopsis thaliana*), amoebozoa (*Dictyostelium discoideum*), fungi (*S. cerevisiae, H. polymorpha, P. pastoris* and *P. methanolica*), metazoa (*Caenorbhabditis elegans, Drosophila melanogaster* and *Mus musculus*).[2,16,17] The above findings are all obtained in model organisms used for molecular genetic cell biological studies. The situation is less clear when looking at model organisms within excavata and chromalveolata. Often, these organisms are chosen because they are important parasites relevant to human disease. Based on ultrastructural observations, macroautophagy has been suggested to occur in both *Leishmania donovani* (excavata) and *Tetrahymena thermophila* (chromalveolata) upon induction of cell death. These findings have, however, not been confirmed by molecular means.[18,19] Since the genome of *L. donovani* is not being sequenced a meaningful search for presence of autophagic genes within this organism is not feasible yet.[20] We searched for homologues of the core autophagic machinery in genomes within excavata and

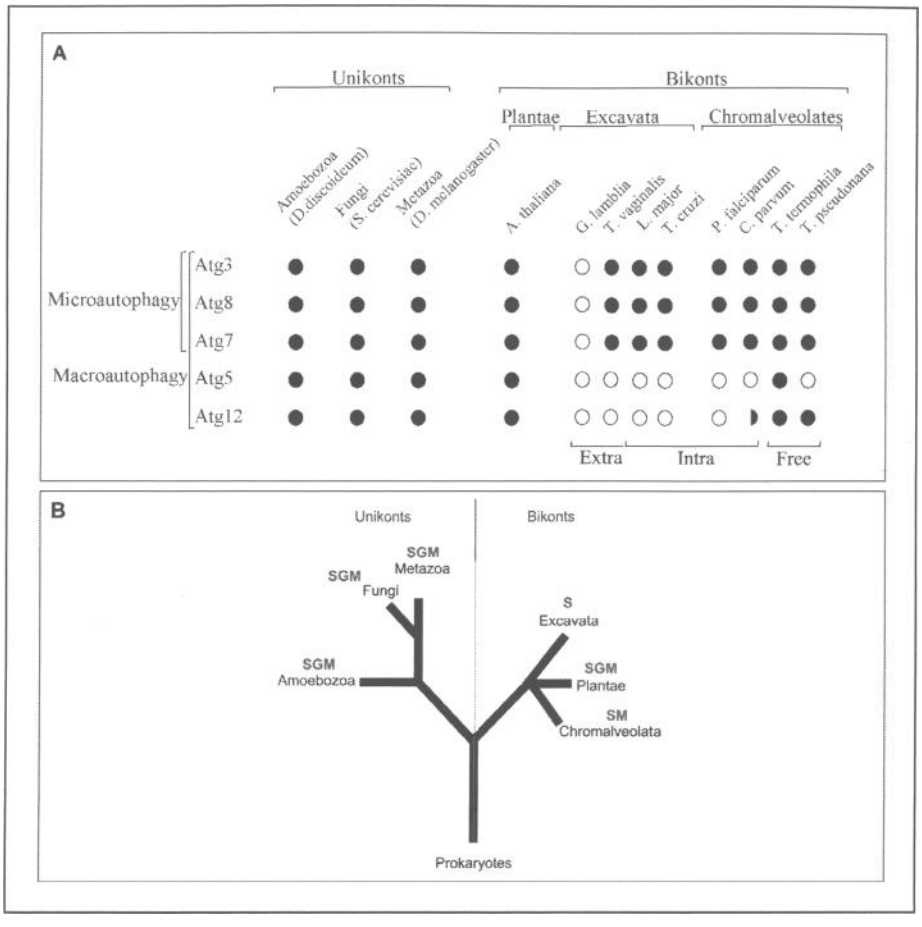


Figure 3. Evolutionary distribution of autophagy. A) The presence or absence of the Atg12 pathway, Atg8 pathway, and Atg7 needed for both pathways in each taxon, are illustrated (full circle: strong evidence of at least one homolog, half circle: weak evidence of a homolog, empty circle: failure to detect a homolog). The query sequences of *S. cerevisiae, D. melanogaster,* and *A. thaliana* were used to search the relevant genomes: *G.lamblia, T. vaginalis, C. parvum* and *T. pseudonana* at the NCBI (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi); *L. major, T. cruzi, P. falciparum* at the Sanger Institute;[34] and *T. thermophila* at the TIGR (http://tigr.org). The following genomes are considered complete: *T. cruzi*,[21] *L. major*,[36] *P. falciparum*,[22] *C. parvum*,[37] *T. pseudonana*, [38] and *T. vaginalis*(according to TIGR the sequencing centre). Blast settings: tblastn of protein sequences against genome database, BLOSUM62 substitution matrix, gapped blast, default gap opening and extension cost.[35] (Intra-intracellular parasite, Extra-extracellular parasite, Free-free swimming organism). B) Schematic tree summarising evidence for autophagy within each taxon. Evidence for autophagy based on ultrastructural studies (electron microscopy-M), functional studies (genetics-G), and genomic presence of genes (sequence-S) is plotted. The following genome sequencing efforts are not yet complete: *G. lamblia, T. thermophila.*

chromalveolata (Fig. 3A). We find that the genomes of *T. thermophila* and *T. pseudonana*, that are both free-swimming chromalveolates, encode homologs for the proteins of both the ATG8 and the ATG12 pathways. Interestingly, the situation is different in intra- and extracellular parasites within both taxa. The genomes of *T. cruzi*, *L major* and *T. vaginalis*, and the chromalveolate *P. falciparum*, encode only homologs for the proteins of the ATG8 pathway.[21,22]

Thus, there is a clear tendency for parasitic species whether excavata or chromalveolata to have lost the ATG12 pathway. *G. lamblia* appears to have also lost the proteins of the ATG8 pathway. This is perhaps not so surprising given that *G. lamblia* is known to lack organelles, such as mitochondria, peroxisomes and lysosomes, normally found in eukaryotes.[23]

The Atg12 machinery is strictly necessary for macroautophagy, but has to our knowledge not been shown to be necessary for microautophagy. By extension, it suggests that microautophagic, but not macroautophagic capacity is present within several parasites. Interestingly, one of the most studied functions of macroautophagy is to provide free amino acids upon starvation conditions by unspecific consumption of cytosol. The extracellular parasites *G. lamblia* and *T. vaginalis* and the intracellular parasites *L. major*, *T. cruzi*, *P. faliciparum* and *C. Pavrum* all spend parts or all their lifecycle intracellularly, or bathed in body fluids of the host(s). Presumably then, they are living in an environment where supply of amino acids is not limiting. It could be that macroautophagy for this reason is not strictly necessary for these highly specialized parasites. In line with this idea, neither members of the Atg12 nor the Atg8 group were found by BLAST analysis of the highly reduced genome of the obligate intracellular parasite *Encephalitozoon cuniculi*, which is considered to be a member of a fungi sister group.[24]

In summary, present available genomic information as well as ultrastructural and genetic evidence exists for both macro- and microautophagy within bikonts and unikonts (Fig. 3B). The core autophagy machinery was therefore present in the LCEA. Partial or complete loss of autophagic capacity, most likely have occurred secondarily in parasitic species within both bikonts and unikonts. We postulate that macro- and microautophagy will be present in most nonparasitic eukaryotes.

Discussion

Sampling core molecular machineries used for distinct cellular processes across eukaryotic taxa allows deduction of the minimal set of these molecules present in the LCEA. Caution should be shown when interpreting the original function of these machineries, since cooption of molecules for new functions is one of the main driving forces of evolution. This caveat can be partially counteracted by ensuring that the molecules involved perform the same functions within both unikonts and bikonts by genetic studies, since it would be less likely that the same function would be acquired independently.

Using this approach, it has been proposed that the LCEA contained a complex endomembrane system involved in both exocytic and endocytic events.[25] Autophagy is intimately linked with endocytosis. Not only do the two pathways end up in the same acidic degradative compartment, there is also evidence for fusion of autophagosomes with late endocytic compartments before the cargo is released in lysosomes.[26] It is therefore important to understand the extent to which the endocytic machinery was present in the LCEA. The core components of endocytosis including specific components of early endosomal, as well as recycling and sorting endosomal compartments were all present in the LCEA (see the chapter by Field, Gabernet-Castello, Dacks in this book). Importantly, the HOPS SNARE complex of proteins necessary for the fusion of autophagosomes with lysosomes was also present.[27] Thus, the wide pan-eukaryotic distribution of the core autophagic machinery and the essential endosomal machinery with which it interacts argues for the presence of autophagy in the LCEA.

An interesting and related question is: what was the original function of autophagy? Perhaps the most widely known role of autophagy is its involvement in survival under starvation conditions. Unspecific macroautophagy of cytoplasm can be induced in plants, amoeba, fungi, insects and mammals allowing intracellular components to be rapidly recycled and reused during starvation.[2,10,16] This allows near normal activity of cells even though steady supply of nutrients is not available. It is conceivable that keeping up cellular activity rather than forming spores during starvation conditions gave a selective advantage in early eukaryotic evolution by allowing cells to localise to new foraging locations while consuming intracellular pools of amino acids.

Other suggested functions of autophagy, that have recently gained functional evidence from genetic studies, are: immune response to bacterial infections, promotion of longevity, removal of intracellular protein aggregates, prevention of neurodegeneration and finally the need for autophagy to control cell death and cancer.[2,28,29] While it is hard to speculate how the latter roles were relevant to early eukaryotic life it is easier to hypothesise an important role for autophagy in preventing infections. Some intracellular bacterial pathogens, like *Shigella* and *Streptococcus* escape the phagosome or endosome and multiply in the cytoplasm. Autophagy can counteract such infections by recapturing and degrading these escaped bacteria.[30,31] Thus, autophagy may represent a part of the very first innate immune system.

From a pragmatic point of view, the necessity for autophagy seems to complement the two other intracellular degradation pathways, the proteasomal and endocytic pathway. Proteasomes degrade intracellular short-lived proteins tagged by polyubiquitin. Endocytic and autophagic trafficking shuttles membrane associated molecules or organelles to be degraded in the lysosome. Endocytic trafficking ensures that the cell is able to digest transmembrane proteins within the lysosomal lumen. Some transmembrane cargo, like the Egf receptor, need to be monoubiquitinated to follow the degradation route.[32] Finally, autophagy is responsible for degradation of long-lived proteins, turnover of intracellular membrane-bound organelles and protein aggregates. The appearance of endocytosis and autophagy correlate with appearance of the LCEA. Their simultaneous acquisition could have arisen because of the need for degradation pathways to deal with complex intracellular membrane compartments not degraded by the proteasome.

The wide distribution of autophagy within unikonts and bikonts argues for a fundamental need for autophagy in nonparasitic species. It will be interesting to see to what extent the molecular control and use of autophagy has been adapted to meet different needs in different taxa.

Acknowledgements

The authors would like to thank Per O. Seglen, Harald Stenmark and Karine Lindmo for critically reading the manuscript. T. E. Rusten is supported by the Norwegian Research Council.

References

1. Cuervo AM. Autophagy: In sickness and in health. Trends Cell Biol 2004; 14(2):70-77.
2. Levine B, Klionsky DJ. Development by self-digestion: Molecular mechanisms and biological functions of autophagy. Dev Cell 2004; 6(4):463-477.
3. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436(7054):1113-1118.
4. Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 2005; 6(5):361-375.
5. Pickart CM, Cohen RE. Proteasomes and their kin: Proteases in the machine age. Nat Rev Mol Cell Biol 2004; 5(3):177-187.
6. Ohsumi Y. Molecular dissection of autophagy: Two ubiquitin-like systems. Nat Rev Mol Cell Biol 2001; 2(3):211-216.
7. Klionsky DJ. Autophagy. Georgetown, Texas, USA: Landes Bioscience, 2004.
8. Luiken JJ, van den Berg M, Heikoop JC et al. Autophagic degradation of peroxisomes in isolated rat hepatocytes. FEBS Lett 1992; 304(1):93-97.
9. Klionsky DJ, Cregg JM, Dunn Jr WA et al. A unified nomenclature for yeast autophagy-related genes. Dev Cell 2003; 5(4):539-545.
10. Kuma A, Hatano M, Matsui M et al. The role of autophagy during the early neonatal starvation period. Nature 2004; 432(7020):1032-1036.
11. Noda T, Suzuki K, Ohsumi Y. Yeast autophagosomes: De novo formation of a membrane structure. Trends Cell Biol 2002; 12(5):231-235.
12. Klionsky DJ, Emr SD. Autophagy as a regulated pathway of cellular degradation. Science 2000; 290(5497):1717-1721.
13. Mizushima N, Noda T, Yoshimori T et al. A protein conjugation system essential for autophagy. Nature 1998; 395(6700):395-398.
14. Mukaiyama H, Baba M, Osumi M et al. Modification of a ubiquitin-like protein Paz2 conducted micropexophagy through formation of a novel membrane structure. Mol Biol Cell 2004; 15(1):58-70.

15. Strømhaug PE, Bevan A, Dunn Jr WA. GSA11 encodes a unique 208-kDa protein required for pexophagy and autophagy in Pichia pastoris. J Biol Chem 2001; 276(45):42422-42435.
16. Scott RC, Schuldiner O, Neufeld TP. Starvation-induced autophagy is suppressed by TOR signaling independently of S6K and ameliorates the effects of loss of TOR function. Dev Cell 2004; 7(2):167-178.
17. Rusten TE, Lindmo K, Juhâasz G et al. Programmed autophagy in the Drosophila fat body is induced by ecdysone through regulation of the PI3K pathway. Dev Cell 2004; 7(2):179-192.
18. Bera A, Singh S, Nagaraj R et al. Induction of autophagic cell death in Leishmania donovani by antimicrobial peptides. Mol Biochem Parasitol 2003; 127(1):23-35.
19. Christensen ST, Chemnitz J, Straarup EM et al. Staurosporine-induced cell death in Tetrahymena thermophila has mixed characteristics of both apoptotic and autophagic degeneration. Cell Biol Int 1998; 22(7-8):591-598.
20. Bernal A, Ear U, Kyrpides N. Genomes OnLine Database (GOLD): A monitor of genome projects world-wide. Nucleic Acids Res 2001; 29(1):126-127.
21. El-Sayed NM, Myler PJ, Bartholomeu DC et al. The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 2005; 309(5733):409-415.
22. Gardner MJ, Hall N, Fung E et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 2002; 419(6906):498-511.
23. Lanfredi-Rangel A, Attias M, de Carvalho TM et al. The peripheral vesicles of trophozoites of the primitive protozoan Giardia lamblia may correspond to early and late endosomes and to lysosomes. J Struct Biol 1998; 123(3):225-235.
24. Katinka MD, Duprat S, Cornillot E et al. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature 2001; 414(6862):450-453.
25. Jekely G. Small GTPases and the evolution of the eukaryotic cell. Bioessays 2003; 25(11):1129-1138.
26. Gordon PB, Seglen PO. Prelysosomal convergence of autophagic and endocytic pathways. Biochem Biophys Res Commun 1988; 151(1):40-47.
27. Pulipparacharuvil S, Akbar MA, Ray S et al. Drosophila Vps16A is required for trafficking to lysosomes and biogenesis of pigment granules. J Cell Sci 2005; 118(Pt):3663-3673.
28. Pattingre S, Tassa A, Qu X et al. Bcl-2 antiapoptotic proteins inhibit Beclin 1-dependent autophagy. Cell 2005;122(6):927-939.
29. Ravikumar B, Vacher C, Berger Z et al. Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. Nat Genet 2004; 36(6):585-595.
30. Nakagawa I, Amano A, Mizushima N et al. Autophagy defends cells against invading group A Streptococcus. Science 2004; 306(5698):1037-1040.
31. Ogawa M, Yoshimori T, Suzuki T et al. Escape of intracellular Shigella from autophagy. Science 2005; 307(5710):727-731.
32. Raiborg C, Rusten TE, Stenmark H. Protein sorting into multivesicular endosomes. Curr Opin Cell Biol 2003; 15(4):446-455.
33. Hochstrasser M. Evolution and function of ubiquitin-like protein-conjugation systems. Nat Cell Biol 2000; 2(8):E153-157.
34. Hertz-Fowler C, Peacock CS, Wood V et al. GeneDB: A resource for prokaryotic and eukaryotic organisms. Nucleic Acids Res 2004; 32:D339-343 (Database issue).
35. Altschul SF, Madden TL, Schèaffer AA et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997; 25(17):3389-3402.
36. Ivens AC, Peacock CS, Worthey EA et al. The genome of the kinetoplastid parasite, Leishmania major. Science 2005; 309(5733):436-442.
37. Abrahamsen MS, Templeton TJ, Enomoto S et al. Complete genome sequence of the apicomplexan, Cryptosporidium parvum. Science 2004; 304(5669):441-445.
38. Armbrust EV, Berges JA, Bowler C et al. The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. Science 2004; 306(5693):79-86.

CHAPTER 10

Origin and Evolution of the Centrosome

Michel Bornens* and Juliette Azimzadeh

Summary

In this brief account we specifically address the question of how the plasma membrane-associated basal body/axoneme of the unicellular ancestor of eukaryotes has evolved into the centrosome organelle through the several attempts to multicellularity. We propose that the connection between the flagellar apparatus and the nucleus has been a critical feature for leading to the centriole-based centrosome of metazoa, the Spindle Pole Body of fungi, or to the absence of any centrosome in seed plants. We further suggest that the evolution of this connection could be reflected in the evolution of the centrin proteins. We then review evidence showing that the evolution of the centrosome-based tubulin network has been correlated with the evolution of the cortical actin-based cleavage apparatus. Finally we argue that this coevolution had a major impact on the cell individuation process and on the evolution of multicellular organisms. We conclude that only the metazoan lineage evolved multicellularity without loosing the ancestral association of three basic cellular functions of the basal body/axoneme or the derived centrosome organelle, namely sensation, motion and division.

Introduction

The origin of the flagellar apparatus has recently been discussed in the light of the evolution of tubulin, dynein and kinesin[1,2] (see chapter by David Mitchell) or in the light of the evolution of intracellular coated vesicle transport.[3,4] Many important insights and references can be found in these contributions. In the present chapter, we will depart from the models proposed in these contributions on one important aspect. Both of these models assume that a cytoplasmic microtubule nucleating centre preceded the flagellum. In this view the evolutionary scenario reflects what is observed for axoneme growth in contemporary organisms. We will rather argue that the microtubule network centred on a cytoplasmic organelle is a secondary type of organisation derived from the plasma membrane-associated basal body/axoneme and microtubule pellicle during the evolution of multicellularity. How the plasma-membrane-associated microtubule network originated is not addressed here.

The centriole/basal body structure and the associated (9+2) flagellum are very ancient inventions, present at the apparition of the early eukaryotic cells.[5] Before being a central body (a centro-some) as observed in most metazoan cells, the ancestral organelle was a plasma membrane-associated body from which most of the microtubule cytoskeleton was organized in a sub-membrane pellicle in the cell body having a critical role for cell organisation, the other part being organized as a (9+2) axoneme in the ciliary or flagellar extension. Contemporary examples of such organization include Kinetoplastids, in which the whole cell division process relies on the duplication of the centriole/basal body.[6,7]

*Corresponding Author: Michel Bornens—UMR 144-CNRS-Institut Curie 75248 Paris Cedex 05, France. Email: mbornens@curie.fr

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

Motion, sensation and division are three major functional modules that have been associated with the ancestral organelle (see ref. 8 for a discussion of this aspect). Whether each module was recruited independently or not is an important question. A detailed analysis of molecular pathways participating in the three modules should reveal the extent of their overlap. Whole genome evolution might further reveal how the recruitment of the three modules has taken place during flagellum evolution. In any case, sensation and motion need to be coupled to be efficient. This is observed in bacteria as well, although the bacterial flagellum is quite distinct from the eukaryotic flagellum and the coupling between sensation and movement rests on very different mechanisms.[9] Motion and sensation are two essential facets of cell polarity in eukaryotic cells, a critical feature of cell organisation and activity. The two daughter cells formed by cell division must inherit a polarized organization to be viable. A parsimonious solution for this requirement is that the basal/body axoneme and the associated cytoskeleton directly control the bipolarity of the dividing cell and orchestrate the division process. Recruiting motion, sensation and division modules on the centriole/basal body axoneme and the associated polarised cortical microtubule array could have given optimal survival value to primitive eukaryotic cells, by ensuring them with an efficient and robust machinery for maintaining cell individuation through the division process.[8] Any further evolution of the basal body/axoneme would have been the result of selective pressure on the genomic wiring involved in the maintenance of efficient cell motion, cell sensation or cell division.

The Centrosome-Nucleus Connection

An important feature of cell polarity is the positioning of the nucleus. In most unicellular organisms, there exists a structural association between the nucleus and the basal body/axoneme. In the early days of modern cell biology, association between the nucleus and the flagellar apparatus was seen as a structure organelle, the karyo-mastigont, involving the basal body, the nucleus and the Golgi complex, able to replicate as a unit, together with the genetic material.[10] Interestingly, the nucleus-basal body connection in *C. reinhardtii* was also seen in the past as the neuro-motor.[11] Although the whole concept of karyo-mastigont might appear odd and without molecular characterization, or perhaps pertinent only for some particular unicellular organisms, it is noteworthy that in animal and fungal cells, the centrosome/SPB is also associated with the nucleus.[12,13] This feature participates in the control of nucleus positioning during cell motion and cell division, which involves a conserved dynein-based pathway connecting the centrosome-associated nucleus to the cell cortex.[14] This pathway has important physiological implications during development, particularly for brain ontogenesis.[15] The centrosome-nucleus association could have additional important implications. It has recently been elegantly demonstrated that the maintenance of heterozygous state in *S. cerevisiae* depends on a cellular meiotic mechanism, which rests on the conservative replication of the nucleus-associated SPB.[16]

In order to address the origin and evolution of the centrosome/SPB of higher eukaryotes from the basal body axoneme of the unicellular ancestors, we will investigate the fate of the nucleus-basal body connection in the light of the evolution of the centrin family, which participates in this connection.

Centrosome Has Evolved with Multicellularity

Multicellularity evolved in several lineages independently[17] and has led to very different types of multicellular states. There are clades in which all species are multicellular: their life cycle always involves a multicellular state which in some cases can be transient, as a response to stringent environmental conditions, whereas in others it is permanent, with a transient unicellular stage for reproduction. Classically, these groups involve animals, plasmodial and cellular slime molds, land plants and red algae. We will see that in each of these clades the multicellular state is different from that of the others, either with respect to its genesis or to cell-cell organization. Other clades contain unicellular and colonial/multicellular species. These include

numerous clades such as choanoflagellates, the sister group of animals. Finally, in the supergroup of chromalveolata, diatoms and ciliates are two clades in which most species are unicellular with rare multicellular species.[17]

Multicellularity is thought to have evolved from a colonial state. The selective advantages of multicellularity are the matter of speculations.[17] It has been suggested that escape from predation may have favoured a colonial state, with some experimental evidence. Escape from starvation might also have been a general trigger: cell aggregation, the construction of fruiting bodies and the differentiation into stress-resistant spores is a common response to starvation.[18] A most supported proposal, the so-called flagellar synthesis constraint, argues that motion and division are mutually exclusive due to the fact that the basal body cannot participate in cell division and flagellar swimming at the same time.[19] This rule is not an absolute one, however. Several patterns have been selected during evolution for the basal/body axoneme to act as a division organ. For example, *Chlorogonium elongatum*, a *Chlamydomonas*-related chlorophyte keeps swimming during cell division by modifying the anchorage of the beating flagella on the cell body while basal bodies dissociate and migrate towards the nucleus.[20] A different example comes from *Trypanosoma*, in which the flagellar membrane is tightly associated to the cell body membrane in almost all its length. The flagellar apparatus controls cell organisation throughout cell division, which starts at the duplicating basal body and terminates at the tip of the flagellum.[21,22] Most of the unicellular flagellates, however, loose their flagellum during the division process. The colonial state would have allowed some cells to divide while others maintain motility for the whole colony. The balance between motile and immotile cells in a primitive multicellular organism may even have played a role at the origin of gastrulation.[17] Thus the basal body/axoneme organ essential for the polar organization of unicellular organisms could also have been critical for triggering the colonial state. The success of multicellularity has certainly required many other features that will not be addressed here. We will only attempt to analyse the fate of the basal body/axoneme in the major clades where multicellularity has evolved, namely Opisthokonts, Amoebozoa and Plantae. We will not deal with chromalveolata nor with clades like Rhizaria in which some unicellular organisms can have a central body without a flagellum, as molecular data are lacking.

The Evolution of the Centrin Genes

Recent phylogenetic studies indicate that the last common ancestor of eukaryotes was a cell with a single cilium (an unikont), possessing an endosymbiont-derived mitochondrion, one axoneme, most likely one single centriole with a cone of root microtubules and the cellular machinery to form pseudopodia.[23] This view places the root between unikonts and bikonts, a view also supported by a recent extensive analysis of myosin evolution[24] but which is still debatable[25] (see chapter by Brinkmann and Philippe).

To analyse the fate of the basal body/axoneme among divergent species, we will look at the evolution of centrin, a very ancient protein closely related to calmodulin and specifically associated with basal body/axoneme or centrosomal structures over a broad range of eukaryotes. The centrin and the calmodulin proteins are 'eukaryotic signature proteins' (ESP), which participate in a complex calcium signalling system. ESPs are 347 proteins that have no significant homology in Archaea and Bacteria.[26] Centrin proteins are known to be associated with centriole/basal-body/axoneme structures in many divergent species and could form connecting structures between the motile apparatus and most of the cellular compartments, particularly the plasma membrane and the nucleus.[27,28] A comprehensive view of their functions within centrosomes has still to be established as distinct roles have been reported. Centrin proteins are required in the SPB/centrosome duplication process.[29-33] They are also constituents of calcium-sensitive fibres that connect the basal bodies to one another and to the nucleus in diverse flagellate green algae[27] and are implicated in basal body localisation and segregation.[34] However, they are also present in taxa that lack any type of centrosome, like land plants, where they could participate in other functions. Centrins have been shown to act as regulators in very

diverse cellular processes including DNA repair and RNA export, and shown to be present in the yeast nuclear pore complex.[35-41]

The evolution of centrin and centrin-binding proteins has been analysed elsewhere.[8] We will briefly summarize it before proposing an interpretation for the origin of the centrosome/SPB. An up-dated tree is shown in Figure 1.

The two founding members of the centrin family, i.e., CrCenp (*C. reinhardtii*) and Cdc31p (*S. cerevisiae*) define two subfamilies to which many of the known centrin genes belong. The first subfamily contains centrins from green algae, lower land plants, insects, and vertebrate homologues of the human ubiquitous isoform, centrin 2, and of centrin 1 and 4, two isoforms specific for vertebrate ciliated epithelia.[42-47] The second subfamily is formed by fungal, ScCdc31p-like centrin proteins and vertebrate homologues of human centrin 3, the other ubiquitous human isoform.[48]

The widespread but not universal distribution of ScCdc31p and CrCenp subfamilies indicates that the duplication of the common ancestor gene happened early in eukaryotic evolution. Both isoforms are present in the bikont protist *Giardia intestinalis,* showing that the absence of one or the other isoform in some taxa is due to secondary loss. The duplication of the ancestral centrin gene occurred before the separation of unikonts, which encompass amoebozoa and opisthokonts (fungi, animals, choanozoa), and bikonts, which encompass green plants and major protozoan groups (Fig. 2). One of the two isoforms has been lost early in some taxa, or has become highly divergent, suggesting that functional constraints on that isoform have been relaxed in these taxa, while the other isoform is conserved. For example, only the ScCdc31p-like isoform is present in higher fungi, whereas this isoform has become highly divergent in green algae (Cen3' in Fig. 2). Among the chlorobionts, flowering plants have an additional divergence: they not only lost the ScCdc31p-like isoform but have evolved a divergent subfamily CrCenp-like isoform (Cen2' in Fig. 2). In other taxa, including vertebrates, both isoforms have been conserved. Additional divergence seems, however, to have taken place in metazoa. In Ecdysozoa the ScCdc31p-like isoform has been lost in the fly and in nematodes, while the CrCenp-like centrin diverged substantially in the worms (Fig. 1).

The distribution of CrCenp subfamily members strongly suggests that these genes are specifically required for basal body/axoneme-related function.[8] Higher fungi have lost the CrCenp subfamily secondarily as indicated by the existence of both ScCdc31p and CrCenp subfamilies in Chytridiomycetes fungi, which form flagellated gametes.[49] Thus, similarly to the divergence of the CrCenp subfamily in angiosperms, the loss of CrCenp subfamily in fungi correlates with the loss of the motile apparatus. The ScCdc31p-like centrins are clearly associated with the centrosome/SPB duplication in some organisms. However, their precise functions in divergent organisms with structurally different centrosomes and specific controls of cell cycle progression is far from clear. The loss of the ScCdc31p-like isoform from flies and nematodes is more puzzling (Fig. 1). This suggests that centrin's function in centrosome duplication is not conserved. If we assume that the core process of centrosome duplication has been conserved, as suggested by the conservation of other centrosome-related genes,[50] one is left with the possibility that the ScCdc31p-like centrins function during centrosome duplication in some organisms but not in others.

The Centrosome-Nucleus Connection and the Duplication Process

As stressed above, the nucleus-basal body connection is a critical element for cell polarity. Accordingly, a link between the centrosome and the nucleus has been conserved in many divergent organisms. The link must be preserved continuously during the cell cycle, also during centrosome reproduction. Here we discuss the possibility that centrin requirement for centrosome reproduction is due to its direct participation in the structures and mechanisms that ensure nucleus-centrosome/SPB connection through cell division. This is suggested by the fate of centrin-containing structures that connect the basal body/axoneme to the nucleus in many unicellular systems during cell division (see for example refs. 51, 52). Several centrin-binding

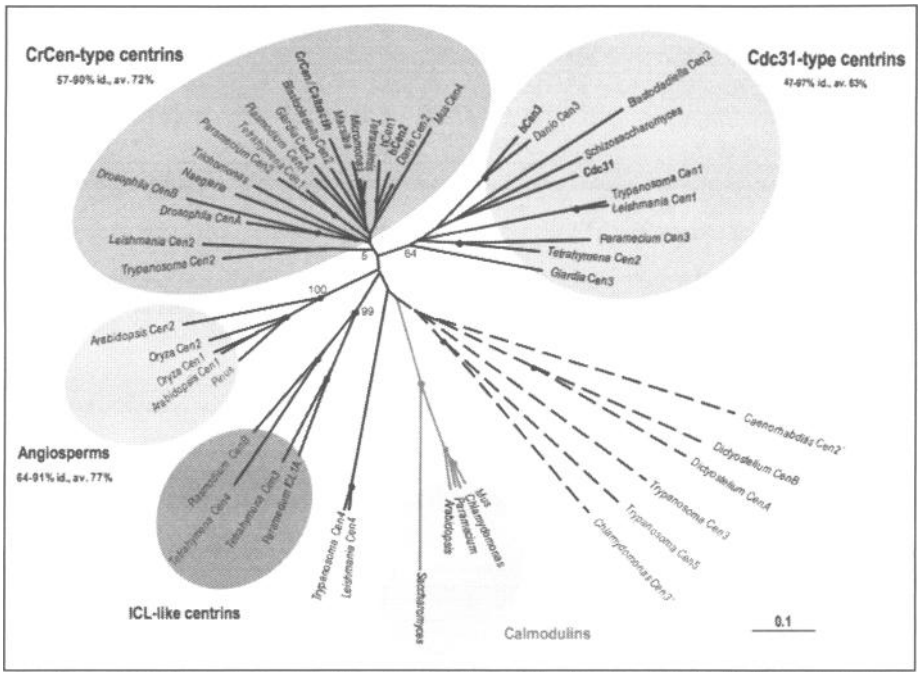


Figure 1. Phylogenetic tree of the centrin family based on the neighbor-joining method (Saitou and Nei, 1987) using the CLUSTAL W program (Thompson et al, 1994) and drawn using Phylodendron - Version 0.8d (D.G. Gilbert) at www.es.embnet.org. NJ tree was constructed from alignment of protein sequences covering the four EF-hand domains (residues 29-162 for *Chlamydomonas reinhardtii* centrin). The following centrin sequences were used: OPISTHOKONTS: **Metazoa**: i) Chordates: *Danio rerio* Cen2 (CF269323, BQ450470), Cen3 (BM141295); *Homo sapiens* Cen1, 2, 3 (Q12798, P41208, O15182); *Mus musculus* Cen4 (NP_665824); ii) Ecdysozoa: Arthropods: *Drosophila melanogaster* CenA, B (CG17493, CG31802); Nematodes: *Caenorhabditis elegans* (NM_066585); **Fungi**: i) Ascomycota: *Saccharomyces cerevisiae* Cdc31p (S47549); *Schizosaccharomyces pombe* SpCdc31p (T41061); ii) Chytrids: *Blastocladiella emersoni Cen1, 2* (Q4F6W6, Q4F6W5); MYCETOZOANS: *Dictyostelium discoideum* CenA, B (Q54X77, Q54IY3); ALVEOLATES: i) Ciliates: *Paramecium tetraurelia* ICL1A (CR932086), Cen2a (CR932099), Cen3a (CR932089); *Tetrahymena termophila* Cen1, 2, 3, 4 (41.m00190, 65.m00239, 105.m00148, 51.m00221); ii) Apicomplexans: *Plasmodium falciparum* CenA, B (NP_703273, NP_702332); EUGLENOZOANS: *Leishmania donovanii* Cen1, 2, 4 (LinJ34.1880, LinJ07.0760, LinJ22.1230); *Trypanosoma brucei* Cen1, 2, 3, 4, 5 (Tb10.6k15.1830, Tb927.4.2260, Tb927.8.1080, Tb927.7.3410, Tb11.01.5470); DIPLOMONADS: *Giardia lamblia* CenA, B (U59300, U42428); HETEROLOBOSEA: *Naegleria gruberi* (U21725); PARABASALIDS: *Trichomonas vaginalis* (CAB55607); GREEN PLANTS: i) Ulvophytes: *Chlamydomonas reinhardtii* Cen (P05434), Cen3 (C_1460027); *Micromonas pusilla* (CAA58718); *Tetraselmis striata* (P43646); ii) Streptophytes: Ferns: *Marsilea vestita* (U92973); Angiosperms: Monocots: *Pinus taeda* (CX646981.1); Dicots: *Arabidopsis thaliana* CenA, B (CAB16762, T45582); Monocots: *Oryza sativa* CenA, B (BAC79876). **Calmodulins**: AtCaM1 (P25854); CrCaM (P04352); MmCaM1 (AAH54805); PtCaM (P07463); ScCaM (P06787). Bootstrap values are displayed as percentages at the principal nodes. Dots indicate bootstrap values ≥ 75 %. Note the very low bootstrap value at the Cen2 node, which could reflect a grouping of those sequences due to overall sequence conservation rather than real phylogeny. Dotted lines indicate a likely artefactual positioning due to long branch attraction (Gribaldo and Philippe, 2002). Range of protein identity within the different subgroups is indicated.

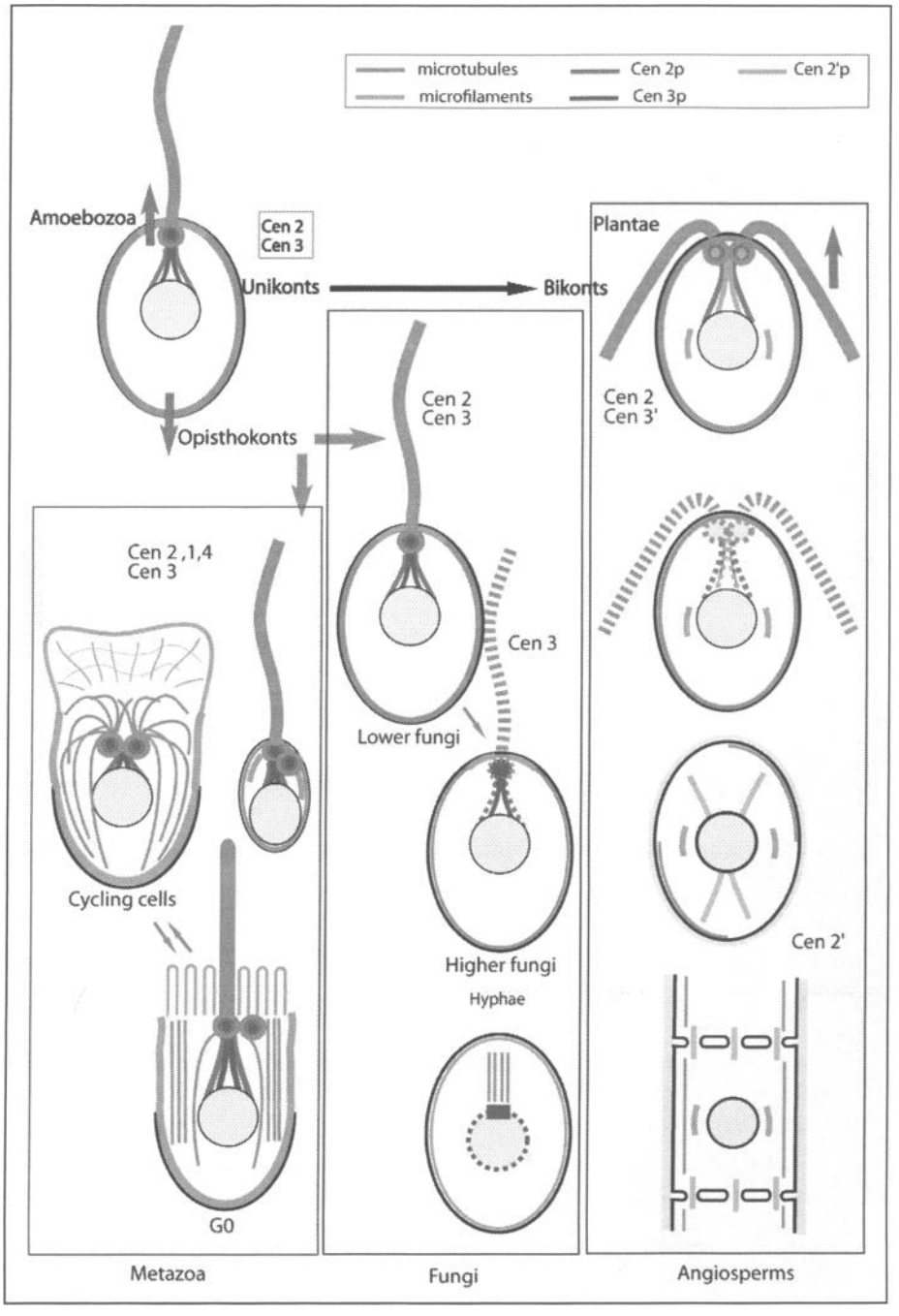


Figure 2. Cartoon tentatively illustrating a plausible scenario for the evolution of the connection between the basal body/axoneme and the nucleus in opisthokonts (animals and fungi) and in flowering plants (see text). Red arrows indicate the swimming direction. A color version of this figure is available online at www.Eurekah.com.

proteins have now been identified, including nuclear proteins participating in DNA repair such as XPC.[35,39-41,53-55] Centrin-binding proteins, which apparently participate in functions at the centrosome/SPB, include the yeast Sfi1 and related proteins in other species.[55] Sfi1p contains multiple internal repeats, each repeat being able to bind Cdc31p. In budding yeast, Sfi1p colocalises with Cdc31p to the half bridge, and apparently provides the half-bridge with an oriented back-bone structure where Sfi1p molecules are aligned side by side in the plane of the nuclear envelope, the N-termini being associated with the SPB whereas the C-termini form the other end of the half-bridge.[56] Remarkably, the half-bridge to bridge transition that precedes the formation of the satellite that nucleates the new SPB would correspond to the assembly of a new half-bridge having a mirror image structure with respect to the other half-bridge. Thus the first event in the SPB duplication event would be the duplication of the half bridge which connects the SPB to the nucleus. Before nucleating the assembly of a new SPB itself, it is the assembly of the new link to the nucleus periphery that would be triggered. This would ensure that the new SPB/centrosome is connected to the nucleus. Indeed, the SPB itself is inserted into the nuclear envelope in *S. cerevisiae*, but this is a late event in the process: the new SPB assembles in the cytoplasm before inserting into the nuclear envelope. In other systems, like the fission yeast *S. pombe*, the SPB maintains a cytoplasmic location, albeit very close to the nuclear envelope. In higher systems, the possibility that centriole nucleation could take place at the nuclear periphery has often been suggested in the past from observations in marine or amphibian eggs in which centriole assembly was artificially triggered.[57] Since Sfi1p has homologues in many divergent systems,[8,55] it will be important to analyse its precise localisation in each case, as a structure functionally equivalent to the yeast half-bridge could exist in other centrosomes, as proposed by Adams and Kilmartin.[58] One would expect this potentially conserved structure to participate in the centrosome-nucleus connection during interphase. Interestingly, hSfi1p binds both hCen2p and hCen3p in vitro. This could account for the fact that both centrins seem to be involved in mammalian centrosome duplication.[32,59] Like for centrin genes, *Drosophila* and *C. elegans* stand apart for centrin-binding proteins: their genomes do not contain identifiable *SFI1* homologue. Thus they must be compared with some caution with other systems: initiation of centrosome duplication might not involve quite the same molecular mechanisms than in other opisthokonts (see ref. 8). For the same reason, Ecdysozoa could help revealing the real function of Sfi1-centrin-containing structures in other systems.

Centrosome Has Evolved in Unikonts

If centrin requirement for centrosome reproduction is due to its direct participation in the structures and mechanisms that ensure nucleus-centrosome/SPB connection through cell division, the evolution of centrin proteins should reflect the evolution of centrin-containing structures that connect the motile apparatus to the nucleus in unicellular systems. A plausible scenario for the evolution of the connection between the basal body/axoneme and the nucleus in opisthokonts (animals and fungi) and in flowering plants is summarized in Figure 2. It involves a coevolution of the plasma membrane-associated actin system. Amoebozoa also evolved a centrosome and a cortical actin network. They are not included in this survey as many questions are still unanswered for this clade.

The ancestral eukaryote possessed one axoneme, one single centriole with a cone of root microtubules and the cellular machinery to form pseudopodia.[24] We assume the actin-dependent machinery for pseudopodia and for endo- or exocytosis was located at the base of the flagellum. This is the casein *Trypanosoma* where the flagellar pocket is the only site for endocytosis, most of the cell cortex being covered with a tightly packed corset of microtubules. Can we correlate the evolution of the two ancestral isoforms of centrin with the evolution of cell organization with respect to motile behaviour?

We assume that the CrCenp subfamily (Cen2p) is strictly correlated with the presence of the motile apparatus whereas ScCdc31p subfamily (Cen3p) has a more specific role in the association of the motile apparatus with the nucleus. Admittedly, this is an oversimplification

as centrin-binding proteins like Sfi1 can bind both types of centrin isoforms[55] suggesting another possibility, namely that they both participate in the same structures whose organisation could possibly depend on the ratio between centrin isoforms. However the absence of Sfi1 in the proteome of the *Chlamydomonas* flagellum[60] suggests that Sfi1 does not function with CrCenp in axoneme structure or activity, whereas the correlative loss of the ScCdc31p isoform and Sfi1 proteins observed in some clades (Ecdysozoa, seed plants, see below) suggests a common involvement of these two proteins in the same structures or functions. We postulate that these structures maintain the association of the motile apparatus with the nucleus. The CrCenp isoform can participate in these structures as well, as established in *C. reinhardtii*. We only postulate that the participation of ScCdc31p isoform brings a specific control on the association of the motile apparatus to the nucleus, by providing the connecting structures with properties whose precise nature is not important at that stage. ScCdc31p-like proteins could for example allow to establish specific regulatory cross-talk with the network involved in dynein-dependent nuclear positioning and migration.[13]

From this postulate, we interpret the high divergence of ScCdc31p subfamily in the unicellular algae (Cen3'p) as a release of functional constraints on this isoform reflecting a modification of the connection between the motile apparatus and the nucleus. This could be due to the transition from uni- to bi-flagellate organisation, to the leading position of the motile apparatus with respect to the direction of swimming, or to the formation of a cell wall. The wall formation would be accompanied by a reduction of the cortical actin network which has a limited role in this group compared to the others. During the evolution of angiosperms (right box in Fig. 2) the basal body/axoneme has been lost, thus releasing the constraints on Cen2p leading to the evolution of a new sub-family (Cen2'p), and the loss of connections between the cortex and the nucleus. The γ-tubulin-dependent machinery for microtubule nucleation is distributed everywhere on the cell cortex or on the nuclear envelope. The conserved gene network involved in dynein-dependent nuclear positioning and migration in fungi or during neuronal migration in humans[13] is lost together with the microtubule nucleating centrosome or SPB organelle.[61] Cell polarity and cell motility are lost altogether. Cell division does not take place by acto-myosin-dependent fission and cell individuation is not complete: the multicellular organism is a symplasm.

In the opisthokonts, the loss of the flagellum in the branch leading to higher fungi (central box in Fig. 2) is correlated with the loss of Cen2p. The basal body structure is lost altogether but the connection between a derived dominant microtubule nucleating organ and the nucleus in which Cen3p is playing a specific role is preserved together with centrin-binding proteins like Kar1p and Sfi1p. The absence of the flagellum released the SPB-microtubule network from its peripheral distribution. It takes a centralized organization, becoming a cytoplasmic microtubule network organized about a SPB/centrosome. The gene network involved in dynein-dependent nuclear positioning and migration ensures a dynamic microtubule-dependent connection between the more centrally located SPB-nucleus complex and the whole cortex where the actin system is now located and provides cells with polarity. The presence of a chitin wall precludes extensive development of the actin network and cell motility. However, complete cell division by acto-myosin-dependent fission can take place in unicellular yeasts. In the multicellular state, fungi form coenocytic hyphae or septate hyphae with large pores, in which cells are not fully separated.

In the branch leading to metazoa, both centrin isoforms are conserved. A more versatile cellular organization of the microtubule network with respect to the cell cortex is observed: reversible disassembly of the axoneme can take place without loosing the basal bodies. This allows the conversion between centrosome-associated centrioles and axoneme-associated basal bodies either in the male germ line at each generation, preserving swimming motility, or in somatic cells in a cell cycle-dependent manner to serve sensory or motile functions. The gene network involved in dynein-dependent nuclear positioning ensures a dynamic microtubule-dependent connection between the more centrally located centrosome-nucleus

complex and the whole cortex where the actin-dependent machinery has considerably developed, allowing cell migration on solid substrates or on other cells. The cross-talk between the actin and microtubule networks sets an elaborate mechanism for maintaining cell polarity, allowing a whole spectrum of cell organizations, including the neuronal polarity in which nuclear migration is critical for brain development. Ancestral swimming motility is preserved in the germ line or in somatic ciliated cells. Cell individuation by fission is complete. Animals form cohesive organisms by sequential divisions and have conserved all the functions associated with the basal body/axoneme of the ancestral unicellular eukaryote, namely sensation, motion and division.

Interestingly, evolutionary constraints on centrin proteins seem to have been relaxed in Ecdysozoa. The worm *C. elegans*, which has no recognizable centrin genes (Fig. 1), has lost any motile flagellum and has only conserved ciliated sensory neurones. The aflagellate sperm centrioles are reduced in size and organization but form a robust centrosome upon egg fertilization, and the centrosome is essential for triggering egg polarity and organizing successive egg divisions.[50,62] The fly *Drosophila*, which has only two CrCenp-like isoforms (Fig. 1), has both sensory neurons and flagellate sperm cells. However sperm cells do not swim actively: flagella are very long and are not compartmentalized.[63] The embryo centrosome is formed around a pair of short and incomplete centrioles. Interestingly centrosomes in *Drosophila* may not be as essential as they are in other species: acentriolar cell lines can be isolated[64] and more recently, the development of acentrosomal embryos has been demonstrated upon inactivation of the PLK4 homologue of Drosophila.[65]

Comparing the biogenesis of the basal body/centriole in Ecdysozoa and in other animals might give important insights into cell and tissue morphogenesis.

Conclusion

In addition to the microtubule and actin networks the two ancestral centrin sub-families and their interacting partners seem to represent a very ancient system linking the motile organ to the cell body and the nucleus. The CrCenp sub-family is required for axoneme activity. A specific participation of the ScCdc31p sub-family in the association between the motile apparatus and the nucleus could have been critical for the 'invention' of the centrosome/SPB during multicellularity. Its involvement in centrosome duplication would reflect the need for ensuring that the daughter centrosomes maintain or reestablish an association with the dividing nuclei during cell division.

A coevolution of the centrosome and of the cortical acto-myosin network has taken place with a profound impact on cell division and on the evolution of multicellular organisms. Only metazoa have conserved all the functions associated with the basal body/axoneme of the ancestral unicellular eukaryote. In Ecdysozoa however, the motile apparatus and its connection to the nucleus have undertaken a specific evolution.

References

1. Mitchell DR. Speculations on the evolution of 9+2 organelles and the role of central pair microtubules. Biol Cell 2004; 96(9):691-6.
2. Mitchell DR. The evolution of eukaryotic cilia and flagella as motile and sensory organelles. In: Gàspàr Jékely, ed. Origins ansd Evolution of Eukaryotic Endomembranes and Cytoskeleton. 2006, (Eureka.com).
3. Jékely G, Arendt D. Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. Bioessays 2006; 28(2):191-8.
4. Jékely G. Origin of eukaryotic endomembranes - A critical evaluation of different model scenarios. In: Gàspàr Jékely, ed. Origins and Evolution of Eukaryotic Endomembranes and Cytoskeleton. Austin: Landes Bioscience, 2007, (Eurekah.com).
5. Baroin A, Perasso R, Qu L et al. Partial phylogeny of the unicellular eukaryotes based on rapid sequencing of a portion of 28S ribosomal RNA. Proc Natl Acad Sci US 1988; 85(10):3474-8.
6. Gull K. Protist tubulins: New arrivals, evolutionary relationships and insights to cytoskeletal function. Curr Opin Microbiol 2001; 4(4):427-32.

7. Gull K. The biology of kinetoplastid parasites: Insights and challenges from genomics and post-genomics. Int J Parasitol 2001; 31(5-6):443-52.
8. Azimzadeh J, Bornens M. The centrosome in evolution. In: Nigg EA, ed. Centrosomes in Development and Disease. Weinheim: Wiley-VCH, 2004:93-122.
9. Bray D, Duke T. Conformational spread: The propagation of allosteric states in large multiprotein complexes. Annu Rev Biophys Biomol Struct 2004; 33:53-73.
10. Janicki C. Centrioles and kinetosomes: Form, function, and evolution. In: Chapman MJ, Colan MF, Margulis L, eds. The Quarterly Review of Biology. 2000; 75(4):409-29, (1915, quoted).
11. Salisbury JL. The lost neuromotor apparatus of Chlamydomonas rediscovered. J Protozool 1988; 35(4):574-7.
12. Bornens M. Is the centriole bound to the nuclear membrane? Nature 1977; 270(5632):80-2.
13. Morris NR. Nuclear migration. From fungi to the mammalian brain. J Cell Biol 2000; 148(6):1097-101.
14. Reinsch S, Gonczy P. Mechanisms of nuclear positioning. J Cell Sci 1998; 111:2283-2295.
15. Tanaka T, Serneo FF, Higgins C et al. Lis1 and doublecortin function with dynein to mediate coupling of the nucleus to the centrosome in neuronal migration. J Cell Biol 2004; 165(5):709-21.
16. Taxis C, Keller P, Kavagiou Z et al. Spore number control and breeding in Saccharomyces cerevisiae: A key role for a self-organizing system. J Cell Biol 2005; 171(4):627-40.
17. King N. The unicellularancestry of animal development. Dev Cell 2004; 7:313-325
18. Gross JD. Developmental decisions in Dictyostelium discoideum. Microbiol Rev 1994; 58(3):330-51.
19. Buss LW. Evolution, development, and the units of selection. Proc Natl Acad Sci USA 1983; 80:1387-1391.
20. Hoops HJ, Witman GB. Basal bodies and associated structures are not required for normal flagellar motion or phototaxis in the green alga Chlorogonium elongatum. J Cell Biol 1985; 100(1):297-309.
21. Bastin P, Pullen TJ, Sherwin T et al. Protein transport and flagellum assembly dynamics revealed by analysis of the paralysed trypanosome mutant snl-1. J Cell Sci 1999; 112(Pt 21):3769-77.
22. Gull K. Host-parasite interactions and trypanosome morphogenesis: A flagellar pocketful of goodies. Curr Opin Microbiol 2003; 6(4):365-70.
23. Cavalier-Smith T. Only six kingdoms of life. Proc Roy Soc Lond 2004; B271:1251-1262.
24. Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005; 436(7054):1113-8.
25. Brickmann H, Philippe H. The diversity of eukaryotes and the root of the eukaryotic tree. In: Gàspàr Jékely, ed. Origins and Evolution of Eukaryotic Endomembranes and Cytoskeleton. Austin: Landes Bioscience, 2006, (Eurekah.com).
26. Hartman H, Fedorov A. The origin of the eukaryotic cell: A genomic investigation. Proc Natl Acad Sci US 2002; 99(3):1420-5, (Erratum in: Proc Natl Acad Sci USA. 2002; 99(7):4752).
27. Salisbury JL, Baron AT, Sanders MA. The centrin-based cytoskeleton of Chlamydomonas reinhardtii: Distribution in interphase and mitotic cells. J Cell Biol 1988; 107(2):635-41.
28. Geimer S, Melkonian M. The ultrastructure of the Chlamydomonas reinhardtii basal apparatus: Identification of an early marker of radial asymmetry inherent in the basal body. J Cell Sci 2004; 117(13):2663-74.
29. Huang B, Watterson DM, Lee VD et al. Purification and characterization of a basal body-associated Ca(2+)-binding protein. J Cell Biol 1988; 107:121-131.
30. Spang A, Courtney I, Grein K et al. The Cdc31p-binding protein Kar1p is a component of the half bridge of the yeast spindle pole body. J Cell Biol 1995; 128(5):863-77.
31. Paoletti A, Bordes N, Haddad R et al. Fission yeast cdc31p is a component of the half-bridge and controls SPB duplication. Mol Biol Cell 2003; 14(7):2793-808.
32. Salisbury JL, Suino KM, Busby R et al. Centrin-2 is required for centriole duplication in mammalian cells. Curr Biol 2002; 12(15):1287-92.
33. Koblenz B, Schoppmeier J, Grunow A et al. Centrin deficiency in Chlamydomonas causes defects in basal body replication, segregation and maturation. J Cell Sci 2003; 116(13):2635-46.
34. Taillon BE, Adler SA, Suhan JP et al. Mutational analysis of centrin: An EF-hand protein associated with three distinct contractile fibers in the basal body apparatus of Chlamydomonas. J Cell Biol 1992; 119(6):1613-24.
35. Araki M, Masutani C, Takemura M et al. Centrosome protein centrin 2/caltractin 1 is part of the Xeroderma Pigmentosum group C complex that initiates global genome nucleotide excision repair. J Biol Chem 2001; 276(22):18665-72.
36. Fischer T, Rodriguez-Navarro S, Pereira G et al. Yeast centrin Cdc31 is linked to the nuclear mRNA export machinery. Nat Cell Biol 2004; 6(9):840-8.

37. He CY, Pypaert M, Warren G. Golgi duplication in Trypanosoma brucei requires Centrin2. Science 2005; 310(5751):1196-8.
38. Molinier J, Ramos C, Fritsch O et al. CENTRIN2 modulates homologous recombination and nucleotide excision repair in Arabidopsis. Plant Cell 2004; 16(6):1633-43.
39. Nishi R, Okuda Y, Watanabe E et al. Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. Mol Cell Biol 2005; 25(13):5664-74.
40. Pulvermuller A, Giessl A, Heck M et al. Calcium-dependent assembly of centrin-G-protein complex in photoreceptor cells. Mol Cell Biol 2002; 22:2194-2203.
41. Sullivan DS, Biggins S, Rose MD. The yeast centrin, Cdc31p, and the interacting protein kinase, Kic1p, are required for cell integrity. J Cell Biol 1998; 143:751-765.
42. Lee VD, Huang B. Molecular cloning and centrosomal localization of human caltractin. Proc Natl Acad Sci USA 1993; 90(23):11039-43.
43. Errabolu R, Sanders MA, Salisbury JL. Cloning of a cDNA encoding human centrin, an EF-hand protein of centrosomes and mitotic spindle poles. J Cell Sci 1994; 107(Pt 1):9-16.
44. Wolfrum U, Salisbury JL. Expression of centrin isoforms in the mammalian retina. Exp Cell Res 1998; 242(1):10-7.
45. Hart PE, Glantz JN, Orth JD et al. Testis-specific murine centrin, Cetn1: Genomic characterization and evidence for retroposition of a gene encoding a centrosome protein. Genomics 1999; 60(2):111-20.
46. Laoukili J, Perret E, Middendorp S et al. Differential expression and cellular distribution of centrin isoforms during human ciliated cell differentiation in vitro. J Cell Sci 2000; 113(Pt 8):1355-64.
47. Gavet O, Alvarez C, Gaspar P et al. Centrin4p, a novel mammalian centrin specifically expressed in ciliated cells. Mol Biol Cell 2003; 14(5):1818-34.
48. Middendorp S, Paoletti A, Schiebel E et al. Identification of a new mammalian centrin gene, more closely related to Saccharomyces cerevisiae CDC31 gene. Proc Natl Acad Sci US 1997; 94(17):9141-6.
49. Ribichich KF, Gomes SL. Blastocladiella emersonii expresses a centrin similar to Chlamydomonas reinhardtii isoform not found in late-diverging fungi. FEBS Lett 2005; 579(20):4355-60.
50. Leidel S, Gonczy P. Centrosome duplication and nematodes: Recent insights from an old relationship. Dev Cell 2005; 9(3):317-25.
51. Grunow A, Lechtreck KF. Mitosis in Dunaliella bioculata (Chlorophyta): Centrin but not basal bodies are at the spindle poles. Journal of Phycology 2001; 37(6):1030-1043.
52. Daunderer C, Schliwa M, Graf R. Dictyostelium centrin-related protein (DdCrp), the most divergent member of the centrin family, possesses only two EF hands and dissociates from the centrosome during mitosis. Eur J Cell Biol 2001; 80(10):621-30.
53. Jaspersen SL, Giddings Jr TH, Winey M. Mps3p is a novel component of the yeast spindle pole body that interacts with the yeast centrin homologue Cdc31p. J Cell Biol 2002; 159(6):945-56.
54. Biggins S, Rose MD. Direct interaction between yeast spindle pole body components: Kar1p is required for Cdc31p localization to the spindle pole body. J Cell Biol 1994; 125(4):843-52.
55. Kilmartin JV. Sfi1p has conserved centrin-binding sites and an essential function in budding yeast spindle pole body duplication. J Cell Biol 2003; 162(7):1211-21.
56. Li S, Sandercock AM, Conduit P et al. Structural role of Sfi1p-centrin filaments in budding yeast spindle pole body duplication. J Cell Biol 2006; 173(6):867-77.
57. Kallenbach RJ, Mazia D. Origin and maturation of centrioles in association with the nuclear envelope in hypertonic-stressed sea urchin eggs. Eur J Cell Biol 1982; 28(1):68-76.
58. Adams IR, Kilmartin JV. Spindle pole body duplication: A model for centrosome duplication? Trends Cell Biol 2000; 10(8):329-35.
59. Middendorp S, Kuntziger T, Abraham Y et al. A role for centrin 3 in centrosome reproduction. J Cell Biol 2000; 148(3):405-16.
60. Pazour GJ, Agrin N, Leszyk J et al. Proteomic analysis of a eukaryotic cilium. J Cell Biol 2005; 170(1):103-13.
61. Lawrence C, Morris, Meagher R et al. Dyneins have run their course in plant lineage. Traffic 2001; 2(5):362-363.
62. Cowan CR, Hyman AA. Centrosomes direct cell polarity independently of microtubule assembly in C. elegans embryos. Nature 2004; 431(7004):92-6.
63. Avidor-Reiss T, Maer AM, Koundakjian E et al. Decoding cilia function: Defining specialized genes required for compartmentalized cilia biogenesis. Cell 2004; 117(4):527-39.
64. Szollosi A, Ris H, Szollosi D et al. A centriole-free Drosophila cell line. A high voltage EM study. Eur J Cell Biol 1986; 40(1):100-4.
65. Bettencourt-Dias M, Rodrigues-Martins A, Carpenter L et al. SAK/PLK4 is required for centriole duplication and flagella development. Curr Biol 2005; 15(24):2199-207.

CHAPTER 11

The Evolution of Eukaryotic Cilia and Flagella as Motile and Sensory Organelles

David R. Mitchell*

Abstract

Eukaryotic cilia and flagella are motile organelles built on a scaffold of doublet microtubules and powered by dynein ATPase motors. Some thirty years ago, two competing views were presented to explain how the complex machinery of these motile organelles had evolved. Overwhelming evidence now refutes the hypothesis that they are the modified remnants of symbiotic spirochaete-like prokaryotes, and supports the hypothesis that they arose from a simpler cytoplasmic microtubule-based intracellular transport system. However, because intermediate stages in flagellar evolution have not been found in living eukaryotes, a clear understanding of their early evolution has been elusive. Recent progress in understanding phylogenetic relationships among present day eukaryotes and in sequence analysis of flagellar proteins have begun to provide a clearer picture of the origins of doublet and triplet microtubules, flagellar dynein motors, and the 9+2 microtubule architecture common to these organelles. We summarize evidence that the last common ancestor of all eukaryotic organisms possessed a 9+2 flagellum that was used for gliding motility along surfaces, beating motility to generate fluid flow, and localized distribution of sensory receptors, and trace possible earlier stages in the evolution of these characteristics.

Evidence for the Presence of a 9+2, Motile, Sensory Organelle in the Last Common Eukaryotic Ancestor

As summarized in Figure 1, typical cilia and flagella (hereafter called flagella, there being no consistent structural or functional difference between organelles with these two designations) are motile projections oriented perpendicular to the cell surface, but they vary in length, in number per cell, and in the patterns of motility that they produce. They are composed of a cylinder (the axoneme) of nine doublet microtubules surrounding two single microtubules and are covered by the cell membrane. Between each pair of flagellar doublets are rows of axonemal dynein ATPases, which power the bending of these organelles, and extending toward the center of the cylinder are radial spokes, which touch upon a central apparatus and regulate axonemal dyneins. This central element consists of two single microtubules, assembled from a unique nucleating site,[1] plus many interconnecting microtubule-associated proteins (reviewed in ref.

*Corresponding Author: David R. Mitchell—Department of Cell and Developmental Biology, SUNY Upstate Medical University, 750 E. Adams St., Syracuse, New York 13210, U.S.A. Email: mitcheld@upstate.edu

Eukaryotic Membranes and Cytoskeleton: Origins and Evolution, edited by Gáspár Jékely.

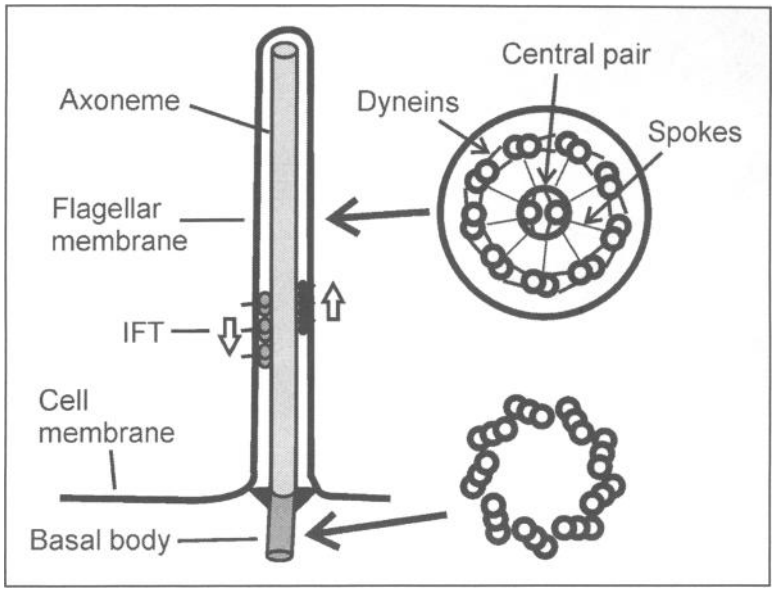


Figure 1. Diagram of structures common to all motile cilia and flagella. Longitudinal view to the left shows the relationship between the axoneme and basal body, and the location of intraflagellar transport (IFT) motors between axonemal doublet microtubules and the flagellar membrane. Transition fibers attached to the basal body separate the flagellar membrane domain from the rest of the cell membrane. Cross sectional views to the right show structures in flagella, including the nine outer doublet and two single central pair microtubules (top) and the nine triplet microtubules of basal bodies (bottom).

2). Together they form a structure that provides a cylindrical surface apposing the ends of the radial spokes.[3] The nine doublets assemble from a much shorter cylinder of nine triplet microtubules, the basal body or centriole, which is anchored to the cell surface and stabilized in the cytoplasm by other cytoskeletal elements. Basal bodies that anchor flagella are often interchangeable during the cell cycle with centrioles,[4] and these two names should be considered as two functional descriptions for the same structure. Between the doublets and the membrane are particles associated with intraflagellar transport (IFT), a process important for flagellar assembly and protein trafficking in this cellular compartment.[5,6] In the outward (anterograde) direction, IFT is powered by kinesins of the kinesin2 family; in the inward (retrograde) direction, power is provided by dyneins of the cytoplasmic dynein 2 (DHC1b) family. Many flagella act as sensory antennae through localization of receptors to the flagellar membrane. In extreme cases, termed primary cilia or sensory cilia, the motile function has been discarded and with it the dyneins, radial spokes, the central pair complex, and other proteins needed for bend formation. IFT is still required for the assembly and maintenance of these primary cilia, which play important sensory roles in metazoan organisms.[6]

Some attempts have been made in the past to identify intermediates in the evolution of flagella by looking within existing branches of eukaryotes for organisms that may have diverged before the complete 9+2 flagellum had evolved. However, improved methods of analysis and the recent burst of sequence data are rapidly transforming long-held views of eukaryotic phylogeny to new schemes in which there are many branches that diverged within a relatively short period of time[7,8] (Fig. 2). Many of these branches are represented today by single-celled protists, so that the true diversity of eukaryotes cannot be appreciated without some understanding of the relationships among these often less-studied organisms. Thus the nearest

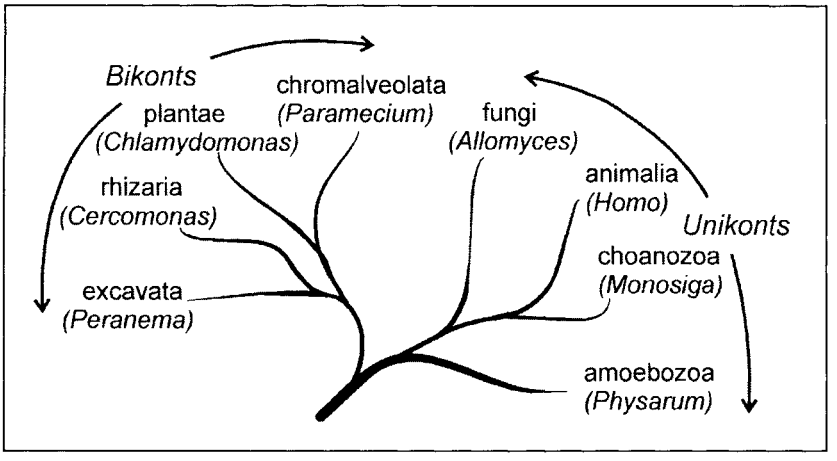


Figure 2. Diagram of probable evolutionary divergence that generated all existing branches of eukaryotic organisms. Under the name of each branch or clade is a the name of a representative genus in that clade that contains species with typical motile 9+2 flagella. Based on recent studies of rare gene fusion events, as well as more traditional sequence comparisons, the entire tree is divided into two superclades, unikonts and bikonts.

relatives of animals (metazoans) are single-celled choanoflagellates.[9,10] Fungi (some of which were once considered primitive because of their simplicity) turn out to be another twig of this same branch, the opisthokonts or unikonts. *Allomyces*, a chytridiomycete fungus with flagellated gametes and zoospores, is just one example of a fungal cell that swims, like a sperm, with a typical 9+2 flagellum. Amoebozoa, containing such amoeboflagellates as *Physarum*, are probably (based on a shared gene fusion and on mitochondrial sequences) the only other unikont twig, and branch somewhat earlier than fungi.[10-13] The other major superclade of eukaryotes, the bikonts, encompasses a great variety of flagellated and amoeboid organisms, including green plants and green and red algae (plantae), ciliates, dinoflagellates and their kin (chromalveolata), euglenids, trypanosomatids, diplomonads and their sister taxa (excavata), and the radiolaria, cercozoa, etc. (rhizaria).[14,15] In the resulting tree (Fig. 2), one should note that most of the model organisms under intense study during the past twenty years reside on one branch (unikont), but fortunately for studies of flagellar evolution, additional attention has been focused on a few flagellated bikont organisms, most especially the green alga *Chlamydomonas reinhardtii*.

One striking conclusion of these recent phylogenetic studies is that every extant branch of eukaryotes includes organisms with motile, 9+2 flagella. Even proteins of the central pair apparatus, such as products of the *Chlamydomonas PF6, PF16, PF20, KLP1*, and *CPC1* genes, have been conserved between algae and humans.[16] From this we can only conclude that these organelles had evolved prior to the divergence of all extant eukaryotic clades from a common ancestor. In addition, IFT proteins, which are central to flagellar assembly and to the display of sensory receptors and flagellar surface motility, are also present in flagella from distant branches of eukaryotic phylogeny (e.g., trypanosomes,[17] insects,[18] and green algae[19]), and therefore must have evolved prior to the beginnings of eukaryotic radiation.[18] The microtubule rootlet structures that stabilize basal bodies in the cytoplasm do vary phylogenetically and therefore the nature of those that might have been present in the last common eukaryotic ancestor are difficult to determine,[20,21] but all of the elements of triplet microtubules, and the accessory proteins needed for basal body formation, must also have been present at the base of this tree.

Eukaryotes likely developed the nucleus, endomembrane system, and cytoskeleton, and then used the phagocytic ability that was provided by the combined cytoskeletal and

endomembrane systems to obtain the precursors to mitochondria, long before the evolution of flagella. The framework of doublet microtubules upon which flagella depend must have evolved from simpler single microtubules which, as essential elements of the mitotic machinery, would have been needed to segregate a genome enclosed in a nucleus. Likewise, dyneins as microtubule-based motors undoubtedly functioned as transport motors on cytoplasmic and mitotic microtubules long before their use was adapted to flagella. If all of the essential elements of eukaryotic cells were in place for so long before the advent of flagella, one must ask why there are no branches of existing eukaryotes that lack flagella. The simplest explanation is that the branch of early eukaryotes that first developed a functional 9+2 flagellum possessed a tremendous selective advantage over its competitors, and was the only eukaryote whose descendants survive today.

Evolution of Tubulin, Dynein and Kinesin

The closest prokaryotic relative of tubulins, FtsZ, functions during bacterial septation, and FtsZ homologs continue to perform a similar role in chloroplasts and perhaps some mitochondria.[22-24] In early eukaryotes, FtsZ gene duplication and modification led to alpha and beta tubulin, which form a stable dimer that retains FtsZ properties such as polymerization, GTP-binding, and GTP hydrolysis-dependent conformational change, but which gained the added ability to interact laterally to form tubes. Gamma tubulin is also ubiquitous and likely emerged early, to function as a nucleating site that helps determine microtubule polarity and distribution. Additional tubulin isoforms delta and epsilon are ubiquitous among organisms with triplet microtubules, and the formation of triplet microtubules, essential for the function of basal bodies, has been shown to require both delta and epsilon tubulin in *Chlamydomonas*[25,26] and both epsilon[27] and the less ubiquitous eta tubulin[28] in *Paramecium*. The presence of these tubulin isoforms in members of both the unikont and bikont clades[17,29] argues for their evolution prior to the divergence of eukaryotes.

The dynein motors that power both flagellar beating and retrograde IFT movements are members of the superfamily of AAA ATPases. In dyneins, six individual AAA domains have become fused into a single large polypeptide,[30] but only four of these six domains in dyneins retain the signature sequences of nucleotide binding pockets.[31] DNA pumping ATPases of archaea (*HerA*) and bacteria (*FtsK*), also AAA ATPases, are needed during prokaryotic cell division for correct daughter chromosome segregation,[32] and it would be tempting to assume that an interaction between FtsK and FtsZ could have evolved directly into an interaction between dynein and tubulin. However, dyneins apparently evolved from an entirely different branch of the AAA superfamily from the DNA pumping ATPases. The closest eukaryotic relatives of dyneins are midasins, similarly giant eukaryotic AAA ATPases, which function at the nuclear pore in 60S ribosome export. The closest prokaryotic homologs of dyneins and midasins are members of the MoxR family, single AAA domain ATPases that function as chaperones in the assembly of large protein complexes, such as methanol dehydrogenase and nitric oxide reductase.[30] It would thus appear that dyneins evolved as microtubule motors in the early eukaryotic lineage, and that their prokaryotic ancestors were proteins that performed conformational work linked to ATP binding and hydrolysis.

Sequence comparisons among extant dynein heavy chains divide dyneins into two broad families, cytoplasmic and axonemal.[33] Most organisms have only two cytoplasmic dyneins, one devoted to general cytoplasmic microtubule-based movements, present in all eukaryotes, and one for retrograde IFT movement on axonemal microtubules, absent from organisms such as *Saccharomyces cerevisiae* that lack axonemes. Each is thought to function as a homodimer of catalytic heavy chains. Organisms with motile flagella also have a large family of axonemal dyneins that can be further divided into three subfamilies, outer row dyneins, I1 inner row dyneins, and additional diverse inner row dyneins (reviewed in ref. 34). Outer row dyneins diverged before the common ancestor into two isoforms (alpha and beta) that form heavy chain heterodimers. A third isoform that diverged more recently is found in *Chlamydomonas*

(plantae) and *Tetrahymena* (chromalveolata) but not in sea urchins or fruit flies (animalia). Outer row dyneins bind in a continuous row with 24 nm spacing (every third 8 nm tubulin dimer), whereas inner row dynein isoforms occur once every 96 nm along each doublet microtubule.[35] The I1 inner row dynein is a typical heterodimer of two heavy chain subunits, and both isoform subfamilies are represented in the genomes of all organisms that retain motile flagella. I1 dyneins appear to have become established early as major targets of signal-dependent regulation of flagellar bending parameters.[36] The many additional inner row dyneins present in, for example, sea urchins and ciliates, appear to have diverged more recently; several of the ciliate inner row dyneins are more closely related to each other than to any of the urchin inner row dyneins. Structural, genetic and biochemical analyses in *Chlamydomonas* indicate that these additional inner row dyneins function as monomers, rather than the dimers typical of all other dyneins, and some isoforms may be differentially distributed along the length of the organelle.[37] Their sequence relationships suggest that the last common eukaryote may have had a single isoform of this monomeric inner row dynein.

Among the members of the very large and diverse superfamily of kinesin ATPases are at least two families with members that function as flagellar proteins. The small kinesin9 family is represented by a sequence expressed in ciliated cells of mammals[38] and by the *Chlamydomonas* Klp1 protein.[39] Klp1 has been localized to the central pair complex[39] and shown to be important for normal flagellar motility in that organism.[40] Although evidence for the role of kinesin9 members in mammalian cells is not available, their expression patterns support an early evolving flagellar function for this protein. The kinesin2 family is larger and functionally more diverse. While some members of the kinesin2 family are anterograde motors for IFT, others are anterograde motors in other cytoplasmic compartments such as neurons of metazoans. The presence of kinesin2 homologs in ciliates and flagellates, but not in nonflagellated fungi, suggests that kinesin2 coevolved with axonemes and was only coopted for other transport functions in recent metazoan evolution.[6]

The Origins of 9+2 Flagella

Given the evidence summarized above, the last common eukaryotic ancestor had a motile 9+2 flagellum, was anchored on a basal body of triplet microtubules, and required IFT for its assembly. How did such a complex system evolve? Clearly it must have been preceded by a microtubule cytoskeleton with dynein and kinesin motors. Strong arguments have been made that the driving force for the evolution of a microtubule cytoskeleton and its associated motor proteins was the ability to accurately segregate a large, nuclear membrane-enclosed genome by mitosis. Although the complicated checks and balances used to assure mitotic fidelity vary widely in extant eukaryotes, some aspects of mitosis have been sufficiently conserved and are so central to the process that they must have been present at an early stage. Other aspects of mitosis, assumed to be ancient because of their simplicity, reveal a minimal apparatus but may not reflect the ancestral condition. Although model organisms such as fungi have provided many details of such minimal systems, the tremendous variety of extant mitotic mechanisms (as reviewed in ref. 41) should not be forgotten. Mitosis requires a microtubule organizing center (MTOC) that duplicates once per cell cycle, a connection between each chromatid and one of the duplicated MTOCs, and separation of the MTOCs with their associated chromatids. These MTOCs vary structurally from the simple nuclear membrane-embedded spindle pole bodies of some unicellular organisms to the complex centriole-containing centrosomes of many metazoan cells. Most organisms form two microtubule arrays during mitosis, one oriented toward the chromosomes to link each chromatid to its MTOC, and a second that assembles between the two MTOCs to form a scaffold for MTOC separation. In G_1 phase of the cell cycle, prior to DNA and MTOC duplication, the primitive cell would have had a single MTOC, with a single array of microtubules directed away from the nucleus that defined the polarity of the cell. It is this cytoplasmic microtubule array that most likely provided the raw material for evolution of the flagellar axoneme.[42-44]

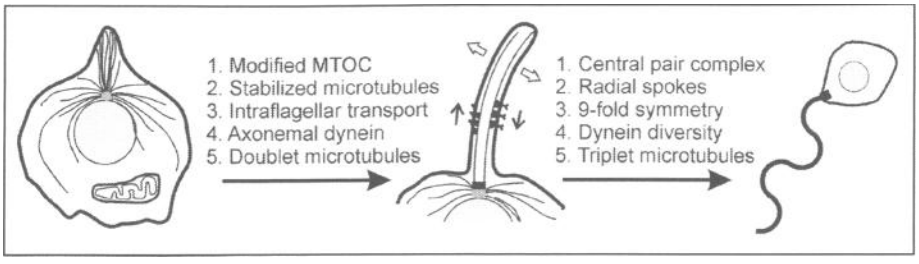


Figure 3. Proposed steps in the transition from an early eukaryote, with a polarized morphology based on asymmetric placement of a microtubule organizing center, but lacking flagella (left), through an intermediate with a protoflagellum that supported gliding and limited bending (center), to the last common eukaryotic ancestor, with a fully developed, motile 9+2 flagellum (right).

A polarized array of microtubules projecting from one side of the nucleus, as seen in most cells today, need only become linked into a bundle to provide an organelle that could distort the cell membrane and form a protoflagellum (Fig. 3). Microtubules radiating from the MTOC that were not incorporated into this protoflagellar bundle would continue to provide a cytoskeleton for general cytoplasmic transport and organization of the endomembrane system, and the interaction of motors such as kinesin and dynein with this microtubule cytoskeleton would provide directed motility of associated vesicles. Such movement along the protoflagellar bundle could direct exocytosis and endocytosis to a specific region of the cell membrane, creating a new membrane domain. The similarity of IFT proteins to proteins involved in vesicle trafficking[18] and the similarity of IFT kinesin and dynein to cytoplasmic versions of these motors, argues that axonemes evolved from proteins that were already in use in microtubule-based vesicular transport systems.

Early eukaryotes, lacking any other means of locomotion, were presumably benthic amoeboid cells and could not yet swim. Microtubule-based motors moving along a parallel bundle of microtubules in the protoflagellum would have provided at least two specific advantages to these organisms. Simple coupling of retrograde movement to transmembrane proteins would convert the protoflagellum to a feeding organelle, bringing particles that adhered to its surface back toward the cell body by retrograde IFT for subsequent phagocytosis. Alternatively, if substrate adhesion through IFT-associated protoflagellar transmembrane proteins was strong, and cell body adhesion proportionately weak, retrograde IFT could support gliding motility. Flagellar gliding as a means of locomotion is common in many pelagic, benthic and soil flagellates today, and such surface motility has also been described for metazoan cilia, suggesting that it was either an early adaptation of the IFT system, or one that has happened repeatedly during subsequent evolution.[44] As the coupling mechanisms between flagellar adhesion molecules and gliding motors have not been widely studied, their evolutionary history remains unknown.

Movement of secretory vesicles along a polarized microtubule bundle would also provide a polarized distribution of cell surface molecules, including receptors. While the localization of receptors to ciliary and flagellar surfaces has been documented in both protists and in metazoan sensory cilia, including chemosensory cilia in *C. elegans*, kidney cilia in vertebrates, and the highly modified cilia in sensory neurons of the vertebrate retina, evidence that specific receptors have been localized to this membrane domain since before eukaryotic divergence is only fragmentary (reviewed in ref. 44). Receptor localization does not require the complex structure of a flagellum, but the ability to define a polarized cell surface domain for sensory signaling may have been one of the driving forces in early flagellar evolution.

Flagella are first and foremost organelles that beat, and their complex structure would not have evolved without strong selection for motility, as witnessed by the rapid loss of much of

this complexity in nonmotile sensory cilia, and the complete loss of flagella in nonmotile organisms such as yeasts. However strong this advantage of a beating flagellum might have been to early eukaryotes, intermediate stages must have existed that provided intermediate levels of motion; a sudden jump from a benthic, amoeboid or gliding organism to one that can swim by flagellar beating is not plausible. To understand the advantages of less vigorous bending motility, one need only look at flagellar function among existing flagellates. Many organisms use flagella to generate feeding currents that increase the frequency with which food particles (bacteria, other eukaryotes, or detritus) can be ingested, while others use flagellar movements to aid in trapping food particles.[45,46] Feeding currents are common in organisms such as choanoflagellates, which attach to the substrate with stalks and use flagellar beating to create currents past the stationary cell, in mastigamoebae, which create currents while continuing to move by amoeboid activity, and in many biflagellates such as bodonids, which create currents with an anterior flagellum and glide on a posterior flagellum. Even a modest ability to vibrate or wave a protoflagellum could have provided the initial selective advantage that drove further development of single microtubules into doublet microtubules, and favored diversification of dyneins that could take advantage of this new doublet microtubule track (Fig. 3).

Along with increased motility came the increased need to anchor the axoneme, which may have driven both centriole/basal body evolution and the development of links between the flagellar base and the cell membrane. These links would segregate the flagellar compartment from the rest of the cytoplasm, requiring further refinements in the IFT sorting/trafficking mechanism so that standard vesicle fusion occurred at the flagellar base, and IFT movement transported both membrane and nonmembrane components to and from the flagellar compartment. In addition, such links create a boundary between the flagellar membrane and the rest of the plasma membrane, sequestering receptors and adhesion molecules into a unique membrane domain.

The ability to bend does not require an axoneme with 9-fold symmetry, and many axonemes have been discovered that depart from this pattern, yet most of these departures appear to be more recent modifications of an ancestral 9+2 pattern. As argued in more detail elsewhere, I propose that motility regulation by a central apparatus provided a strong selective advantage to the organism in which it evolved, and that the most successful regulatory mechanism was based on an apparatus built on a scaffold of two central microtubules, with regulatory signals transmitted through radial spokes of a defined length.[44,47] The geometry of this regulatory mechanism presumably favors an outer cylinder of precisely nine doublet microtubules, with the distance between doublets determined by the reach of dyneins that must span each interdoublet gap. Whether central pair regulation was based on a fixed central pair orientation, as found today in metazoans such as bivalves,[48] sea urchins,[49] and ctenophores[50] and possibly in excavates such as euglenids,[51] or a rotating central pair,[52] as commonly found in green algae such as *Chlamydomonas*[53] and *Micromonas*,[54] and chromalveolates such as *Paramecium*[55] and *Synura*,[56] remains to be determined. Central pair rotation may allow regulation of bends in different beat planes, and therefore be a more flexible regulatory system for organisms whose survival is most highly dependent on rapid changes in flagellar beat parameters.[47,53] The origin of radial spokes remains, at this time, one of the greater mysteries of flagellar evolution, as related proteins have not been identified in other microtubule-associated regulatory complexes.

Diversification of Flagellar Structure and Function during Eukaryotic Radiation

Many changes to the basic (if complex) 9+2 flagellum that was present in the last common eukaryotic ancestor are seen in some present day organisms, whereas others appear to have retained the original model with few alterations. Changes include additions, such as mastigonemes that project from the membrane surface and increase effective hydrodynamic resistance, and accessory structures that increase axoneme stiffness (paraflagellar rods in

euglenoids, outer dense fibers or extra microtubules in spermatozoa; for a more extensive survey of structural changes in spermatozoa, see Baccetti, ref. 57). Simplifications or deletions of structures no longer used by an organism or cell type include loss of the outer row dyneins, loss of the central apparatus and radial spokes (motile 9+0), loss of the central apparatus, radial spokes, and some doublet microtubules (6+0, 3+0). In the case of strictly nonmotile sensory cilia, all components necessary for motility (central pair, spokes, dyneins) may be absent, leaving only the membrane and the nine doublet microtubules to support IFT and receptor localization. Motile flagella that lack the central pair and radial spokes appear to have a simplified bending pattern that only accommodates helical bending waves, and are found on parasitic flagellates that do not depend on flagellar motility for locomotion in a complex environment,[58] on nodal cilia in early vertebrate embryos, where they generate a unidirectional fluid current essential for establishing left-right asymmetry,[59] and on certain vertebrate spermatozoa.[60] Some insect sperm that lack central pair microtubules have cylinders of 12 or 14 doublets, or spirals of hundreds of doublets,[57] suggesting further that the original standard of a cylinder of nine doublet microtubules was selected to accommodate the radial spoke-central pair regulatory complex, and that if the central pair is not needed, successful axonemes can evolve with alternative doublet patterns.

More difficult to catalog are modifications of existing parts to meet new demands, often observable not at the structural level but as differences in average length, beat frequency, or waveform, or as the ability to change beat frequency, beat direction, or waveform in response to signaling cascades. These signaling pathways may in turn begin with stimulation of flagellar surface receptors, or with cascades that are transmitted from elsewhere in the cell. The evolution of some of these changes, especially those affecting signaling pathways, will be difficult to trace until many more genomes have been sequenced, and proteomic analysis confirms the location and function of putative flagellar gene products. Some changes probably occurred only once, and surveying the distribution of organisms that retain such features may clarify phylogenetic relationships. Other changes have occurred independently more than once and can be considered convergent evolution. For example, the ability to generate ATP within the flagellar compartment has evolved in several independent ways. Glycolysis is an important source of energy for sperm motility in mammals[61] where fermentable sugars come directly from seminal fluid,[61,62] but has not been reported in vertebrate cilia and flagella other than sperm tails and some types of nonmotile cilia (e.g., the outer segments of mammalian photoreceptor cells[63]). Mammalian sperm-specific isoforms of glycolytic enzymes such as glyceraldehyde 3-phosphate dehydrogenase[64] and enolase[65] have likely evolved recently, as the glycolytic enzymes identified in *Chlamydomonas* flagella[66] are not closely related and appear to have been targeted for flagellar use specifically in algae. Completely different methods of flagellar ATP generation occur in other organisms, including phosphocreatine/creatine phosphokinase shuttles in sea urchin[67] and mammalian[68] spermatozoa, and in chicken photoreceptor outer segments,[69] and a phosphoarginine/arginine phosphokinase shuttle in *Paramecium* cilia.[70] Overall, the basic motile machinery of 9+2 organelles appears to be ancient and highly conserved, whereas the signaling cascades that regulate motility and accessory structures that modify its output have changed to suit specific organismal needs.

Summary

Typical 9+2 flagella likely evolved from bundling and extension of cytoplasmic microtubules that assembled on a microtubule organizing center and that generated a polarized cellular morphology. Close apposition of the plasma membrane created a separate membrane domain that could be used to localize receptors for sensory signal transduction, and required simultaneous evolution of intraflagellar transport to maintain this polarized structure. Membrane-associated IFT-based movement provided a mechanism for gliding motility, and addition of axonemal dynein motors allowed this extension to bend and generate currents past the cell. Formation of doublet microtubules allowed elaboration of dyneins to improve

motility, and the addition of the radial spoke-central pair regulatory system provided responsive dynein control. The strong selective advantage of a motile 9+2 flagellum may have resulted in rapid diversification of the last common eukaryotic ancestor into all existing branches of eukaryotic organisms.

Acknowledgements

Robert Bloodgood contributed through thoughtful discussions on flagellar gliding, David Asai shared recent results on dynein evolution, and Gaspar Jekely provided insight into the evolutionary connections between IFT and vesicle trafficking.

References

1. McKean PG, Baines A, Vaughan S et al. Gamma-tubulin functions in the nucleation of a discrete subset of microtubules in the eukaryotic flagellum. Curr Biol 2003; 13(7):598-602.
2. Smith EF, Yang P. The radial spokes and central apparatus: Mechano-chemical transducers that regulate flagellar motility. Cell Motil Cytoskeleton 2004; 57:8-17.
3. Mitchell DR. Reconstruction of the projection periodicity and surface architecture of the flagellar central pair complex. Cell Motil Cytoskeleton 2003; 55:188-99.
4. Beisson J, Wright M. Basal body/centriole assembly and continuity. Curr Opin Cell Biol 2003; 15(1):96-104.
5. Rosenbaum J. Intraflagellar transport. Curr Biol 2002; 12(4):R125.
6. Scholey JM. Intraflagellar transport. Annu Rev Cell Dev Biol 2003; 19:423-43.
7. Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. Curr Biol 2004; 14(17):R693-R696.
8. Baldauf SL. The deep roots of eukaryotes. Science 2003; 300(5626):1703-6.
9. Cavalier-Smith T, Chao EE. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. J Mol Evol 2003; 56(5):540-63.
10. Lang BF, O'Kelly C, Nerad T et al. The closest unicellular relatives of animals. Curr Biol 2002; 12(20):1773-8.
11. Bapteste E, Gribaldo S. The genome reduction hypothesis and the phylogeny of eukaryotes. Trends Genet 2003; 19(12):696-700.
12. Stechmann A, Cavalier-Smith T. Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. J Mol Evol 2003; 57(4):408-19.
13. Steenkamp ET, Baldauf SL. Origin and evolution of animals, fungi and their unicellular allies (Opisthokonta). In: Hirt RP, Horner DS, eds. Organelles, Genomes and Eukaryote Phylogeny. Boca Raton: CRC Press, 2004:109-29.
14. Nikolaev SI, Berney C, Fahrni JF et al. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci USA 2004; 101(21):8066-71.
15. Cavalier-Smith T, Chao EEY. Molecular phylogeny of centrohelid heliozoa, a novel lineage of bikont eukaryotes that arose by ciliary loss. J Mol Evol 2003; 56(4):387-96.
16. Pazour GJ, Agrin N, Leszyk J et al. Proteomic analysis of a eukaryotic cilium. J Cell Biol 2005; 170:103-13.
17. Berriman M, Ghedin E, Hertz-Fowler C et al. The genome of the African trypanosome Trypanosoma brucei. Science 2005; 309(5733):416-22.
18. Avidor-Reiss T, Maer AM, Koundakjian E et al. Decoding cilia function: Defining specialized genes required for compartmentalized cilia biogenesis. Cell 2004; 117(4):527-39.
19. Li JB, Gerdes JM, Haycraft CJ et al. Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell 2004; 117(4):541-52.
20. Cavalier-Smith T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol 2002; 52(Pt 2):297-354.
21. Moestrup O. The flagellate cytoskeleton. In: Leadbeater BCS, Green J, eds. The Flagellates. London: Taylor and Francis, 2000:69-94.
22. McKean PG, Vaughan S, Gull K. The extended tubulin superfamily. J Cell Sci 2001; 114(15):2723-33.
23. Amos LA, van den EF, Lowe J. Structural/functional homology between the bacterial and eukaryotic cytoskeletons. Curr Opin Cell Biol 2004; 16(1):24-31.
24. Gitai Z. The new bacterial cell biology: Moving parts and subcellular architecture. Cell 2005; 120(5):577-86.
25. Dutcher SK. The tubulin fraternity: Alpha to eta. Curr Opin Cell Biol 2001; 13:49-54.
26. Dutcher SK. Elucidation of basal body and centriole functions in Chlamydomonas reinhardtii. Traffic 2003; 4(7):443-51.

27. Ruiz F, Krzywicka A, Klotz C et al. The SM19 gene, required for duplication of basal bodies in Paramecium, encodes a novel tubulin, eta-tubulin. Curr Biol 2000; 10(22):1451-4.
28. Dupuis-Williams P, Fleury-Aubusson A, de Loubresse NG et al. Functional role of epsilon-tubulin in the assembly of the centriolar microtubule scaffold. J Cell Biol 2002; 158(7):1183-93.
29. Dutcher SK. Long-lost relatives reappear: Identification of new members of the tubulin superfamily. Curr Opin Microbiol 2003; 6(6):634-40.
30. Iyer LM, Leipe DD, Koonin EV et al. Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol 2004; 146(1-2):11-31.
31. Asai DJ, Koonce MF. The dynein heavy chain: Structure, mechanics and evolution. TICB 2001; 11(5):196-202.
32. Iyer LM, Makarova KS, Koonin EV et al. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: Implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res 2004; 32(17):5260-79.
33. Gibbons BH, Asai DJ, Tang W-JY et al. Phylogeny and expression of axonemal and cytoplasmic dynein genes in sea urchins. Mol Biol Cell 1994; 5:57-70.
34. Asai DJ, Wilkes DE. The dynein heavy chain family. J Euk Microbiol 2004; 51(1):23-9.
35. Porter ME. Axonemal dyneins: Assembly, organization, and regulation. Curr Opin Cell Biol 1996; 8:10-7.
36. Porter ME, Sale WS. The 9+2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility. J Cell Biol 2000; 151(5):F37-F42.
37. Mitchell DR. Chlamydomonas flagella. J Phycol 2000; 36:261-73.
38. Miki H, Setou M, Hirokawa N. Kinesin superfamily proteins (KIFs) in the mouse transcriptome. Genome Res 2003; 13(6B):1455-65.
39. Bernstein M, Beech PL, Katz SG et al. A new kinesin-like protein (Klp1) localized to a single microtubule of the Chlamydomonas flagellum. J Cell Biol 1994; 125:1313-26.
40. Yokoyama R, O'Toole E, Ghosh S et al. Regulation of flagellar dynein by a central pair kinesin. Proc Natl Acad Sci USA 2004; 101:17398-403.
41. Kubai DF. The evolution of the mitotic spindle. Int Rev Cytol 1975; 43:167-227.
42. Cavalier-Smith T. The evolutionary origin and phylogeny of eukaryote flagella. Symp Soc Exp Biol 1982; 35:465-93.
43. Cavalier-Smith T. The evolutionary origin and phylogeny of microtubules, mitotic spindles and eukaryote flagella. Biosystems 1978; 10(1-2):93-114.
44. Mitchell DR. Speculations on the evolution of 9+2 organelles and the role of central pair microtubules. Biol Cell 2004; 96(9):691-6.
45. Arndt H, Dietrich D, Auer B et al. Functional diversity of heterotrophic flagellates in aquatic ecosystems. In: Leadbeater BCS, Green J, eds. The Flagellates. London: Taylor and Francis, 2000:240-68.
46. Sleigh MA. Trophic strategies. In: Leadbeater BCS, Green J, eds. The Flagellates. London: Taylor and Francis, 2000:147-65.
47. Mitchell DR. Regulation of eukaryotic flagellar motility. Am Inst Phys Conf Proc 2005; 555:130-6.
48. Gibbons IR. The relationship between the fine structure and direction of beat in gill cilia of a lamellibranch mollusc. J Biophys Bioch Cyt 1961; 11:179-205.
49. Sale WS. The axonemal axis and calcium-induced asymmetry of active microtubule sliding in sea urchin sperm tails. J Cell Biol 1986; 102:2042-52.
50. Tamm SL, Tamm S. Ciliary reversal without rotation of axonemal structures in ctenophore comb plates. J Cell Biol 1981; 89:495-509.
51. Melkonian M, Robenek H, and Rassat J. Flagellar membrane specializations and their relationship to mastigonemes and microtubules in Euglena gracilis. J Cell Sci 1982; 55:115-35.
52. Omoto CK, Gibbons IR, Kamiya R et al. Rotation of the central pair microtubules in eukaryotic flagella. Mol Biol Cell 1999; 10(1):1-4.
53. Mitchell DR, Nakatsugawa M. Bend propagation drives central pair rotation in Chlamydomonas reinhardtii flagella. J Cell Biol 2004; 166(5):709-15.
54. Omoto CK, Witman GB. Functionally significant central-pair rotation in a primitive eukaryotic flagellum. Nature 1981; 290:708-10.
55. Omoto CK, Kung C. Rotation and twist of the central-pair microtubules in the cilia of Paramecium. J Cell Biol 1980; 87(1):33-46.
56. Jarosch R, Fuchs B. On the rotation of fibrils in the synura-flagellum (author's transl). Protoplasma 1975; 85(2-4):285-90.
57. Baccetti B. Evolutionary trends in sperm structure. Comp Biochem Physiol A 1986; 85(1):29-36.
58. Prensier G, Vivier E, Goldstein S et al. Motile flagellum with a "3 + 0" ultrastructure. Science 1980; 207:1493-4.

59. Nonaka S, Tanaka Y, Okada Y et al. Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. Cell 1998; 95(6):829-37.
60. Gibbons BH, Gibbons IR, Baccetti B. Structure and motility of the 9 + 0 flagellum of eel spermatozoa. J Submicrosc Cytol 1983; 15:15-20.
61. Mukai C, Okuno M. Glycolysis plays a major role for adenosine triphosphate supplementation in mouse sperm flagellar movement. Biol Reprod 2004; 71(2):540-7.
62. Lardy HA, Phillips PH. The interrelation of oxidative and glycolytic processes as sources of energy for bull spermatozoa. American Journal of Physiology 1941; 133:602-9.
63. Hsu SC, Molday RS. Glycolytic enzymes and a GLUT-1 glucose transporter in the outer segments of rod and cone photoreceptor cells. J Biol Chem 1991; 266(32):21745-52.
64. Westhoff D, Kamp G. Glyceraldehyde 3-phosphate dehydrogenase is bound to the fibrous sheath of mammalian spermatozoa. J Cell Sci 1997; 110(Pt 15):1821-9.
65. Gitlits VM, Toh BH, Loveland KL et al. The glycolytic enzyme enolase is present in sperm tail and displays nucleotide-dependent association with microtubules. Eur J Cell Biol 2000; 79:104-11.
66. Mitchell BF, Pedersen LB, Feely M et al. ATP production in Chlamydomonas reinhardtii flagella by glycotytic enzymes. Mol Biol Cell 2005; 16:4509-18.
67. Tombes RM, Brokaw CJ, Shapiro BM. Creatine kinase-dependent energy transport in sea urchin spermatozoa. Flagellar wave attenuation and theoretical analysis of high energy phosphate diffusion. Biophys J 1987; 52(1):75-86.
68. Huszar G, Sbracia M, Vigue L et al. Sperm plasma membrane remodeling during spermiogenetic maturation in men: Relationship among plasma membrane beta 1,4-galactosyltransferase, cytoplasmic creatine phosphokinase, and creatine phosphokinase isoform ratios. Biol Reprod 1997; 56(4):1020-4.
69. Wallimann T, Wegmann G, Moser H et al. High content of creatine kinase in chicken retina: Compartmentalized localization of creatine kinase isoenzymes in photoreceptor cells. Proc Natl Acad Sci USA 1986; 83(11):3816-9.
70. Noguchi M, Sawadas T, Akazawa T. ATP-regenerating system in the cilia of Paramecium caudatum. J Exp Biol 2001; 204(6):1063-71.

Index

A

α-solenoid 43, 75, 76
Acidisome 88
Acritarch 3, 4, 6, 8-10, 14
Actin 11, 47, 51, 62, 65, 68, 70, 86, 92, 97, 98-110, 119, 125-127
Actin-binding proteins 97, 104, 105, 109
Actin cytoskeleton 51, 92, 97, 100, 105-107, 109
Actin depolymerization factor (ADF) 98-103, 107, 109
Actin nucleation 100, 103, 106, 108-110
Actobindin 105
Adaptin 91-93
Allomyces 132
Alpha proteobacteria 21
Alveolates 3, 28, 123
Aminopeptidase I 113
Amitochondriate eukaryotes 1, 26, 30
Amoebozoa 10, 28-30, 91, 98, 99, 102, 103, 105-107, 114, 121, 122, 125, 132
Amphipathic helix 47
Angiosperms 24, 110, 122, 123, 126
Animals 1, 8, 10, 14, 15, 21, 24, 27, 31, 120-122, 124, 125, 127, 132
AP180 92, 94
Apicomplexa 28, 29, 98, 99, 103
Apoptosis 86, 106, 109
Aquifex 23
Archaea 1, 20, 21, 25, 27, 29, 50, 55, 57, 97, 101, 108, 121, 133
Archean 1, 3, 4, 11
Archeoglobus 22
Archezoa 20, 21, 26, 27, 31
Arp2/3 complex 103, 106, 107, 109, 110
Atg12 113-116
Atg8 113-116
Autogenous 38, 39, 42, 44-47, 50
Autophagosome 38, 111, 113, 116
Autophagy 111-117
Axoneme 119-122, 124-127, 130, 134, 136

B

β-propeller 43, 75, 76
β-thymosin 100
β-trefoil fold 105
Bacteria 1, 3, 11, 13, 21, 25, 29, 41, 50-52, 55, 85, 97, 99, 101, 108, 109, 112, 117, 120, 121, 133, 136
Bacterial septation 133
Bangiomorpha pubescens 7-9
Bangiophyte red algae 1, 4, 13
Barbed end 103, 104, 109
Basal body 119-122, 124-127, 130-132, 134, 136
Bayesian methods 23
Bikonts 2, 29-31, 84, 101, 102, 104-107, 116, 117, 121, 122, 132, 133
Biomarkers 1-4, 11, 15, 30
Biopolymers 2, 3, 8, 11
Brown algae 1, 8, 10, 28

C

Caenorhabditis elegans 74, 77, 84, 88, 123, 125, 127, 135
Calmodulin 121, 123
Calponin homology domain 99, 104, 110
Cambrian 10
Capping protein 104, 107, 109, 110
Carpediemonas 29, 88
CASP 63
Caulobacter 97, 108
Caveolae 86
Caveolin 86, 91
Cdc31p 122, 123, 125
Centrin 119-123, 125-127
Centrosome 119-122, 125-127
Cercozoans 86
Chaperone-mediated autophagy 111, 113
Chlamydomonas reinhardtii 84, 91, 88, 120, 122, 123, 126, 132
Chlorogonium elongatum 121
Chloroplast 2, 11, 14, 133
Choanoflagellates 27, 121, 132, 136
Cholesterol 86
Chromalveolates 10, 28-30, 84, 86, 87, 89, 91, 92, 94, 104, 114-116, 121, 132, 134, 136
Chromatin 45, 47, 76
Chytridiomycete 122, 132
Ciboulot 105
Cilia 29, 44, 47, 49, 50, 87, 130, 131, 135-137

Ciliate 3, 27-29, 86, 121, 123, 132, 134
Ciona intestinalis 84, 88, 91
Clathrin 43, 47, 51, 62, 63, 66-68, 74, 75, 76, 86, 91-93
Coat protein 62, 73-75, 91
Coat protein complexes 73-75
Coatomer 63, 66, 67
Coenocytic hyphae 126
Complexification 15, 20, 21, 31
COPI 43, 62-69, 74-76
COPII 43, 51, 62-64, 66-69, 74-76, 101, 109
Corallochytreans 27
Coronin 102, 106
Cortactin 106
Cotranslational transport 52, 53, 55, 57
CrCenp 122, 125-127
Crescentin 11, 97
Cryptococcus neoformans 84, 88
Cryptophytes 28
Cyanidioschyzon 84, 88, 94
Cyanidioschyzon merolae 84, 88, 89, 94, 91
Cyanobacteria 3, 7, 8, 11, 13, 14
Cytopharynx 86, 87
Cytoproct 87, 88
Cytoskeleton 1, 4, 7, 11, 13, 14, 31, 38, 41, 47, 49, 51, 57, 77, 89, 92, 97, 100, 104, 105-110, 119, 120, 132, 134, 135

D

Dab2 92
Deinococcus 23
Diatoms 8, 14, 28, 121
Dictyostelium discoideum 99, 108, 114, 102, 123
Dinoflagellates 3, 28, 29, 132
Diplomonad 27, 29, 104, 123, 132
Diplomonadida 98, 99
Domains of life 21, 27
Drosophila melanogaster 74, 84, 88, 114, 115, 123
Dynamin 91
Dynein 65, 101, 109, 119, 120, 126, 130, 131, 133-138

E

Ecdysozoa 24, 122, 123, 125-127
Encephalitozoon cuniculi 54, 116
Endocytosis 45-47, 66, 68, 74-76, 84-86, 88, 91-94, 106, 111, 116, 117, 125, 135
Endomembrane 11, 38-41, 43-47, 49, 58, 73-77, 89, 93, 116, 132, 133, 135
Endophilin 47, 51
Endoplasmic reticulum (ER) 38, 40, 43, 45, 47, 52, 55-58, 61, 62, 64-70, 73-76, 93, 113
Entamoeba histolytica 77, 84, 88, 99
Eps15 92, 93
Epsin 47, 51, 92, 93
Escherichia coli 52, 55, 99
ESCRT 92, 93
Eubacteria 40, 44, 75
Euglenozoa 29, 88
Excavata 29, 84, 91, 92, 94, 98, 114, 116, 132
Excavates 10, 29, 86, 88, 136
Excystment 2, 4, 5, 7, 11, 13

F

Fatty acid ester 44
Filament nucleation 97, 109
Filament polymerization 100
Filament severing 101
Filopodia 100
Fimbrin 99, 104, 110
Flagella 1, 2, 31, 108, 121, 127, 130-137
Flagellar apparatus 119-121
Flagellar pocket 86, 88, 125
Formin 99, 103, 109, 110
FtsA 101
FtsK 133
FtsZ 11, 47, 51, 97, 133
Fucosyltransferase 63
Fungi 1, 4, 8-10, 13, 14, 21, 27, 50, 54, 65, 86, 88, 89, 91-94, 98, 100, 102-107, 110, 114, 116, 119, 122-126, 132, 134

G

Galactosyl transferase 63
GARP 91, 93
Gastrulation 121
Gelsolin 99, 101, 102, 105, 107, 109
GGAs 92
Giantin 63
Giardia intestinalis 122
Giardia lamblia 54, 84, 88, 89, 91, 99, 115, 116, 123
Gliding motility 101, 130, 135, 137
Glycosylation 58, 61, 64, 65
Golgi complex 61-64, 66-70, 86, 88, 89, 91, 120
GOS28 64

Grypania 10, 13
GTPase 46, 47, 50, 52, 53, 55-57, 66-69, 73-81, 86, 88, 91, 92, 99, 101, 103, 106-108, 110
GTPase activating protein (GAP) 56, 76, 77

H

Heterodimeric capping protein 104
Heterolobosean amoeboflagellate 88
Homo sapiens 74, 77, 84, 88, 102, 123
Homoplasy 22
HOPS 91, 116
Hrs 93
Hsp70 101
Hydrogenosomes 1, 27

I

Ichthyosporeans 27
Intraflagellar transport 44, 50, 130, 131, 137
Isoprenyl ether 44

J

Jakobids 29, 88

K

Karyo-mastigont 120
Kinesin 67, 106, 110, 133-135
Kinetoplastida 88, 98, 99, 103, 104

L

Lamellipodia 100
Last common ancestor of extant eukaryotes (LCAEE) 27, 30, 31
Last common eukaryotic ancestor (LCEA) 93, 114, 116, 117, 130, 132, 134, 136, 138
Last universal common ancestor (LUCA) 21
Leishmania donovani 114
Leishmania major 54, 99, 115, 116
Liposomes 47, 51, 58
Long branch attraction 20, 23, 27, 123
Lysosome 38, 51, 66, 73-75, 86, 88, 89, 91, 93, 111, 113, 114, 116, 117

M

Macroautophagy 113, 114, 116
Malawimonas 29, 88
Mannosidase 63
Mastigamoebae 136
Mastigonemes 136
Maximum likelihood (ML) 23, 24, 50
Maximum parsimony 21, 27
Membrane ruffles 100
Membrome 74, 78
Metazoa 84, 86, 89, 91-93, 97, 98, 100, 102-107, 110, 113, 114, 119, 122, 123, 126, 127
MglB 99, 101, 107
Microautophagy 113, 114, 116
Microfossils 2-5, 7-11, 13, 14
Microsporidia 27, 31, 62, 64-68
Microtubule 47, 49, 51, 65, 106, 110, 119, 120, 126, 127, 130, 132-137
Microtubule organizing center (MTOC) 134, 137
Mitochondria 1, 2, 11, 20, 21, 27, 29-31, 39, 41, 44, 47, 49, 50, 65, 69, 73, 97, 107, 108, 113, 116, 133
Mitosomes 1, 27
MoxR 133
MreB 11, 47, 99, 101, 107
Multicellularity 1, 8, 13, 14, 70, 119-121, 127
Multivesicular body 93
Mus musculus 114, 123
Myosin 2, 29, 50, 65, 70, 79, 80, 84, 99, 100, 105-107, 110, 121, 126, 127

N

Nodal cilia 137
Nuclear pore complex (NPC) 38, 40-45, 50, 74, 76, 93, 122
Nucleariids 27
Nucleation 97
Nucleus 1, 2, 11, 30, 31, 39-41, 43-45, 47, 49, 50, 66, 68, 69, 74, 111, 112, 119-122, 124-127, 132-135
Nyctotherus 27

O

Oomycetes 28, 29
Opisthokonta 27, 29, 88, 92-94, 105
Opisthokonts 8, 10, 29-31, 84, 86, 92, 94, 121-126, 132
Oxygen 1, 4, 11, 21, 30, 50

P

Palaeovaucheria 8
Paramecium 85-88, 99, 123, 133, 137
Paranosema grylli 64
ParM 11, 47
Paulinella chromatophora 27
Peroxisome 21, 50
Pexophagy 113
Phagocytosis 11, 14, 41, 45-47, 58, 67, 68, 74, 84-87, 107, 135
Phagotrophy 41, 49, 64
Phosphatidylethanolamine 113
Phosphoinositides 92
Phylogenomic 20, 21, 24, 25, 27, 28
Phylogeny 2, 20-22, 24-29, 31, 50, 57, 111, 123, 131, 132
Physarum 132
Phytophthora ramorum 84, 88
Pichia methanolica 113
Pichia pastoris 70, 113
PI-kinases 92
Pinocytosis 84, 86
Plantae 27, 29, 84, 86, 89, 92, 94, 98, 121, 132, 134
Plants 10, 21, 27, 31, 62, 65, 70, 91, 97, 100, 101, 103, 104, 106-109, 111, 114, 116, 119-126, 132
Plasmodium falciparum 84, 88, 102, 115, 123
PLK4 127
Polyubiquitin 112, 117
Precambrian 1, 2, 15
Predation 1, 8, 11, 13-15, 45, 49, 121
Procollagen-I 62, 63
Profilin 98, 100, 101, 103, 105, 107-109
Proteasome 112, 113, 117
Protein conducting channel 40, 44, 52, 55, 57
Proterocladus 8
Proterozoic 1-3, 7, 10, 11, 13-15
Protocoatomer 43, 44
Pseudopodia 86, 121, 125

R

Rab GTPase 73, 74, 77-81
Radial spokes 130, 131, 136-138
Retortamonads 29, 88
Rhizaria 10, 29, 84, 121, 132
Rho GTPases 103, 106, 107, 110

S

Saccharomyces cerevisiae 50, 54, 74, 77, 84, 86, 88, 93, 102, 109, 113, 114, 115, 120, 122, 123, 125, 133
Sar1 47, 58, 62, 66, 67, 69, 74, 76, 109
Satka favosa 6, 7
Schizosaccharomyces pombe 62, 77, 114, 115, 123, 125
Sec23 66, 76, 99, 101, 102, 109
Secondary simplification 20, 31, 89
Secretion 44, 46, 58, 65, 67, 68, 70
SecY 42, 44, 45, 55
Septate hyphae 126
SFI1 125, 126
Shigella 117
Shuiyousphaeridium macroreticulatum 6, 7
Sialyltransferase 63
Signal recognition particle 52, 53
Small subunit ribosomal RNA (SSU rRNA) 20, 21
SNARE 62-65, 68, 69, 73, 74, 77, 78, 81, 88, 89, 91, 116
Spermatozoa 137
Spindle pole body 119
Spire 21, 110, 105
Spirochete 44
SRα 25, 52, 55-57
SRβ 47, 52, 55-58, 67
SRP 52-58
SRP receptor 52, 55
SRP54 25, 52, 53, 55-57
Sterols 3, 11, 69
Stonins 92
Stramenopiles 8, 28
Streptococcus 117
Symbiosis 8, 39, 40, 44, 50, 108
Symplasm 126
Syntaxin 65, 68, 88, 89, 91, 93
Syntrophy 40, 49

T

Tappania plana 4, 5, 7, 11, 13
Tetrahymena thermophila 84, 88, 91, 104, 114, 115
Thalassiosira pseudonana 84, 88
Thermotoga 23
Thermus thermophilus 23, 24
Toxofilin 100
Toxoplasma 99, 100
Treadmilling 100

Tree of Life 20, 21, 25, 27, 88
Trichocysts 88
Trichomonads 26, 27, 29
Trichomonas vaginalis 77, 99, 115, 116, 123
Trimastix 29, 88
Trypanosoma 54, 84, 86, 88, 94, 99, 121, 123, 125, 132
Tubulin 11, 47, 51, 97, 101, 119, 126, 133, 134

U

Ubiquitin 92, 93, 111, 113, 114
Unikont 1, 2, 29, 84, 101-103, 105-107, 116, 117, 121, 122, 125, 132, 133

V

Valeria lophostriata 4, 5, 7, 11
Vesicle transport 58, 88, 93, 119
Villin headpiece (VHP) 98, 99, 102, 104, 105, 107, 108
VSVG 62, 63

W

WASP homology domain 2 (WH2) 98-100, 104-108, 110
WAVE 106, 108