Evolution from the Galapagos

Book · January 2013

DOI: 10.1007/978-1-4614-6732-8

4 authors, including:

Gabriel Trueba
Universidad San Francisco de Quito (USFQ)
284 PUBLICATIONS **1,717** CITATIONS
SEE PROFILE

Stephen J. Walsh
University of North Carolina at Chapel Hill
194 PUBLICATIONS **6,055** CITATIONS
SEE PROFILE

Carlos F. Mena
Universidad San Francisco de Quito (USFQ)
63 PUBLICATIONS **1,134** CITATIONS
SEE PROFILE

Social and Ecological Interactions in the Galapagos Islands

Gabriel Trueba
Carlos Montúfar *Editors*

Evolution from the Galapagos

Two Centuries after Darwin

Springer

Social and Ecological Interactions in the Galapagos Islands

For further volumes:
http://www.springer.com/series/10427

Gabriel Trueba • Carlos Montúfar
Editors

Evolution from the Galapagos

Two Centuries after Darwin

Springer

Editors
Gabriel Trueba
Universidad San Francisco de Quito
Quito
Ecuador

Carlos Montúfar
Universidad San Francisco de Quito
Quito
Ecuador

ISSN 2195-1055 ISSN 2195-1063 (electronic)
ISBN 978-1-4614-6731-1 ISBN 978-1-4614-6732-8 (eBook)
DOI 10.1007/978-1-4614-6732-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013939741

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

In Memory of Lynn Margulis (1938–2011)

Foreword

Social and Ecological Interactions in the Galapagos Islands—a New Galapagos Book Series edited by Stephen J. Walsh & Carlos F. Mena and published by Springer Science+Business Media.

As the second book in the Galapagos Book Series, Gabriel Trueba, Professor at the Universidad San Francisco de Quito, has organised an exciting collection of papers on evolution written by a number of highly qualified authors that collectively describe important topics, issues, and approaches to the study of evolution, which strongly resonate in the Galapagos Islands of Ecuador. As the "Cradle of Evolution," the Galapagos Islands have achieved international recognition and acclaim through the ground-breaking work of Charles Darwin and the work that has followed by scientists, such as Peter and Rosemary Grant in the Galapagos Islands and their students, including Carlos Valle, a Galapagueño from Santa Cruz Island and a Professor at the Universidad San Francisco de Quito. Several other scholars, showcased in this book, have made important contributions to evolutionary biology and other associated areas of research that are described in various chapters of this book.

Building on the success of the initial book that launched the new Galapagos Book Series, "***Science and Conservation in the Galapagos Islands: Frameworks and Perspectives*** (***2013***)," Stephen J. Walsh & Carlos F. Mena (Editors), this book addresses topics that push the evolutionary envelop and advance our understanding in critical areas. Achieved through a World Summit on Evolution that was held in the Galapagos Islands, the following chapters contribute to the study of evolution and enhance the Galapagos Book Series in a number of important ways.

We look forward to expanding the Galapagos Book Series with other outstanding contributions to the study of the Galapagos Islands and the Galapagos Marine Reserve. We will continue to select themes that are important to the study of the social, terrestrial, and marine subsystems and their linked effects in the Galapagos Islands as a template to the study of other similarly challenged island ecosystems around the globe.

University of North Carolina at Chapel Hill, USA — Stephen J. Walsh
Universidad San Francisco de Quito, Ecuador — Carlos F. Mena
Galapagos Book Series Editors—Springer Science+Business Media

Foreword

Evolution; Galapagos; two words which conjure up images, both vivid and inspiring, in the minds of biologists everywhere. Let us take evolution first; and of course Charles Darwin, whose name is synonymous with evolution, but let us not forget Alfred Russel Wallace (the centenary of his death will be celebrated in 2013) who independently came up with similar ideas to those of Darwin.

As is apparent in this book, remarkable developments have taken place in our understanding of evolution since the papers of Darwin and Wallace were read at the Linnean Society of London by the Secretary on July 1st 1858. I still find it surprising that Thomas Bell, President of the Linnean Society at that time, made no reference to this work in the Annual Report, thus implying that he failed to see the significance of those papers, but it is easy with the gift of hindsight to make such a comment over 150 years later!

It is widely known that Darwin's visit to the Galapagos in 1835 and the collections he made there, when he was on the voyage of the Beagle, provided him with material to develop his ideas on evolution, culminating in the publication of his book 'On the Origin of Species by Natural Selection or the Preservation of Favoured Races in the struggle of Life'. With the massive advances in molecular and computer techniques, the advent of bioinformatics, sequencing techniques and genomics over the last half century, some chapters in this book address problems not conceived by Darwin: the book brings readers bang up to date to the cutting edge of scientific knowledge.

New perspectives in the role of symbiosis are discussed, and the birth of a new science, symbiogenetics, to study all aspects of genetic integration of heterologous genomes, is seen as a new discipline within studies on evolution.

It is fascinating to consider that Darwinian selection could have played a role at the molecular level before life began. Darwin realised that little was understood about the origin of life, and in the last 160 years much has been learnt about the synthesis of organic compounds from the so-called 'primeval broth'. Despite new methodologies facilitating a much greater understanding of the intricacies and mechanisms of evolution, we still do not understand how the nucleic acid based genetic synthesis of extant life may have originated.

There are still many questions to be answered and much to be done in describing the organisms on earth, how they evolved and how they are related. It is still a guess

as to how many species on earth remain to be described, but the estimate that only about 5% of fungi are known illustrates the enormity of the task ahead. It is vital that this task is taken seriously, especially when one considers cases of extinction and how rapidly they are taking place. It is of crucial importance that we should know what is on Earth before it disappears forever.

But what is so special about the Galapagos archipelago? In general, island ecosystems and organisms provide opportunities for the study of evolution; the Galapagos, with its endemic fauna, flora and fungi provides an ideal ecosystem for such studies as is so clearly exemplified in this book. However, the existence of the Galapagos archipelago as an extraordinary ecosystem poses its own problems. Not surprisingly it has been for quite some time, and will continue to be, a uniquely magnetic pull for a considerable fraction of the human population, whether they are Darwinian disciples, tourists, natural historians, artists or photographers, and this pressure raises many conservation issues. The introduction of exotic species, which may colonise, act as competitors to endemic species and/or introduce pathogens such as viruses, bacteria or protozoa, are just some matters to cause us concern. Today, not one of the 19 large islands is free from introduced species. Nevertheless, the valuable work of the Galapagos National Park and Charles Darwin Foundation is rightly acknowledged, for both organisations have focused on protection and restoration of native endangered species and habitats and the control and eradication of exotic species.

I would like to heartily congratulate all those associated with this multi-authored book; those whose concept it was and all the contributors who have succeeded in producing a scholarly work to celebrate the bicentenary of the birth of Charles Darwin. I have no doubt that readers will find this book both illuminating and stimulating, taking us from the fundamental, mind-changing papers of Darwin and Wallace to the technologies of gene-sequencing, origins of organisms, how they evolved and the associated mechanisms, and their relationships to one another.

The authors are clearly demonstrating that ahead of us all lies the outline of a path for future research, which will further our understanding of life on earth and how it evolved.

Vaughan Southgate
President of the Linnean Society of London

Preface

Why Another Multi-Authored Evolution Book

This volume includes the collection of some of the most significant lectures that well-known experts presented at our two international "summits on evolution" (2005, 2009) as updated and revised chapters. The meetings took place on one of the large islands of the Galapagos archipelago (San Cristobal) at GAIAS (*Galapagos Institute for the Arts and Sciences*) of the Universidad San Francisco de Quito (*USFQ*), Ecuador. The chapters are ordered chronologically from the past to present Earth eons; they start from the origins of life and come towards the present concerns of mammalian evolution and recent ecological problems.

Modern life sciences comprise a vast subject broken up and taught as academic disciplines with a poor, sometimes non-existent record of integration. Different kinds of organisms at scales from the monstrous to the microbe come within its purview. Most evolution meetings reflect this fragmentation; they concentrate on certain modern forms of life and usually ignore both the fossil record of these and many other cognate, relevant issues. Ultimately all life shares common ancestry, as Darwin so perceptively described, and the diversity of the living evolved in similar ways in response to environmental, including climate, change. Yet organisms assignable to different domains (subphyla and phyla) exhibit evolutionary peculiarities. The role of lateral gene transfer, for example, is easily documented among prokaryotes whereas it seems far less conspicuous in animals. Yet the ultimate eukaryotic ancestors of the later evolved themselves by a massive horizontal gene transfer process recognisable as "genome acquisition". The study of some evolutionary processes like symbiosis have been neglected and even ignored by many interested in natural history. We believe that our integrative approach is necessary for a coherent view of the grand sweep of the evolution of life that is so entwined with the Earth's geological history.

The main goal of the two *Galapagos Summits on Evolution* has been to bring together scientists and graduate students engaged in the study of evolution, from life's origin to its current diversity. Due to their historical significance, the Galapagos are a unique venue for promoting comprehensive research on evolution and ecology and to make the research results available to students and teachers everywhere, but

especially from developing countries. As shown by the enthusiastic attendance at both summits and the many suggestions to keep them continuing, our meetings have opened new opportunities for students from Ecuador and other Latin American countries to be inspired by some of the most brilliant minds in evolutionary science. We hope to publish in English to make this work available to the widest possible readership and to distribute the documentary film.

Quito, Ecuador — Gabriel Trueba
Quito, Ecuador — Carlos Montúfar

Acknowledgements

The authors would like to thank Micah Dunthorn and Antonio Lazcano for contributing to organise the academic sessions for the two summits; to Tom Kunz and Micah Dunthorn for their contribution in the writing of the graduate student scholarship grant for the National Science Foundation; to Diego Quiroga, the director of the Galapagos Institute for the Arts and Sciences (GAIAS) for providing funds and for his commitment to the success of these events, the USFQ informatics department, to the USFQ catering services, to Paula Cordova, Veronica Barragan and Maria Teresa Perez for their invaluable help in the organisation of the summits; to Santiago Gangotena and Carlos Montufar for their unconditional support to these events; to Lynn Margulis, Antonio Lazcano, Jorge Chiriboga, Celeste A. Asikainen and Dianne Bilyak for their help in putting this book together.

Contents

Contributors

Sandra L. Baldauf Department of Systematic Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

Martin Carr Department of Biology, University of York, Heslington, York, UK

Michael F. Dolan Department of Geosciences, University of Massachusetts, Amherst, MA, USA

Edward Feil Department of Biology and Biochemistry, University of Bath, Bath, UK

Peter R. Grant Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

Rosemary Grant Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

Ricardo Guerrero Department of Microbiology, University of Barcelona, Barcelona, Spain

Amparo Latorre Genetics, Evolution of Genomic Complexity and Comparative Genomics, University of Valencia, Valencia, Spain

Antonio Lazcano School of Sciences, Universidad Nacional Autónoma de México (UNAM), Mexico

Lynn Margulis Department of Geosciences, University of Massachusetts, Amherst, MA, USA

Andrés Moya Department of Genetics, University of Valencia, Valencia, Spain

Maria Romeralo Department of Systematic Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

Bruce Scofield Psychology and Astronomy, Kepler College, Seattle, WA, USA

Department of Geosciences, University of Massachusetts Amherst, Amherst, MA, USA

Vaughan Southgate Department of Zoology, Natural History Museum, London, UK

Carlos A. Valle Biological and Environmental Sciences College, Universidad San Francisco de Quito, Quito, Ecuador

Shuhai Xiao Department of Geosciences, Virginia Tech, Blacksburg, VA, USA

Ada Yonath Structural Biology Department, Weizmann Institute of Science, Rehovot, Israel

Part I
Historical Perspectives

Chapter 1
Darwin–Wallace Paradigm Shift

Ten days that never shook the world

Ricardo Guerrero and Lynn Margulis[†]

Few visitors who wander between the hanging flags into the ample courtyard fail to be impressed by Burlington House, a splendid architectural complex. Most pass this place near Picadilly in central London on their mandatory visit to the aluminum sculpture *Eros* (*Love*). In fact, *Eros* is the "*The Angel of Christian Charity*"; he balances himself on one foot in the Circus a few hundred yards away from Burlington House. The Royal Academy of Art and other cultural societies such as Society of Antiquaries of London, the Royal Society of Chemistry, the Geological Society of London, where the world's first geological map is exhibited, the Royal Astronomical Society, and, since 1854, the Linnean Society are all accommodated just beyond the huge hanging flags.

During summer months, southeast England is humid and often warm, as it was at the end of June 1858, where Charles Robert Darwin (b. February 12, 1809, Shrewsbury, Shropshire, d. April 19, 1882, Down House, Kent) had his home and lived most of his adult life, in the county of Kent. There, Darwin enjoyed and documented the natural world at a propitious distance from the city. Surrounded by his loving and devoted wife Emma (née Wedgwood) and his prodigious descendants—she bore him ten children, seven of whom survived to adulthood—he led a calm life of study and research. Apparently, for reasons of health, he seldom left Down House. He possibly had contracted Chagas disease, a trypanosomiasis with recurring symptoms, in South America during his around-the-world voyage of the *Beagle* (1831–1836). Darwin, during his lifetime, already a well-known naturalist and a Fellow of the Linnean

[†] Lynn Margulis died on November 22, 2011.

R. Guerrero (✉)
Department of Microbiology, University of Barcelona, Barcelona, Spain
University of Massachusetts-Amherst, Amherst, USA
e-mail: rguerrero@iec.cat

L. Margulis
Department of Geosciences, University of Massachusetts-Amherst, Amherst, USA
Balliol College, Oxford University, Oxford, UK
e-mail: lmargulis@nsm.umass.edu

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_1, © Springer Science+Business Media New York 2013

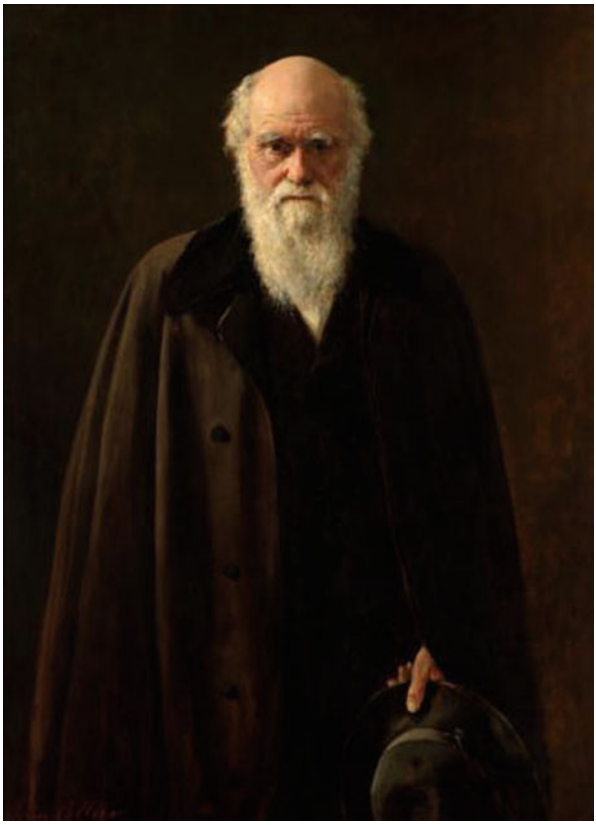


Fig. 1.1 Charles Robert Darwin's exhibited oil portrait that hangs in the great meeting room of the Linnean Society of London, Burlington House

Society, enjoyed a voluminous correspondence with intellectuals around the globe. One correspondent, an Englishman, a young wanderer Alfred Russel Wallace (1823–1913), wrote Darwin about adventures, observations, and collections in the East Indies. Enthusiastic Wallace, replete with scientific curiosity and a spirit to observe for himself, merited Darwin's distant letters (Fig. 1.1) (Burkhardt 1996).

During the last days of June 1858, a series of events occurred with consequences not only for the history of science but also for human history worldwide. On June 18, Darwin received a letter that Wallace had written in the South Seas in February. Wallace asked Darwin—given his influence in London's scientific circles—to publish his "essay" enclosed with the letter called *On the Tendency of Varieties to Depart Indefinitely from the Original Type*. Wallace put forward essentially the hypothesis on the origin of species that Darwin had sketched as early as the beginning of the 1840s. Ayala (2007), Mayr (1964) Darwin, recognizing the validity of his ideas on natural selection decided immediately to publish Wallace's essay. Beforehand, he consulted his two great mentors and friends: geologist Sir Charles Lyell and botanist Joseph Hooker. Both, well aware that Darwin had developed the concept for many years, argued against prior publication by Wallace. To deny Darwin's priority would seriously diminish the accomplishment of their admired friend. Darwin had written extensively on "descent with modification" and "natural selection." Indeed, by then,

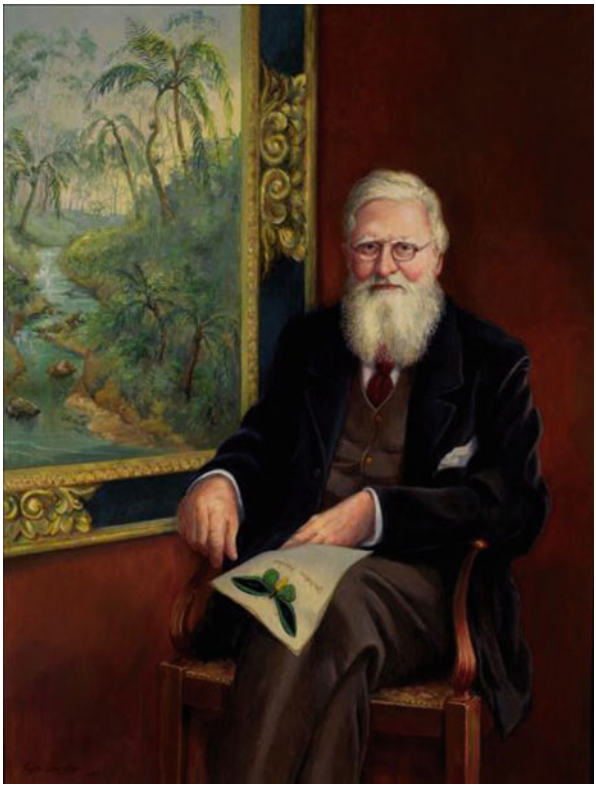


Fig. 1.2 Alfred Russel Wallace's exhibited oil portrait that hangs in the great meeting room of the Linnean Society of London, Burlington House

he had completed numerous chapters of his book. His belabored manuscript *On the Origin of Species*, contained myriad observations on the origin and perpetuation of variation in plants and animals, but it progressed only slowly. Lyell and Hooker in unison urged Darwin to present his and Wallace's papers simultaneously at the forthcoming meeting of the Linnean Society (Fig. 1.2).

A circumstance catapulted the readings into the Linean hall; only 10 days later, both papers were presented by neither author. The president of the Society from 1849 to 1853, the Scottish botanist Robert Brown (1773–1858) had died on June 10. Brown had discovered the nucleus of plant cells. He observed what was later called "Brownian movement," explained by Einstein many years later. The last meeting of the 1857–1858 season was scheduled for June 17, but, as a sign of grief and respect for Brown, only business was discussed and the scientific session was postponed. Since Brown, as Council member of the Linnean Society at the time of his death, required an elected replacement within 3 months, an adventitious meeting on Thursday, July 1, was hastily mandated (Gage and Stearn 1988). Lyell and Hooker quickly wrote to the secretary, John Joseph Bennet, and enclosed the Darwin–Wallace papers for presentation. They wrote emphasizing Darwin's prior elaboration of the "species-change-through-time" concept:

My Dear Sir,

The accompanying papers, which we have the honour of communicating to the Linnean Society, and which all relate to the same subject, viz. the Laws which affect the Production of Varieties, Races, and Species, contain the results of the investigations of two indefatigable naturalists, Mr. Charles Darwin and Mr. Alfred Wallace.

These gentlemen having, independently and unknown to one another, conceived the very same very ingenious theory to account for the appearance and perpetuation of varieties and of specific forms on our planet, may both fairly claim the merit of being original thinkers in this important line of inquiry; but neither of them having published his views, though Mr. Darwin has for many years past been repeatedly urged by us to do so, and both authors having now unreservedly placed their papers in our hands, we think it would best promote the interests of science that a selection from them should be laid before the Linnean Society.

Taken in the order of their dates, they consist of:

I. Extracts from a MS. work on Species, by Mr. Darwin, which was sketched in 1839, and copied in 1844, when the copy was read by Dr. Hooker, and its contents afterwards communicated to Sir Charles Lyell. The first Part is devoted to "The Variation of Organic Beings under Domestication and in their Natural State", and the second chapter of that Part, from which we propose to read to the Society the extracts referred to, is headed, "On the Variation of Organic Beings in a state of Nature; on the Natural Means of Selection; on the Comparison of Domestic Races and true Species".*

**This MS. work was never intended for publication, and therefore was not written with care.—C.D. 1858.*

II. An abstract of a private letter addressed to Professor Asa Gray, of Boston, U.S., in October 1857, by Mr. Darwin, in which he repeats his views, and which shows that these remained unaltered from 1839 to 1857.

III. An Essay by Mr. Wallace, entitled "On the Tendency of Varieties to depart indefinitely from the Original Type". This was written at Ternate [coastal town on a volcanic island of the Moluccas, Eastern Indonesia] in February 1858, for the perusal of his friend and correspondent Mr. Darwin, and sent to him with the expressed wish that it should be forwarded to Sir Charles Lyell, if Mr. Darwin thought it sufficiently novel and interesting. So highly did Mr. Darwin appreciate the value of the views therein set forth, that he proposed in a letter to Sir Charles Lyell, to obtain Mr. Wallace's consent to allow the Essay to be published as soon as possible. Of this step we highly approved, provided Mr. Darwin did not withhold from the public, as he was strongly inclined to do (in favour of Mr. Wallace), the memoir which he had himself written on the same subject, and which as before stated, one of us had perused in 1844, and the contents of which we had both of us been privy to for many years. On representing this to Mr. Darwin, he gave us permission to make what use thought proper of his memoir, &c.: and in adopting our present course, of presenting to the Linnean Society, we have explained to him that we are

not solely considering the relative claims to priority of himself and his friend, but the interests of science generally; for we feel it to be desirable that views founded on a wide deduction from the facts, and matured by years of reflection, should constitute at once a goal from which others may start, and that, while the scientific world is waiting for the appearance of Mr. Darwin's complete work, some of the leading results of his labours, as well as those of his able correspondent, should together be laid before the public.

We have the honour to be yours very obediently,

Charles Lyell. Jos. D. Hooker

This letter sent on June 30, a day before the meeting, of course arrived too late for any member, even if he so desired, to read either article. Neither Darwin, ill and grief-stricken by the death of one of his sons, nor Wallace, half the Earth away in the Eastern Moluccas, attended. By the order of delivery, the Lyell–Hooker letter indicates that they both clearly favored Darwin—his name is quoted ten times and Wallace's only four. The outcome was that three papers were read in London on the warm afternoon of July 1: (1) an extract of a chapter from the book that Darwin was preparing (which Lyell and Hooker insisted corresponded to a sketch from 1839, which Darwin gave to Hooker in 1844), (2) an abstract of a private letter addressed to Prof. Asa Gray, of Boston, in October 1857, and (3) Wallace's essay written, they insisted, in 1858.

After probably more than 2 hours of session, the six botany papers postponed from June 17 were presented to very few Linnean Society members on July 1. Linné's affirmation "*Species totae sunt sicut Deus creavit*" remained unscathed. Stability of the plant creation was demonstrated in experiments from Kew Gardens. The last paper on British flora by newly nominated vice-president George Bentham argued yet again for Linné's fixity of species. But, after the reading in absentia of the Darwin–Wallace papers, Bentham, probably embarrassed, canceled his own presentation.

None of the botanists or zoologists perceived the genesis of a new biology. The theory of "species evolution by natural selection" flew right by them. Some were mortally bored by the readings, a few, sensibly perturbed. President Thomas Bell did not call for commentary and the interminable session terminated late "without anything special to mention," as Bell recorded 1 year later in the minutes for July 1.

This failure to detect a scientific revolution does not negate the importance of "learned" scientific academies. Before the first formal scientific society (l'Accademia dei Lincei, i.e., Lincean Academy, founded in 1603 by 18-year-old Federico Cesi near Rome), men of science were isolated. Ideas were exchanged through "snail mail": courier, post-by-horse, or shipping ensured delays and losses of formal scientific letters. Scientific information spread by book publication with costs met by the author-researcher, who necessarily was wealthy and well-connected. Taxidermic specimens, fossils, hand-samples of rocks, drawings, books, maps, articles, letters all needed for publication preparation were unobtainable or accessed with difficulty. Scientific societies emerged by necessity. Bibliographic information and collections became available at the service of common interest groups. When the oldest natural

history society, the Linnean Society of London, was founded in 1788, it began with 3 Honorary Members, 20 Members, 39 Foreign Members, and 11 Associates. More than 2,000 members of different nationalities, professions, and specialties currently pay annual dues. For some of its first 40 years, the Society was the main channel for the communication and publication of research in the natural sciences.

Darwin, nominated by President Thomas Bell and members Sylvester Hanley, Edward Forbes, Robert Brown, John S. Henslow, James J. Bennett, and A. White, had been elected Fellow of the Linnean Society on March 7, 1854. Praiseworthy words were spoken on what Darwin, as new member, could be expected to contribute to the Society. Reality far exceeded those expectations. By then a well-known natural historian, A. R. Wallace joined the Linnean on January 18, 1871. He was nominated by George Bentham, H. T. Stainton, J. D. Hooker, A. W. Bennett, S. Stevens, A. Muller, E. W. H. Holdsworth, A. Newton, W. H. Flower, J. W. Dunning, G. R. Gray, and E. Sheppard (Gage and Stearn 1988).

Thomas Bell, successor of Brown in 1853, as sixth president until 1861, had little idea that the July 1, 1858, session would shift a paradigm. Bell did not live to see a scientific revolution, a theory that, in Christendom and beyond, changed entire belief systems. Nor did he ever see the magnificent life-size portraits of both bearded naturalists—C. R. Darwin and A. R. Wallace—in the meeting room today. The Darwin–Wallace readings on that stuffy, humid July 1 session were not mentioned till almost a year later, on May 24, 1859, in Bell's Annual Report to Fellows of the Society (Guerrero 2008):

> The year [1858] which has passed [. . .] has not, indeed, been marked by any of those striking discoveries which at once revolutionize, so to speak, the department of science on which they bear; it is only at remote intervals that we can reasonably expect any sudden and brilliant innovation which shall produce a marked and permanent impress on the character of any branch of knowledge, or confer a lasting and important service on mankind. A Bacon or a Newton, an Oersted or a Wheatstone, a Davy or a Daguerre, is an occasional phenomenon, whose existence and career seem to be specially appointed by Providence, for the purpose of effecting some great important change in the conditions or pursuits of man.

He could not have been more mistaken. The Darwin–Wallace July 1, 1858, presentation impelled Darwin to complete his book "*On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*," which was finally published on November 24, 1859, by John Murray, London. Darwin's revolutionary book was a sudden success; the whole edition sold out, primarily to book sellers, on the day of its release (Tickell 2008). People from many scientific disciplines (algology to zoology) and walks of life (anarchists to zealots) were "impregnated" by "Darwin's dangerous idea (Dennett 1995; Ryan 2002)." But, by the time the books were in the hands of eager readers and "natural selection" was debated in salons, pubs, academic halls, newspapers, scientific societies and from soapboxes, no one realized that the commotion had begun 17 months before, during 10 frantic days preceding July 1. Except for Darwin, Lyell, Hooker their families and closest friends, those 10 days, to say the least, had not yet "shaken the world."

References

Ayala FJ (2007) Darwin's greatest discovery: design without designer. Proc Natl Acad Sci U S A 104:8567–8573

Burkhardt F (1996) Charles Darwin's letters. A selection 1825–1859. Cambridge University Press, UK

Dennett DC (1995) Darwin's dangerous idea. Evolution and the meaning of life. Penguin Group, Simon & Schuster, USA

Gage AT, Stearn WT (1988) A bicentenary history of the Linnean society of London. Academic Press, UK

Guerrero R (2008) The session that did not shake the world (the Linnean Society, 1st July 1858). Int Microbiol 11:209–212

Mayr E (1964) Introduction to "On the Origin of Species", a facsimile of the first edition. Harvard University Press, USA

Ryan F (2002) Darwin's blind spot. Evolution beyond natural selection. Houghton Mifflin, USA

Tickell C (2008) The theory of evolution: 150 years afterwards. Int Microbiol 11:283–288

Chapter 2
From Copernicus to Darwin (1473–1882)

Carlos Montúfar

A small but fascinating country, Ecuador is probably one of the most interesting natural laboratories on the planet. Compact in its geography but diverse in all aspects, its territory hosts one of the most biodiverse settings available on the planet. Its mainland is covered with lush tropical rainforests, snow-peaked active volcanoes that reach the sky, and unique coastal region rich in marine life. The Galapagos Islands, some 600 miles west, speak for themselves. Throughout its history, its natural beauty as well as its rich human heritage has enchanted travelers, scientists, and adventurers. Endowed with diversity from indigenous tribes in the rain forest to modern universities in urban settings, it still is a natural encyclopedia waiting to be read and discovered.

Stretching our imagination, it is not hard to see how this unique place on the planet has provided scientific information that has contributed to rethink about our place in the universe. In a span of two centuries, adventurous scientific expeditions confirmed and kindled two of the major paradigm shifts in human thought: heliocentrism and evolution. Science cannot rely on common sense. Staring at the night sky from anywhere on the planet and especially from the high mountains of Ecuador, it is not hard to mistakenly conclude that we must be in the center of the universe. Not only are we in its center but also we should be unique, permanent, and immutable in this vast unknown, created by the works of a supreme being. These ideas were so embedded in human thought that it took millennia to challenge them. Philosophers and scientists developed intriguing and ingenious mechanisms and theories to explain current thought on the origin of the universe, the origin of life, and our place in the universe. One must admit the creative genius of humans in elaborating complicated models explaining "the majestic clockwork" of the universe as Bronowsky so elegantly describes in his inspiring book, *The Ascent of Man* (Bronowski 1973).

Human inventiveness has also allowed exceptional discoveries that have changed our view of the universe and ourselves based on keen observation with rudimentary or no instruments. Such is the case of a series of astronomers that through their

C. Montúfar (✉)
Universidad San Francisco de Quito, Quito, Ecuador
e-mail: cmontufar@usfq.edu.ec

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_2, © Springer Science+Business Media New York 2013

observations, experiments, and hypothesis laid the foundations of modern science: Copernicus, Brahe, Galileo, and Kepler. Nicolaus Copernicus (1473–1543) challenged the view of the universe of his time. Deeply embedded in religious thought and described by Ptolemy (Claudius Ptolemaeus, c. 90–c. 168 AD), the geocentric model of the universe had gone unchallenged for centuries, although other Greek philosophers and astronomers had proposed a heliocentric view. In 1543, Copernicus published *De Revolutionibus Orbium Coelestium*, whose heliocentric model was to begin a series of scientific breakthroughs culminating with Darwin's grand Theory of Evolution and modern evolutionary biology. Be it coincidence or simply that Ecuador is a unique natural laboratory, these discoveries were intimately tied to a series of events that occurred in its territory between the eighteenth and nineteenth centuries leading to Darwin's visit to the Galapagos Islands. Following Copernicus, along with the incredible observations of Tycho Brahe (1546–1601) and the work of Johannes Kepler (1571–1630), Galileo Galiei (1564–1642) took forth upon himself to establish once and for all the Copernican model of the universe. History and science owe him a privileged place in the annals of knowledge. With the invention of the telescope, astronomy changed forever. Rigorous observation, experimentation, and measurement became key steps to scientific development. Measuring the Earth and observing life within Ecuador confirmed and sparked two major current scientific paradigms: we are not in the center and we are not immutable. Sir Isaac Newton (1643–1727), despite his genius, could not have guessed that his greatest work, the Theory of Gravitation, was to have a decisive vote in its favor, a few miles from a place that was probably unknown to him: Quito, Ecuador. With Newton, Copernicus, and Copernicus' followers, science had finally a consistent and elegant theory of planetary motion capable of placing man on the moon. And, as is true in case of all sciences, experimental evidence was necessary. Newton's law of universal gravitation predicted that the Earth should be flat at the poles and had a shape of an oblate spheroid. The equatorial radius should be larger than the polar radius. This was contrary to the Cartesian model defended by Giovanni Domenico Cassini (1625–1712), an Italian-born French astronomer (later Jean-Dominique) who, along with his son Jacques, carried out measurements in France suggesting that the Earth was elongated at the poles. Charles C. Gillispie in his book *Science and Polity in France: The End of the Old Regime* (Gillispie 1980) states: "*After the publication of Newton's Principia in 1687, it came to seem that such down-to-earth techniques of land surveying should, in principle, be competent to resolve the largest issue of cosmology, the choice between Cartesian and Newtonian theories of the world*." The Paris Academy of Sciences promoted a series of expeditions to resolve this issue. Gerald James Holton and Stephen G. Brush (Brush 2005) in their book entitled "*Physics, The Human Adventure*: *From Copernicus to Einstein and Beyond*" state: "*The results of this project, perhaps the first ever large cooperative government-funded effort to resolve a scientific question, provided a decisive confirmation of Newton's theory*." The French geodesic mission led by Charles Marie de La Condamine (1701–1774) was to measure the meridian arc at the equator, and Pierre Louis Maupertuis (1698–1759) led the mission to Lapland. La Condamine, Louis Godinand, and Pierre Bouguer arrived in what today is Ecuador in the early eighteenth

century accompanied by two Spanish overseers, Jorje Juan and Antonio de Ulloa. After a long 7-year stay, the results of their measurements, along with the results of Maupertuis in Lapland, confirmed Newton's predictions and discarded the Cartesian hypothesis of the Earth's elongation at the poles (Tristan 1999). La Condamine's return to Europe in 1744 via the Amazon River is considered one of the first scientific explorations of this vast tropical waterway. La Condamine's work influenced a young German naturalist, Alexander Von Humboldt (1769–1859), whose travels in the early nineteenth century allowed him to produce some of the most extensive works on geography and nature ever published. Humboldt's *Personal Narrative* refers repeatedly to La Condamine's journal. In 1802, Humboldt arrived in Ecuador and among other feats, attempted to climb mount Chimborazo, considered the world's highest peak, and reached an altitude of 5,875 m, the highest ever attained by a human at that time. His work as naturalist in the region and in the Ecuadorian territory acclaimed him to world fame. Humboldt is considered the father of ecology and modern geography. Humboldt visited Thomas Jefferson in Virginia, where left a lasting impression. In a letter to Caspar Wistar (Jefferson 1804), Jefferson states: "*I have omitted to state… the extreme satisfaction I have received from Baron Humboldt's communications. The treasures of information which he possesses are inestimableand fill us with impatience for their appearance in print*."

Humboldt's writings certainly influenced the young Charles Darwin (1809–1882) who had read most of the German scientist's works. Humboldt's *Personal Narrative* was a cherished companion in the Beagle and is continuously mentioned with appraisal in Darwin's Diary. In a letter to J. S. Henslow (Darwin 1985), Darwin writes: "*I formerly admired Humboldt, I now almost adore him; he alone gives any notion, of the feelings which are raised in the mind on first entering the Tropics*…." Charles Darwin arrived in what is today San Cristóbal Island in the Galapagos Archipelago in October 1835. Darwin's legacy of his observations in this Archipelago is without a doubt the major paradigm shift of modern science. In 1858, Alfred Russel Wallace (1823–1913), who had also read Humboldt and had traveled in the Amazon, proposed a similar theory of natural selection. Copernicus, Kepler, Galileo, Newton, La Condamine, Darwin, and Wallace are all major contributors to the paradigm shifts that altered the way humans think of their relevance in the universe. They all touched Ecuador, directly or indirectly. The year 2009, 200 years from Darwin's birthday and 150 years from Humboldt's death, provides a unique opportunity to celebrate these magnificent series of historical events.

Albert Einstein, at a young age, read both Humboldt and Darwin (Pais 1982)….

References

Bronowski J (1973) The ascent of man. Little, Brown and Co., Boston

Brush GH (2005) Physics, the human adventure. Rutgers University Press, New Bunswick

Darwin C (1985) Letter to J.S. Henslow 18 May, 1832. University of Kansan Digital Library, 1821–1836. The Correspondence of Charles Darwin, vol I. Cambridge University Press, New York, pp 236–238

Jefferson T (1804) Letter to Caspar Wistar, June 7. University of Kansan Digital Library
Gillispie CC (1980) Science and polity in France at the end of the old regime. Princeton University Press, Princeton
Pais A (1982) Subtle is the lord: the science and the life of Albert Einstein. Oxford University Press, New York
Tristan F (1999) El Proceso con las Estrellas. LibriMundi, Quito

Part II
A Microbial World

Chapter 3
How Did Life Originate?

Antonio Lazcano

Introduction

Like Lamarck, Buffon, and his own grandfather Erasmus, Charles Darwin was convinced that plants and animals had evolved because of natural processes from simple inorganic compounds. Although Darwin rejected the idea that putrefaction of preexisting organic compounds could lead to the appearance of organisms, the pages he excised from his private notebooks show that he was convinced as early as 1837 that "the intimate relation of life with laws of chemical combination, and the universality of latter render spontaneous generation not improbable." Nonetheless, Darwin consciously avoided discussing the origin of life in the *Origin of Species*, where he only wrote that "all the organic beings which have ever lived on this Earth may be descended from some primordial form, into which life was first breathed" (Peretó et al. 2009).

Darwin's biblical reference drew considerable criticism both from friends and foes. In 1863, Richard Owen argued in a review of W. B. Carpenter's *Introduction to the Study of the Foraminifera* that microscopic organisms could periodically appear spontaneously in mud because of an undefined "general polarizing force. " Darwin, added Owen, "could only express" the creative force responsible for the origin of life "in Pentateuchal terms as the primordial form into which life was first breathed!" (Peretó et al. 2009).

A few weeks later, Darwin answered Owen's criticisms in a letter published in the *Athenaeum*, and stated that:

"I hope that you will permit me to add a few remarks on Heterogeny, as the old doctrine of spontaneous generation is now called, to those given by Dr. Carpenter, who, however, is probably better fitted to discuss the question than any other man in England. Your reviewer", wrote Darwin, "believes that certain lowly organized

A. Lazcano (✉)
Facultad de Ciencias, Universidad Nacional Autónoma de México, México City, Mexico
e-mail: alar@ciencias.unam.mx

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_3, © Springer Science+Business Media New York 2013

animals have been generated spontaneously—that is, without pre-existing parents—during each geological period in slimy ooze. A mass of mud with matter decaying and undergoing complex chemical changes is a fine hiding-place for obscurity of ideas. But let us face the problem boldly. He who believes that organic beings have been produced during each geological period from dead matter must believe that the first being thus arose. There must have been a time when inorganic elements alone existed on our planet: let any assumptions be made, such as that the reeking atmosphere was charged with carbonic acid, nitrogenized compounds, phosphorus, &c. Now is there a fact, or a shadow of a fact, supporting the belief that these elements, without the presence of any organic compounds, and acted on only by known forces, could produce a living creature? At present it is to us a result absolutely inconceivable. Your reviewer sneers with justice at my use of the 'Pentateuchal terms,' of one primordial form into which life was first breathed: in a purely scientific work I ought perhaps not to have used such terms; but they well serve to confess that our ignorance is as profound on the origin of life as on the origin of force or matter."

A few years later, Thomas H. Huxley declared in a lecture that "the researchers of the chemistry have revealed that all protoplasm is proteinaceous." Like Huxley, Hooker, Haeckel, and other close friends and followers, Darwin was very much aware how the interplay between chemistry and biology had led to the association of biological phenomena with "living proteins" and with the intracellular gel-like matter that Carl Heinrich Schultz had termed "protoplasm" (Fruton 1986). In 1871, he mailed to Hooker the now famous letter in which he not only referred to Pasteur's experiments on spontaneous generation but also discussed the possibility of the origin of life in a "warm little pond"—but adding a cautious "big if":

My dear Hooker,

I return the pamphlets, which I have been very glad to read.–It will be a curious discovery if Mr. Lowe's observation that boiling does not kill certain molds is proved true; but then how on earth is the absence of all living things in Pasteur's experiments to be accounted for?–I am always delighted to see a word in favour of Pangenesis, which some day, I believe, will have a resurrection. Mr. Dyer's paper strikes me as a very able Spencieran production.

It is often said that all the conditions for the first production of a living organism are now present, which could ever have been present. But if (and oh what a big if) we could conceive in some warm little pond with all sorts of ammonia and phosphoric salts,–light, heat, electricity &c. present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present day such matter will be instantly devoured, or absorbed, which would not have been the case before living creatures were formed.

Henrietta makes hardly any progress, and God knows when she will be well.

I enjoyed much the visit of you four gentlemen, i.e., after the Saturday night, when I thought I was quite done for.

Yours affecty
C. Darwin

At the Shores of a Warm Little Pond

The demise of Darwin's idea of natural selection at the turn of the twentieth century had a major impact in the way the origin of life was explained. As shown by the work of Hermann J. Muller, not all attempted to approach the issue of the emergence of living systems from a Darwinian perspective. A member of Thomas Hunt Morgan's fly room in Columbia University, Muller had been pondering for some time on the autocatalytic properties of chromatin (Ravin 1977). Although the actual nature of genes was yet not understood, in 1922, he speculated that the gene "... reacts in such a way as to convert some of the common surrounding material into an end-product identical in kind with the original gene itself. This action fulfills the chemist's definition of 'autocatalysis' ... but the most remarkable feature of the situation is not this oft-noted autocatalytic action in itself—it is the fact that, when the structure of the gene becomes changed, through some 'chance variation', the catalytic property of the gene may be correspondingly changed, in such a way as to leave it still autocatalytic" (Muller 1922).

The physicist Leonard Troland suggested that the origin of life was the result of the random formation of a self-replicating "genetic enzyme" (Troland 1917), but it did not take long for Muller to transform Troland's views and to argue that the first living system consisted of little more than a mutable gene, or set of genes, endowed with catalytic and autoreplicative properties (Muller 1922, 1926). Muller was not a darwinist, and his appreciation of genetic mutation as the fundamental mechanism of evolutionary novelties developed at a time when the appeal of Darwin's ideas on the role of natural selection had diminished and there was a major split separating geneticist from naturalist. Accordingly, given the appearance of a genetic material capable of replication, mutation, and further replication of mutant forms, Muller argued that "evolution would automatically follow" (Pontecorvo 1982).

In November 1923, however, a small book titled *The Origin of Life* provided an alternative explanation based on a Darwinian perspective (cf. Oparin 1967). It had been written by Alexander Ivanovich Oparin, a young Russian biochemist who, like many of his contemporaries, accepted the idea of a primordial protoplasm but proposed that life had been preceded by a lengthy period of abiotic syntheses and accumulation of organic compounds that had led to the accumulation of what we call today the primitive soup. Like many of his fellow students and colleagues, Oparin was well acquainted with the ideas of Haeckel and others that assumed that the first forms of life had been autotrophic microbes. However, it was impossible for him to accept that life had emerged already endowed with an autotrophic metabolism that included enzymes, chlorophyll, and the ability to synthesize organic compounds from CO_2 and water.

As a young student at the University of Moscow, Oparin had joined the laboratory of Alexei N. Bakh, an eminent scientist and political figure at the Karpov Physico-chemical Institute. There he worked on photosynthesis and, like most biochemists of his generation, quickly adopted the idea that metabolism was the outcome of oxidation and reduction reactions that were coupled inside cells. By then, Oparin

was also a convinced evolutionist. As an undergraduate, he had attended the lectures given regularly by Kliment A. Tymiriazev, a renowned plant physiologist, agronomer, and the main advocate of Darwinism in Russia. Tymiriazev had left the university but organized weekly meetings with students and colleagues in his Moscow apartment. By the time Oparin graduated, he had an academic background that combined natural history, biochemistry, and plant physiology, a knowledge acquired within a research tradition strongly committed to integral approaches in the analysis of natural phenomena. He was not only familiar with nearly all the literature on evolution available in Russia but also, perhaps even more important, with the Darwinian method of comparative analysis and historical interpretation of life features (Lazcano 2010).

As a heterotrophic anaerobe is metabolically simpler than an autotrophic one, Oparin argued, the former would necessarily have evolved first. Thus, based on the simplicity and ubiquity of fermentative reactions, he proposed that the first organisms must have been heterotrophic bacteria that could not make their own food but obtained organic material present in the primitive milieu. Like Oparin, the geochemist Charles Lipman, the microbiologist R. B. Harvey, and the geneticist and polymath John B. S. Haldane also proposed a heterotrophic emergence of life (Bada and Lazcano 2003). Based on experiments by British chemist E. C. C. Baly, who claimed that he had synthesized amino acids (Baly 1924) and sugars by the ultraviolet (UV) irradiation of a solution of CO_2 in water (Baly et al. 1927), Haldane (1929) suggested that the absence of oxygen in a CO_2-rich primitive atmosphere had led to the synthesis of organic compounds and the formation of a "hot dilute soup." The discovery of phages led Haldane to suggest that viruses could represent an intermediate step in the transition from the prebiotic soup to the first heterotrophic cells. Life, wrote Haldane (1929) could have remained "in the virus stage for many millions of years before a suitable assemblage of elementary units was brought together in the first cell."

In 1936, Oparin published a second volume in Russian also called *Origin of Life*, whose English translation became available 2 years later (Oparin 1938). This book is a masterpiece of evolutionary analysis, and in it, Oparin critically revised his original proposal, and argued for a highly reducing primitive milieu in which iron carbides of geological origin would react with steam to form hydrocarbons. Their oxidation would yield alcohols, ketones, aldehydes, etc., that would then react with ammonia to form amines, amides, and ammonium salts. The resulting protein-like compounds and other molecules would form a dilute solution, where they would aggregate to form colloidal systems from which the first heterotrophic microbes evolved. Following Bungenberg de Jong's proposal that the colloid properties of droplets, which he termed coacervates, formed by the spontaneous aggregation of biological macromolecules could explain the properties of protoplasm, Oparin proposed them as precursors of the first cells (Oparin 1938).

It is frequently forgotten that the idea that bacteria lacked genes was surprisingly enduring. As later as 1942, Julian Huxley wrote in *Evolution: The Modern Synthesis* that bacteria "have no genes in the sense of accurately quantized portions of hereditary substances; and therefore have no need for accurate division of the genetic system which is accomplished by mitosis" (Huxley 1942). It is, therefore, not altogether surprising that Oparin assumed that biological inheritance was the outcome of

growth and division of the coacervate drops, and that he considered enzymatic-based assimilation, growth, and reproduction traits of life, but not nucleic acids, whose role as the material basis of inheritance was not even suspected when he published his second book. In hindsight, this may seem naïve, but Oparin's commitment to coacervates resulted in part from his refusal to assume that life can be reduced to a single compound such as the randomly formed "living gene" proposed by Muller.

Prebiotic Chemistry: Was There a Primitive Soup?

Metaphors, Francis Bacon once wrote, should be considered "rather as a pleasure or play of wit that a science." If not misunderstood, however, metaphors help to convey powerful images. Synonymous terms like "primitive soup," "primordial broth," or "Darwin's warm little pond" have led to major misunderstandings in some cases, including the simplistic image of a worldwide ocean rich in self-replicating molecules and accompanied by all sorts of biochemical monomers. However, it refers to parts of the prebiotic environment where the accumulation and interaction of the products of abiotic synthesis may have taken place, including oceanic sediments, intertidal zones, shallow ponds, membrane-bound systems, fresh water lakes, lagoons undergoing wet and dry cycles, and glacial ponds where evaporation, eutectic separations, or other physicochemical mechanisms, such as the adherence of biochemical monomers to active surfaces, could have raised local concentrations and promoted polymerization (Bada and Lazcano 2009).

The Oparin–Haldane proposal is based on the assumption that the syntheses and accumulation of abiotic organic compounds were a necessary precursor for the appearance of life. Experimental evidence in support of the idea that the first living beings were anaerobic heterotrophs came first from Harold C. Urey's laboratory, who had been involved with the study of the origin of the solar system and the chemical events associated with this process. Urey had also considered the origin of life in the context of his proposal of a highly reducing terrestrial atmosphere (Urey 1952). The first successful prebiotic amino acids synthesis was carried out with an electric discharge and a strongly reducing model atmosphere of CH_4, NH_3, H_2O, and H_2 (Miller 1953). The result of this experiment was a large yield of a racemic mixture of amino acids, together with hydroxy acids, short aliphatic acids, and urea. One of the surprising results of this experiment was that the products were not a random mixture of organic compounds; rather, a relatively small number of compounds were produced in substantial yield. Moreover, with a few exceptions, the compounds were of biochemical significance.

The mechanism of synthesis of the amino acids and hydroxy acids formed in the spark discharge experiment was investigated in considerable detail (Miller 1955). The presence of large quantities of hydrogen cyanide, aldehydes, and ketones in the water flask, which were clearly derived from the methane, ammonia, and hydrogen originally included in the apparatus, demonstrated that the amino acids were the outcome of a Strecker-like synthesis that involved aqueous phase reactions of highly

reactive intermediates. Contemporary reassessments of the Miller–Urey experiments demonstrate that prebiotic synthesis of amino acids and amines can also take place via the Bucherer–Bergs pathway and the hydrolysis of HCN polymers, which can also be expected to have been quite efficient (Cleaves et al. 2008).

A few years after the Miller experiment, Juan Oró, who had been studying the synthesis of amino acids from an aqueous solution of HCN and NH_3, reported the abiotic formation of adenine (Oró 1960). The synthesis is indeed remarkable. If concentrated solutions of ammonium cyanide are refluxed for a few days, adenine is obtained in up to 0.5 % yield along with 4-aminoimidazole-5 carboxamide and the usual cyanide polymer (Oró 1960; Oró and Kimball 1961). This reaction, proceeds through the self-condensation of HCN to give diaminomaleonitrile, which according to Oró and Kimball (1961), then reacts with formamidine to give adenine.

The remarkable ease by which adenine can be synthesized by the aqueous polymerization of ammonium cyanide demonstrated the significance of HCN and its derivatives in the primitive environments (Oró 1960) and played a major role in the convergence of prebiotic chemistry and molecular biology. The prebiotic importance of HCN has been further substantiated by the discovery that the hydrolytic products of its polymers include amino acids, purines, and orotic acid, which is a biosynthetic precursor of uracil (Ferris et al. 1978). The reaction of urea with cyanoacetylene or cyanoacetaldehyde, which is a hydrolytic derivative of HCN (Ferris et al. 1968), leads to high yields of cytosine and uracil, especially under simulated evaporating pond conditions, which increase the urea concentration (Robertson and Miller 1995). However, it is possible that not all biochemical components are formed by monomers. Although polyphosphates and phosphate esters are not prebiotic reagents, the abiotic formation of cytidine ribonucleotides by the assembly of the heterocycle base on a sugar phosphate recently reported by John Sutherland's group indicates that the assumption that all prebiotic syntheses proceed by the stepwise joining of its components should be revised (Powner et al. 2009).

It is likely that other geological sources of reductants, such as pyrite, may have been available. The FeS/H_2S combination is a strong reducing agent that has been shown to provide an efficient source of electrons for the reduction of organic compounds under mild conditions. This possibility, which was developed in the context of the pyrite-dependent autotrophic model of the origin of life first suggested by Wächtershäuser (1988), is also consistent with the heterotrophic theory. Although pyrite-mediated CO_2 reduction to amino acids, purines, and pyrimidines is yet to be achieved, the FeS/H_2S combination is a strong reducing agent that has been shown to reduce nitrogen, nitrate, and acetylene, to induce peptide bond formation by the activation of amino acids with carbon monoxide and (Ni, Fe)S, and to promote the synthesis of acetic acid and pyruvic acid, which have also been synthesized from CO under simulated hydrothermal conditions.

The ease of formation of amino acids, purines, and pyrimidines in one-pot reactions supports the contention that these molecules were present in the prebiotic broth. Laboratory simulations suggest that urea, alcohols, and sugars formed by the nonenzymatic condensation of formaldehyde, a wide variety of aliphatic and aromatic hydrocarbons, urea, carboxylic acids, and branched and straight fatty acids,

including some that are membrane-forming compounds, were also components of the primitive soup. The basic tenet of the heterotrophic theory of the origin of life is that the maintenance and reproduction of the first living entities depended primarily on prebiotically synthesized organic molecules. There has been no shortage of discussion about how the formation of the primordial soup took place, but it is very unlikely that any single mechanism can account for the wide range of organic compounds that may have accumulated on the primitive Earth. In addition, an analysis of Oró's 1961 suggestion on the role of cometary nuclei as sources of volatiles to the primitive Earth has lead to the reassessment of the proposal that the exogenous delivery of organic matter by asteroids, comets, and interplanetary dust particles could have played a significant role in the prebiotic accumulation of the compounds necessary for the origin of life (Oró 1961; Chyba and Sagan 1992).

It is thus likely that exogenous sources of organic compounds also contributed to the synthesis of the primitive soup. The major sources of exogenous compounds would appear to be comets and dust, with asteroids and meteorites being minor contributors. Asteroids would have impacted the Earth frequently during the Hadean and early Archean, but the amount of organic material brought in would seem to be small, even if the asteroids are assumed to be Murchison meteorite-type objects. Comets are one of the most promising sources of exogenous compounds (Oró and Lazcano 1997). As summarized elsewhere (Bada and Lazcano 2009), it is reasonable to assume that the atmosphere that developed on the Earth during the period 3.8–4.4 billion years ago was essentially a mix of volatiles delivered by bodies such as cometary nuclei, combined with the products of outgassing processes from the interior of an already differentiated planet.

Given adequate expertise and experimental conditions, it is possible to synthesize almost any organic molecule. However, the fact that a number of molecular components of contemporary cells can be formed nonenzymatically in the laboratory does not necessarily mean that they were also essential for the origin of life or that they were available in the prebiotic environment. The primitive soup must have been a bewildering organic chemical wonderland, but it could not include all the compounds or molecular structures found today in even the oldest prokaryotes. There are a number of other compounds that have been synthesized under primitive Earth conditions and that we have not discussed here. These include dicarboxylic and tricarboxylic acids, fatty acids, fatty alcohols, porphins, and several others. On the other hand, there are some biochemical monomers that do not yet have adequate prebiotic synthesis, such as histidine, pyridoxal, thiamine, among others. It is possible that some of these compounds may not have been synthesized prebiotically, so their occurrence in living systems may have been the result of early metabolic syntheses.

As discussed here, the existence of different abiotic mechanisms by which biochemical monomers can be synthesized under plausible prebiotic conditions is well established. During the past few years, laboratory simulations of prebiotic synthesis have been developing models of specific detailed environments, including those that may have been provided by the surface of clays, volcanic small ponds, and liposomes. These are probably more realistic simulations. Of course, not all prebiotic pathways are equally efficient, but the wide range of experimental conditions under

which organic compounds can be synthesized demonstrates that prebiotic syntheses of the building blocks of life are robust, i.e., the abiotic reactions leading to them do not take place under a narrow range defined by highly selective reaction conditions, but rather under a wide variety of experimental settings.

The remarkable coincidence between the molecular constituents of living organisms and those synthesized in prebiotic experiments is too striking to be fortuitous, and the robustness of this type of chemistry is supported by the occurrence of most of these biochemical compounds in the 4.5 billion-year-old Murchison carbonaceous chondrite and in other carbon-rich meteorites. How first life evolved is not known, but analysis of carbonaceous chondrites and the laboratory simulations of the primitive Earth suggest that prior to the emergence of the first living systems, the prebiotic environment was endowed with (a) a large suite of organic compounds of biochemical significance; (b) many organic and inorganic catalysts (such as cyanamide, metallic ions, sulfur-rich minerals, and clays); (c) purines and pyrimidines, i.e., the potential for template-dependent polymerization reactions; (d) membrane-forming compounds; and (e) the availability of many possible sources of carbon and nitrogen for primordial heterotrophs.

The RNA World: What's in a Name?

The 1936 report by C. F. Bawden, N. C. Pirie, J. D. Bernal, and I. Frankuchen that preparations of the tobacco mosaic virus (TMV) contained "nucleic acid of the ribose type" was eventually followed by the demonstration by the groups led by Gerhard Schramm and Heinz Fraenkel-Conrat that the infectivity of the TMV was due to RNA (Bawden et al. 1936). We do not know when and how viruses like the TMV that have RNA genomes originated. As Holland (1993) wrote, "today's biosphere is DNA-based, and the only known RNA 'life forms' are RNA viruses. These depend on DNA-based host cells for their existence. Whether any existing RNA virus genomes contain sequence vestiges of the earliest RNA life forms is unknown (and probably unknowable), but it is clear that RNA viruses are ubiquitous, extremely successful cellular parasites of considerable disease importance."

However, the awareness of the multiple roles of RNA had a major impact on our understanding of the origin and early evolution of life. It is true as N. C. Pirie remarked during the famous 1957 Origin of Life meeting in Moscow that "I do not think that a discussion of the intimate details of the habits of the tobacco mosaic virus (TMV) has any strict bearing on the origin of life" (Pirie 1959). Combined with the discovery that RNA plays a key role in protein synthesis (Brachet 1942), the evidence that RNA could be a carrier of genetic information had major consequences in the development of the ideas on the origin of life (Brachet 1959; Belozerskii 1959). After noting that the evidence indicated that RNA molecules played a key role in protein synthesis, Belozerskii wrote that "There is no doubt that nucleic acids played an important role in the evolution of the organic world and metabolic reactions. Yet both RNA and DNA could hardly arise simultaneously in the early evolution

of life. It rather seems that ribonucleotides, and then RNA, originated first. DNA came into existence far more recently, as the protoplasm became more differentiated and its functions grew in complexity. It seems that RNA, being associated with the most general processes of life, was formed at an earlier evolutionary stage, while the origin of DNA was associated with the development of more specialized and phylogenetically later features of organisms" (Belozerskii 1959).

Although the idea that RNA came before DNA was further supported by the recognition that ribonucleotides are the metabolic precursors of deoxyribonucleotides and that RNA primers are required for DNA replication, the issue of the origin of the functional relationship between nucleic acids and protein remained an open one. From the late 1960s onward, however, it became clear that the understanding of the origin of life was troubled by the emergence of nucleic acid-directed protein synthesis, which is recognized as a central feature of all extant life, appeared to be an insurmountable problem. As Monod wrote in his *Chance and Necessity*, "... it might be thought that the discovery of the universal mechanisms basic to the essential properties of living beings would have helped solve the problem of life's origins. As it turns out, these discoveries, by almost entirely transforming the question, have shown it to be even more difficult than it formerly appeared" (Monod 1971).

In fact, a possible solution to this issue had been suggested by Rich (1962), Woese (1967), Orgel (1968), and Crick (1968), who independently proposed the idea that the first living entities were based on RNA as both the genetic material and catalyst. As written by Crick (2006), this possibility was largely forgotten. However, the discovery of ribozymes in the early 1980s by the groups of Thomas Cech and of Sidney Altman, gave considerable credibility to the idea that the first living organisms were based on RNA as both the genetic material and as catalyst, an hypothetical stage called the RNA world (Gilbert 1986; Joyce 2002). It is possible that the ribonucleotide moieties of many coenzymes, which are small organic molecules that provide a varied set of reactive groups to catalytic proteins, are in fact molecular fossils of a primordial metabolic state and reflect recruitment processes that diversified the catalytic abilities of ribozymes, providing new venues for an increasingly complex RNA world metabolism (Orgel and Sulston 1971; White 1976).

The surprising widening of the catalytic repertoire of RNA thanks to the SELEX experiments that allow the evolution of new catalytic abilities to appear under selection pressures and to catalyze an increasingly large number of chemical reactions has lent strong support to the possibility of the so-called RNA world and greatly simplifies the understanding of the origin of protein biosynthesis and of the genetic code. Until a few years ago, the origin of the genetic code and of protein synthesis was considered synonymous with the appearance of life itself. This is no longer a dominant point of view.

The experimental evidence demonstrating that ribozymes can mediate four of the basic reactions involved in protein biosynthesis, i.e., amino acid activation, aminoacyl-RNA synthesis, peptide bond formation, and RNA-based coding, suggests that ribosome-mediated protein synthesis first evolved in an RNA world (Kumar and Yarus 2001). In fact, the demonstration that ribosomal peptide synthesis is a ribozyme-catalyzed reaction makes it almost certain that there was once an RNA

world and that protein biosynthesis is one of its evolutionary outcomes. The extraordinary structural and functional complexity of extant ribosomes must have been preceded by simpler structures. The observations that have shown that peptide bond formation occurs in a highly conserved site devoid of proteins (Ban et al. 2000; Nissen et al. 2000) formed by two 60-ribonucleotide L-shaped RNA core units, which appear to be the outcome of an early duplication (Agmon 2009), is consistent with the hypothesis that protein biosynthesis first evolved in an RNA world and that the original protoribosome lacked proteins, i.e., the ribosome-catalyzed nucleic acid-coded protein synthesis is the outcome of Darwinian selection of protein-free RNA-based biological systems and not of mere physicochemical interactions that took place in the prebiotic environment.

This possibility is now widely accepted, but the chemical lability of RNA components suggests that this molecule was not a direct outcome of prebiotic evolution but may have been one of the evolutionary outcomes of what are now referred to as pre-RNA worlds. The RNA world hypothesis, however, does not imply that life should be stripped of its identity and reduced to a mere collection of autocatalytic RNA molecules. There are many definitions of the RNA world, but they do not imply that ribozymes suddenly appeared on the primitive Earth endowed with the miraculous ability to construct a fully functional living being. Although at the time being the hiatus between the primitive soup and the RNA world is discouragingly enormous, the hypothesis that life was preceded by a prebiotic broth still provides the most useful evolutionary framework for addressing the issue of a stepwise (but not necessarily slow) emergence of life on Earth. This does not imply that wriggling autocatalytic nucleic acid molecules were floating in the waters of the primitive oceans, ready to be used as primordial genes, or that the RNA world sprung completely assembled from simple precursors present in the prebiotic soup. Precellular evolution was not a continuous, unbroken chain of progressive transformations steadily proceeding to the first living beings. Many prebiotic *culs-de-sac* and false starts probably took place, with natural selection acting over populations of primordial systems based on genetic polymers simpler than RNA.

The chemical nature of the first genetic polymers and the catalytic agents that may have formed the pre-RNA worlds that bridged the gap between the prebiotic broth and the RNA world are completely unknown and can only be surmised. Modified nucleic acid backbones have been synthesized, which either incorporate a different version of ribose or lack it altogether. Experiments on nucleic acid with hexoses instead of pentoses and on pyranoses instead of furanose (Eschenmoser 1999) suggest that a wide variety of informational polymers is possible, even when restricted to sugar phosphate backbones. One possibility that has not been explored is that the backbone of the original informational macromolecules may have been atactic (e.g., disordered) kerogen-like polymers such as those formed in some prebiotic simulations. There are other possible substitutes for ribose, including open chain, flexible molecules that lack asymmetric carbons. One of the most interesting chemical models for a possible precursor to RNA involves the so-called peptide nucleic acids (PNAs), which have a polypeptide-like backbone of achiral 2-amino-ethyl-glycine, to which nucleic acid bases are attached by an acetic acid (Nielsen 1993). Such molecules form very stable

complementary duplexes, both with themselves and with nucleic acids. Although they lack ribose, their functional groups are basically the same as in RNA, so they may also be endowed with catalytic activity.

The pre-RNA world hypothesis does not imply that genetic polymers could only evolve from simpler genetic polymers in a never-ending succession of genetic takeovers but rather indicates the need to study, under plausible prebiotic conditions, very simple monomers and genetic polymers that could serve as laboratory models of the possible evolutionary precursors of RNA. The assumption that a single molecule once served both as the depositary of information storage and biological catalyst is not necessarily married to a reductionist approach that assumes that life can be assigned to such compound. It is possible that the emergence of life is best understood in terms of the dynamics and evolution of systems of chemical replicating entities endowed with genetic polymers.

Whether these entities were enclosed within membranes is not yet clear, but given the prebiotic availability of amphiphilic compounds, this may have well been the case. It is reasonable, for instance, to assume that the invention of protein synthesis and the encapsulation of reaction machinery needed for replication may have taken place during the RNA world (Schrum et al. 2010). On the other hand, the manifold roles that RNA molecules play, such as riboswitches, microRNAs (miRNAs), and piwi-interacting RNAs (piRNAs), in the regulation of gene expression also support the idea that RNA played a major role in early stages of biological evolution (Yarus 2010). Alarmones, which are relatively simple small ribonucleotide derivatives like cyclic AMP and ZTP that are activated when cells sense stress, including starvation or environmental insults, are made by nucleotide-modifying biosynthetic pathways. Together with a wide variety of RNA-dependent control mechanisms such as riboswitches, miRNAs, and piRNAs, the structural simplicity and biological distribution of alarmones suggest that they may have appeared in RNA/protein cells prior to the evolution of DNA genomes.

It is likely that by the time Archaea, Bacteria, and Eukarya diverged, DNA had already been selected as the cell genetic polymer. The sequence similarities shared by many ancient, large proteins found in all three domains (Becerra et al. 2007) suggest that considerable fidelity existed in the operative genetic system of their common ancestor, but such fidelity is unlikely to be found in RNA-based genetic systems (Lazcano et al. 1992). When did the transition from RNA to DNA cellular genomes take place? As is often the case with metabolic innovations, the dating of such a step cannot be documented from geological data. However, since all extant cells are endowed with DNA genomes, the most parsimonious conclusion is that this genetic polymer was already present in the last common ancestor (LCA), or cenancestor, that existed prior to the divergence of the three primary domains, i.e., the Bacteria, Archaea, and Eukarya. As discussed elsewhere (Delaye et al. 2005), although there have been a number of suggestions that the LCA (or its equivalents) was endowed with genomes formed by small-sized RNA molecules or hybrid RNA/DNA genetic system, there are manifold indications that double-stranded DNA genomes of monophyletic origin had become firmly established prior to the divergence of the three primary domains.

Early Biological Evolution

The extraordinary similarities at very basic biochemical and genetic levels among all know life forms can be interpreted as propinquity of descent, i.e., all organisms are of monophyletic origin. Although complex, multigenic traits must have evolved through a series of simpler states, no evolutionary intermediate stages or ancient simplified versions of ATP production, DNA replication, or ribosome-mediated protein synthesis have been discovered in extant organisms. The molecular details of these universal processes not only provide direct evidence of the monophyletic origin of all extant forms of life but also imply that the sets of genes encoding the components of these complex traits were frozen a long time ago, i.e., major changes in them are strongly selected against.

The variations of traits common to extant species can be easily explained as the outcome of divergent processes from an ancestral life form that existed prior to the separation of the Bacteria, the Archaea, and the Eukarya. Universal gene-based phylogenies ultimately reach such single universal entity, which very likely was part of a population of similar entities that existed throughout the same period. They may have not survived, but some of their genes may have if they became integrated via lateral transfer into the LCA genome. As reviewed elsewhere (Becerra et al. 2007), the cenancestor should be seen as one of the last evolutionary outcomes of a series of ancestral events including lateral gene transfer, gene losses, and paralogous duplications that took place before the separation of the three major cell lineages. Recognition that cellular genomes are historical documents recording at least part of past evolutionary events has allowed important insights into simpler biological systems that appear to have lacked DNA genomes but that can be considered basically RNA/proteins cells far removed, if not in time, at least in complexity with respect to the first living systems. No paleontological remains will bear testimony of its existence, as the search for a fossil of the cenancestor is bound to prove fruitless. From a cladistic viewpoint, the LCA is merely an inferred inventory of features shared among extant organisms, all of which are located at the tip of the branches of molecular phylogenies. It is, however, important to distinguish between ancient and primitive organisms. Organisms located near the root of universal rRNA-based trees are cladistically ancient, but they are not endowed with primitive molecular genetic apparatus nor do they appear to be more rudimentary in their metabolic abilities than their aerobic counterparts (Islas et al. 2003).

The high levels of genetic redundancy detected in all sequenced genomes imply not only that duplication has played a major role in the accretion of the complex genomes found in extant cells but also that prior to the early duplication events revealed by the large protein families, simpler living systems existed, which lacked the large sets of enzymes and the sophisticated regulatory abilities of contemporary organisms. One such group of duplicated genes is what we call paralogous genes, which are sequences that diverge not through speciation but after a duplication event. For instance, in all known cells, protein synthesis requires the presence of two homologous elongation factors EF-Tu and EF-G, which are GTP-dependent enzymes.

Evidence of their common ancestry indicates that prior to the duplication event that produced them, a more primitive, less-regulated version of protein synthesis was taking place with only a single ancestral elongation factor.

Reticulate phylogenies greatly complicate the reconstruction of cenancestral traits. Driven in part by the impact of lateral gene acquisition, as revealed by the discrepancies of different gene phylogenies with the canonical rRNA tree, and in part by the surprising complexity of the universal ancestor as suggested by direct backtrack characterizations of the oldest node of universal cladograms. Inventories of LCA genes include sequences that originated in different precenancestral epochs (Delaye and Lazcano 2000; Anantharaman et al. 2002). The origin of the mutant sequences ancestral to those found in all extant species, and the divergence of the Bacteria, Archaea, and Eukarya were not synchronous events, i.e., the separation of the primary domains took place later, perhaps even much later, than the appearance of the genetic components of their last common ancestor (Delaye et al. 2005). The cenancestor is thus one of the last evolutionary outcomes of a series of ancestral events including lateral gene transfer, gene losses, and paralogous duplications that took place before the separation of Bacteria, Archaea, and Eukarya (cf. Becerra et al. 2007).

Molecular cladistics may provide clues to some very early stages of biological evolution, but it is difficult to see how the applicability of this approach can be extended beyond a threshold that corresponds to a period of cellular evolution in which protein biosynthesis was already in operation, i.e., an RNA/protein world. Older stages are not yet amenable to molecular phylogenetic analysis. A cladistic approach to the origin of life itself is not feasible, as all possible intermediates that may have once existed have long since vanished, and the temptation to do otherwise is best resisted.

Conclusions

Almost a century and a half after Darwin admitted how little was understood about the origin of life, we still do not know when and how the first living beings appeared on Earth. It is difficult to ascertain the earliest traces of life, as most of the rocks from early Archean times that have been preserved have been metamorphosed to a considerable extent. There is no direct evidence of the environmental conditions on the Earth at the time of the emergence of life nor any fossil register of the predecessors of the first cells. Direct information is lacking not only on the chemical composition of the terrestrial atmosphere during the period of the origin of life but also on other general and local environmental conditions that may (or may not) have been important for the appearance of living systems. Multi- and interdisciplinary approaches have led to major advances in our understanding of the synthesis and accumulation of organic compounds of biochemical significance under possible primitive conditions and the catalytic abilities of ribozymes, therefore providing important insights and glimpses on the characteristics of the so-called RNA world. However, how the transition

between the so-called primitive broth and the RNA world took place is completely unknown and can only be surmised. Cladistic approaches to the origin of life itself are not feasible because all possible intermediates that may have once existed have long since vanished. However, the available evidence demonstrates that the chemical gap separating organisms from the nonliving is not insurmountable and that the origin of life was the outcome of a natural process. There are manifold historical records that allow us to reconstruct, with different degrees of precision, the evolutionary processes that preceded life beginings. How these different stages unfolded into one another and how life appeared remain open issues. However, the mere fact that can address this problem in evolutionary terms is, in itself, an extraordinary demonstration of the significance of Darwin's scientific inheritance.

Acknowledgments Support from UC Mexus-CONACYT Fellowship and DGAPA-UNAM is gratefully acknowledged. I am indebted to Dr. Amparo Latorre and Dr. Juli Peretó, of Institut Cavanilles de Biodiversitat i Biología Evolutiva, Universidad de Valencia, for their warm hospitality and support.

References

Agmon I (2009) The dimeric proto-ribosome: structural details and possible implications on the origin of life. Int J Mol Sci 10:2921–2934

Anantharaman V, Koonin EV, Aravind L (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res 30:1427–1464

Bada JL, Lazcano A (2003) Prebiotic soup: revisiting the Miller experiment. Science 300:745–746

Bada JL, Lazcano A (2009) The origin of life. In: Ruse M, Travis J (eds) The Harvard companion of evolution. Belknap/Harvard University Press, Cambridge, pp 49–79

Baly ECC (1924) Photosynthesis. Industrial Eng Chem 16:1016–1018

Baly ECC, Davies JB, Johnson MR, Shanassy H (1927) The photosynthesis of naturally occurring compounds. 1. The action of ultraviolet light on carbonic acid. Proc Roy Soc London 116:197–208

Ban N, Nissen P, Hansen J, Moore PB, Steitz T (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science 289:905–920

Bawden CF, Pirie NC, Bernal JD, Frankuchen I (1936) Liquid crystalline substances from virus-infected plants. Nature 138:1051–1054

Becerra A, Delaye L, Islas A, Lazcano A (2007) Very early stages of biological evolution related to the nature of the last common ancestor of the three major cell domains. Annu Rev Ecol Evol Syst 38:361–379

Belozerskii AN (1959) On the species specificity of the nucleic acids of bacteria. In: Oparin AI, Pasynskii AG, Braunshtein AE, Pavloskaya TE (eds) The origin of life on Earth. Pergamon, New York, pp 322–331

Brachet J (1942) La localisation des acides pentose nucléiques dans les tissues animaux et les oeufs d'Amphibiens en voie de développement. Arch Biol 53:207–257

Brachet J (1959) Les acides nucléiques et l'origine des proteins. In: Oparin AI, Pasynskii AG, Braunshtein AE, Pavloskaya TE (eds) The origin of life on Earth. Pergamon, New York, pp 361–367

Chyba CF, Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic compounds, an inventory for the origin of life. Nature 355:125–132

Cleaves JH, Chalmers JH, Lazcano A, Miller SL, Bada JL (2008) Prebiotic organic synthesis in neutral planetary atmospheres. Orig Life Evol Biosph 38:105–155
Crick FHC (1968) The origin of the genetic code. J Mol Biol 39:367–380
Crick FHC (2006) Foreword. In: Gesteland RF, Cech TR, Atkins FF (eds) The RNA World, vol. XII. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
Delaye L, Lazcano A (2000) RNA-binding peptides as molecular fossils. In: Chela-Flores J, Lemercchand G, Oró J (eds) Origins from the big-bang to biology: proceedings of the first Ibero-American school of astrobiology. Klüwer Academic Publishers, Dordrecht, pp 285–288
Delaye L, Becerra A, Lazcano A (2005) The last common ancestor: what's in a name? Orig Life Evol Biosph 35:537–554
Eschenmoser A (1999) Chemical etiology of nucleic acid structure. Science 284:2118–2124
Ferris JP, Sanchez RP, Orgel LE (1968) Studies in prebiotic synthesis. III. Synthesis of pyrimidines from cyanoacetylene and cyanate. J Mol Biol 33:693–704
Ferris JP, Joshi PD, Edelson EH, Lawless JG (1978) HCN, a plausible source of purines, pyrimidines, and amino acids on the primitive Earth. J Mol Evol 11:293–311
Fruton JS (1986) Protein, enzymes, genes. Yale University Press, New Haven
Gilbert W (1986) The RNA world. Nature 319:618
Haldane JBS (1929) The origin of life. Rationalist Annu 148:3–10
Holland JJ (1993) Replication error, quasispecies populations, and extreme evolution rates of RNA viruses. In: Morse SS (ed) Emerging viruses. Oxford University Press, New York
Huxley J (1942) Evolution: the modern synthesis. Allen & Unwin, London
Islas S, Velasco AM, Becerra A, Delaye L, Lazcano A (2003) Hyperthermophily and the origin and earliest evolution of life. Int Microbiol 6:87–94
Joyce GF (2002) The antiquity of RNA-based evolution. Nature 418:214–221
Kumar RK, Yarus M (2001) RNA-catalyzed amino acid activation. Biochemistry 40:6998–7004
Lazcano A (2010) Historical development of origins of life. In: Deamer DW, Szostak J (eds) Cold spring harbor perspectives in biology: the origins of life. Cold Spring Harbor Press, Cold Spring Harbor
Lazcano A, Fox GE, Oró J (1992) Life before DNA: the origin and early evolution of early Archean cells. In: Mortlock RP (ed) The evolution of metabolic function. CRC, Boca Raton, pp 237–295
Miller SL (1953) A production of amino acids under possible primitive earth conditions. Science 117:528–529
Miller SL (1955) Production of some organic compounds under possible primitive Earth conditions. J Am Chem Soc 77:2351–2361
Monod J (1971) Chance and necessity. Fontana Books, Glasgow
Muller HJ (1922) Variation due to change in the individual gene. Am Nat 56:32–50
Muller HJ (1926) The gene as the basis of life. In: Proceedings of the 1st International Congress of Plant Sciences, Ithaca, pp 897–921
Nielsen P (1993) Peptide nucleic acid (PNA): a model structure for the primordial genetic material? Orig Life Evol Biosph 23:323–327
Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289:920–930
Oparin AI (1938) The origin of life. McMillan, New York
Oparin AI (1967) Proiskhozhedenie Zhizni. Mosckovskii Rabochii, Moscow. English edition: Oparin AI (1924) The origin of life (trans: Bernal JD). Weidenfeld and Nicolson, London
Orgel LE (1968) Evolution of the genetic apparatus. J Mol Biol 38:381–392
Orgel LE, Sulston J (1971) Polynucleotide replication and the origin of life. In: Kimball AP, Oró J (eds) Prebiotic and biochemical evolution. North-Holland, Amsterdam
Oró J (1960) Synthesis of adenine from ammonium cyanide. Biochem Biophys Res Commun 2:407–412
Oró J (1961) Comets and the formation of biochemical compounds on the primitive Earth. Nature 190:442–443

Oró J, Kimball AP (1961) Synthesis of purines under primitive Earth conditions. I. Adenine from hydrogen cyanide. Arch Biochem Biophys 94:221–227
Oró J, Lazcano A (1997) Comets and the origin and evolution of life. In: Thomas PJ, Chyba CF, McKay CP (eds) Comets and the origin and evolution of life. Springer, New York, pp 3–27
Peretó J, Bada JL, Lazcano A (2009) Charles Darwin and the origins of life. Orig Life Evol Biosph 39:395–406
Pirie NC (1959) Session IV: discussion. In: Oparin AI, Pasynskii AG, Braunshtein AE, Pavloskaya TE (eds) The origin of life on Earth. Pergamon Press, New York, pp 370–371
Pontecorvo G (1982) Who was H. J. Muller (1890-1967)? Nature 298:203–204
Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated ribonucleotides in prebiotically plausible conditions. Nature 459:239–242
Ravin AW (1977) The gene as catalysts: the gene as organism. In: Coleman W, Limoges C (eds) Stud Hist Biol 1:1–45
Rich A (1962) On the problem of evolution and biochemical information transfer. In: Kasha M, Pullman B (eds) Horizons in biochemistry. Academic Press, New York, pp 41–51
Robertson MP, Miller SL (1995) An efficient prebiotic synthesis of cytosine and uracil. Nature 375:772–774
Schrum JP, Zhu TF, Szostak JW (2010) The origins of cellular life. In: Deamer DW, Szostak J (eds) Cold spring harbor perspectives in biology: the origins of life. Cold Spring Harbor Press, Cold Spring Harbor, pp 1–15
Troland LT (1917) Biological enigmas and the theory of enzyme action. Am Nat 51:321–350
Urey HC (1952) The planets: their origin and development. University of Chicago Press, Chicago
Wächtershäuser G (1988) Before enzymes and templates, theory of surface metabolism. Microbiol Rev 52:452–484
White HB (1976) Coenzymes as fossils of an earlier metabolic state. J Mol Evol 7:101–104
Woese CR (1967) The genetic code: the molecular basis for gene expression. Harper and Row, New York
Yarus M (2010) Life from an RNA world. Harvard University Press, Cambridge

Chapter 4
A Vestige of an RNA Apparatus With Ribozyme Capabilities Embedded and Functions Within the Modern Ribosome

Ada Yonath

Introduction

Peptide bonds can be formed spontaneously from activated amino acids. However, the efficiency of this spontaneous reaction is rather low, and by its nature, it lacks specificity. By providing the appropriate framework, the ribosomes increase the pace and efficiency of the peptide bond formation reactions as well as its processivity and ensure its fidelity. Ribosomes of all living cells are the universal multicomponent macromolecular assemblies that decode the genetic information and elongate the nascent polypeptide chains under mild as well as extreme conditions. The contemporary ribosomes are ribonucleoprotein assemblies, composed of two subunits of unequal size. A ratio of 2:1 for the ribosomal RNA (rRNA) versus the ribosomal proteins (r-proteins) is maintained throughout evolution, with the exception of the mammalian mitochondrial ribosomes where almost half of the bacterial rRNA is replaced by r-proteins. In prokaryotes, the small subunit (denoted as 30S according to its sedimentation coefficient) contains an RNA chain (called 16S RNA) of about 1,600 nucleotides and 20–21 different proteins (called r-proteins S1, . . . , S21), whereas the large subunit (called 50S in prokaryotes) has two RNA chains (23S and 5S RNA) of about 3,000 nucleotides in total and 31–35 different proteins (called L1, . . . , L35). The molecular weights of different ribosomes range from 2.5 MDa in prokaryotes to 4 MDa in eukaryotes. The main active sites of all ribosomes, namely the decoding site and the peptidyl transferase center (PTC), where peptide bonds are being formed, are composed solely of rRNA. The r-proteins are either supporting the various functional aspects of protein biosynthesis or stabilizing the intricate RNA fold. Despite the significant difference in the sizes of the ribosomes, their functional sites exhibit remarkable conservation in all kingdoms of life. Consistently, the modes of ribosomes operation, as well as the chemical properties, their substrates, and products, appear to be universal.

A. Yonath (✉)
Department of Structural Biology, Weizmann Institute, Rehovot, Israel
e-mail: ada.yonath@weizmann.ac.il

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_4, © Springer Science+Business Media New York 2013

The amino acids bound covalently to the 3′ ends of tRNA molecules are brought to the ribosomes. The three-dimensional structures and the two-dimensional diagrams of all tRNA molecules from all living cells across evolution are alike, although each of them is specific to its amino acid. They are all built mainly of double helical L-shape molecules, made by a stem-elbow-stem (SES) organization, and contain, on one of its ends, the anticodon loop that matches with the complementary three-nucleotide code on the mRNA. About 70 Å away, at their 3′ ends, the tRNA molecules are built of a single strand with the universal sequence of CCA, to which the cognate amino acid is bound by an ester bond. The tRNA molecules are the nonribosomal entities combining the two subunits, as each of their three binding sites, A (aminoacyl), P (peptidyl), and E (exit), resides in both subunits. At the A-sites and P-sites, the tRNA anticodon loops interact with the mRNA on the small subunit, whereas the peptide bonds between the amino acid bound to the A-site tRNA and the peptidyl bound to the P-site tRNA are being formed in the large subunit.

Peptide bonds are formed in the PTC and possess striking architecture. Peptide bond formation is the only chemical reaction occurring within the ribosome. It resides within the universal symmetrical "pocket-like" structural element (Fig. 4.1), an extraordinary feature in the otherwise asymmetric ribosome. This region is composed of highly conserved 180 nucleotides of the rRNA (Bashan et al. 2003; Baram and Yonath 2005; Agmon et al. 2005; Yonath 2009) and provides the framework for the accurate positioning of the ribosomal substrates in the stereochemistry required for peptide bond formation; for substrate-mediated catalysis (Gregory et al. 2004; Bieling et al. 2006; Weinger and Strobel 2006) and for coffining the void consumed by the motions involved in tRNA 3′ end translocation required for the successive peptide bond formations, thus enabling the ribosome polymerase activity (Bashan et al. 2003; Agmon et al. 2005; Sato et al. 2006; Gindulyte et al. 2006). Here, we show that by exploiting structural, biochemical, computational, and comprehensive mutagenesis experiments, the remnant of this machine was identified within the contemporary ribosome.

An Ancient Molecular Machine Is Imbedded in the Modern Ribosome

The overall fold of each half of the ribosomal symmetrical region resembles a common building block of "ancient" as well as "modern" functional RNA molecules of comparable size. These include gene regulators, riboswitches, ribozymes catalyzing phosphodiester cleavage, RNA processors, etc., thus suggesting that an entity resembling the current ribosomal symmetrical region could have existed in the RNA world as a self-folded autonomous entity. This entity could have functioned as an apparatus catalyzing various reactions involved in RNA metabolism, which consequently evolved into a relatively simple machine, capable of catalysis of various chemical reactions, including the formation of peptide bonds and noncoded oligopeptides elongation.

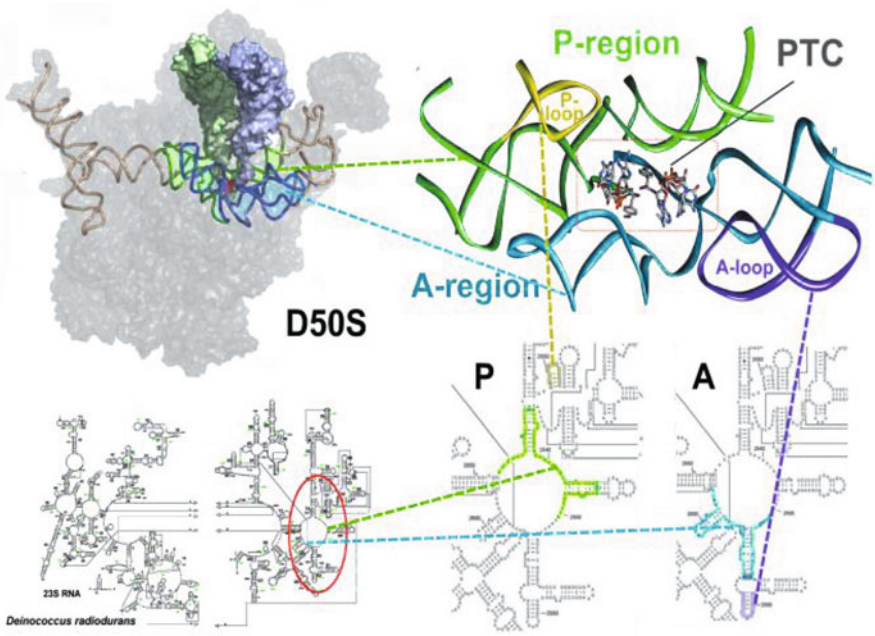


Fig. 4.1 Attempt at construction of the proto-ribosome. In all panels, the two halves of the symmetrical region are colored in *blue* (the A-subregion) and *green* (the P-subregion). *Top left* The entire large ribosomal subunit from *Deinococcus radiodurans* (D50S) (Harms et al. 2001) and the symmetrical region (Bashan et al. 2003) at the heart of it, showing the entire volume occupied by each of the two halves and their RNA extensions. The A-site and P-site tRNA molecules are shown in *blue* and *green*, respectively. The approximate site for peptide bond formation is marked in *red*. *Top right* Zoom on the symmetrical region, showing the RNA fold of its two subregions. The A-loops and P-loops, which create base pairs with C74 of the tRNA 3′ CCA ends of the A-site and the P-site tRNA molecules, are depicted. The location of the two substrates, as found in the modern ribosome (Bashan et al. 2003), is shown within the *red dotted square* in the middle of the symmetrical region. *Bottom left* The two-dimensional representation of the 23S rRNA chain of D50S. The symmetrical region is encapsulated in a *red circle*. The A-loop is shown in *purple* and the P-loop is *olive green*. *Bottom right* The RNA compositions of the A (*blue*) and P (*green*) constructs. The SES conformation of both constructs is evident in the two-dimensional representation (*bottom left*). Each construct was designed on the basis of the features of the symmetrical region

Support for the existence of RNA entities capable of self-replication, folding, and dimerization comes from the recent nonenzymatic synthesis of activated pyrimidine ribonucleotides under prebiotically plausible conditions (Powner et al. 2009; Szostak 2009) and *the* demonstration that RNA oligomers can be obtained nonenzymatically from activated RNA precursors (Pino et al. 2008; Krzyaniak et al. 1994). The dimerization in a symmetrical manner of self-folded motifs of identical, similar, or different sequences, may have occurred spontaneously, or could have been driven by chemical requirements of the RNA world, resulting in RNA machines of "pocket-like" structures. It is conceivable that these machines were capable of hosting various

RNA substrates, which could have been replaced by amino acids conjugated with single or short oligonucleotides, (Ilangasekare et al. 1995; Giel-Pietraszuk et al. 2006; Lehmann et al. 2007; Turk et al. 2010) serving as substrates for peptide bond formation.

There is no evidence for any currently living system that possess ribosomes that are more "primitive" that those of bacteria. Furthermore, the smallest ribosomes existing today, those from bacterial source, are highly complex. This may indicate that because the translation of a genetic code to create coded polypeptides requires significant structural sophistication and once the production of chemically efficient molecules became attainable, there was no reason to maintain semiadvanced apparatuses. Hence, no intermediate protein synthesis machine, which could have evolved during the transition from the ancient RNA world toward the current protein-based life, was preserved.

It appears that currently there are no direct observations for following the transition from the prebiotic era, which is presumed to be dominated by RNA, to the modern protein/DNA/RNA based world. To circumvent our deficiency in possessing direct evidence of this and following the steps in evolution, we searched for structural elements that can be remnants of previous molecular machines. The results of these efforts indicate that the transition from the RNA-only world hinged on the emergence of a mechanism for the production of noncoded polypeptides. Various noncoded polypeptides could have appeared as an ensemble of short oligopeptides; among them, some may have been somewhat functionally meaningful in simple catalysis, in metal transport, or as stabilizers of the machine that created them. These could have been the driving force for the emergence of the genetic codes.

Evolution Pressure May Have Triggered RNA to Produce Proteins

The previously suggested sequence of events stimulated us to pose an intriguing question: what could be the "motivation" (or driving force) of the RNA world to create peptide bonds, thus producing proteins that can perform chemical reactions more efficiently than their creators? A feasible, although less common, answer to this question is based on the suggestion that once amino acids appeared in the RNA world, the existing RNA chains or primitive "machines" operated as ribozymes for binding the "invading" amino acids to themselves (Turk et al. 2010; Pino et al. 2008; Lehmann et al. 2007; Ilangasekare et al. 1995), presumably for controlling and/or for eradicating the invaders. As the mere binding to the RNA chains activated the amino acids, they could have bound to each other, namely participate in peptide bond formation. This was in accordance with the hypotheses that the early forms of translation could occur under less specific astrochemical requirements and that the early ribozymatic activities were performed by very small RNA enzymes.

Based on the high conservation of the ribosomal symmetrical region, which we assigned as the proto-ribosome, as well as on the finding that this region in the

modern ribosome provides all the architectural elements required for amino acid oligomerization, we assumed that it may represent a moiety that existed in the RNA world and functioned in a fashion similar to its precedent within the contemporary ribosome. In other words, it is conceivable that the ancient proto-ribosome in its functionally optimized version is still embedded in the core of the modern ribosome and that the symmetrical region of the modern ribosome originated from the proto-ribosome (Agmon et al. 2006, 2009; Davidovich et al. 2009; Belousoff et al. 2010).

Simple Machinery Within the Complex Ribosome

The assumed sequence of events is in accordance with the discovery of the symmetrical region within the contemporary ribosome and the suggestion that it is the vestige of an ancient RNA apparatus, capable of making chemical bonds. The hypothesis of a self-assembled ribosomal active site, which is still implanted in the internal core of the modern ribosome, was assessed by biochemical experiments and structural methods (Davidovich et al. 2009; Belousoff et al. 2010). In these experiments, we aimed at scrutinizing an ancient self-folded RNA entity that could have functioned in chemical reactions required by the RNA world. For the construction of experimental systems that can resemble the proto-ribosome, we have designed several RNA constructs belonging to two major groups: one based on sequences found in contemporary ribosomes (called here as "native" and "mutated" versions) and the second is composed of semirandom sequences. In both systems, the RNA chains are supposed to assume the SES fold, which is a common RNA fold in modern as well as believed-to-be-ancient molecules (Fig. 4.1), including the ribosomal symmetrical region.

In these experiments, we first aimed at revealing the tendency for self-folding and dimerization of RNA chains, which yielded preliminary evidence supporting the existence of a dimeric proto-ribosome, and provided hints for a feasible pathway for acquiring the structural elements necessary for pocket-like formation. Consistently, it is conceivable that "pocket-like" RNA entities were assembled spontaneously from a pool of RNA chains. We therefore proposed the proto-ribosome conception as the link between the RNA world and modern life.

Unexpected Findings and Their Possible Interpretation

Rather unexpectedly, we found that some, albeit not all, RNA chains with sequences resembling those observed in the current ribosome are capable of forming dimers that may adopt a "pocket-like" structure (Davidovich et al. 2009). For example, dimer formation was observed for the RNA chains that were designed using the sequence of one side of the symmetrical region (called P, Fig. 4.1). Furthermore, by a procedure mimicking site-directed mutagenesis, we showed that the tendency for

dimerization, a prerequisite for obtaining the catalytic center, is linked to the fold of the two components, thus indicating that functional selection at the molecular level already existed in the prebiotic era.

Marked preference of selected sequences to dimerize was observed, although all of the constructs are quite similar, albeit not identical. This nonuniform tendency to dimerize may indicate that survival of the fittest and natural selection seem to play a major role in the prebiotic world, although these properties are commonly related to the evolution of species. In addition, the finding that the P-subregion dimerizes but the A-subregion does not may indicate that initially the proto-ribosome was a dimer of two halves, each of which resembled the P-subregion. Such a pocket could have been created by gene duplication or gene replication, where the half-pocket RNA construct was used as a template for its own replication (Eigen 1993; Lincoln and Joyce 2009; Woese 2001; Yarus 2002).

Further Evolution Steps

It is likely that the more stable constructs of the "pocket-like" molecular dimers may have survived under various environmental conditions. Once some of the randomly produced polypeptides were found to be useful in the RNA world, those RNA pockets that could accommodate suitable substrates at the appropriate stereochemistry, enabling peptide bond formation, could have been evolutionarily favored. As it is assumed that in the prebiotic era, RNA chains could self-replicate (Eigen 1993; Lincoln and Joyce 2009; Woese 2001; Yarus 2002), it is conceivable that phenotypes with favorable properties could have been synthesized in many copies. It is likely that some of these phenotypes were originated by fusion of two different or duplication of two identical sequences, resembling gene elongation events (Fani and Fondi 2009).

Among the products of these early amino acid elongation processes, those molecular entities possessing central, albeit primitive, catalytic and/or synthetic properties became the templates for enhanced production (Belousoff et al. 2010), survived evolution pressures, and underwent natural selection. Among the key tasks performed by the initial oligopeptides is stabilizing the proto-ribosome and/or other components confined in its surrounding, within assemblies that could evolve into "proto cells." Subsequently, the proto-ribosomes underwent optimization from nongenetic peptide bond formation toward performing genetically driven translation.

Based on these findings, as well as on an analysis of the inter-RNA interactions within the ribosome events (Bokov and Steinberg 2009), we suggest that regardless of the sequence of the initial construct, the proto-ribosome has later evolved into the multicomponent molecular machine, the ribosome. The following pathways could be involved: (i) optimization of the pocket for efficient peptide bond formation, so that it can accommodate two similar, albeit not identical, substrates in a stereochemistry suitable for this reaction; (ii) gradual addition of chemical elements that stabilize the proto-ribosome machine, such as RNA chains and/or noncoded oligopeptides or polypeptides; (iii) coevolvement of the RNA apparatus to more sophisticated

machine with its substrates; example can be the A-loops and P-loops (Fig. 4.1) that, in the modern ribosome, interact with C74 of the 3′ end of each of the A-site and the P-site tRNA molecules, i.e., the cytosine located furthest away from the components involved in peptide bond formation, namely the amino acid and growing peptide chain; (iv) the appearance of an entity not yet characterized that can represent a proto-genetic code.

Conclusions

The high-resolution structures of ribosomal particles facilitated advance research in ribosome function and in its possible ancestor. Analysis of the presumed early steps in peptide bond formation indicated that the ribosome is a naturally occurring ribozyme that outlived the transition from the prebiotic RNA world to the contemporary protein nucleic acid life. Furthermore, studies on the suggested extension from the early primitive synthesis processes to the modern version of protein biosynthesis shed light on intrinsic structural requirements and enabled visualization of a feasible path from the primordial world to contemporary genetic code translation. These studies also indicated that functional selection at the molecular level existed already in the prebiotic era, thus suggesting the Darwinian selection could have played a major role at the molecular level, even before life began. Altogether, our structural analysis proposes that a spontaneously created self-folded dimeric RNA apparatus could have evolved into a proto-ribosome, an event that is likely to mark the early steps in the transition from the RNA world into the present form of life.

Acknowledgments We thank Ilana Agmon and all members and guests of the ribosome group at the Weizmann Institute for continuous interest, fruitful discussions, experimental design, and excellent technical assistance. Support was provided by the US National Institute of Health (GM34360) and the Kimmelman Center for Macromolecular Assemblies. AY holds the Martin and Helen Kimmel Professorial Chair.

References

Agmon I, Bashan A, Zarivach R, Yonath A (2005) Symmetry at the active site of the ribosome: structural and functional implications. Biol Chem 386:833–844

Agmon I, Bashan A, Yonath A (2006) On ribosome conservation and evolution. Israel J Ecol Evol 52:359–374

Agmon I, Davidovich C, Bashan A, Yonath A (2009) Identification of the prebiotic translation apparatus within the contemporary ribosome. http://precedings.nature.com/documents/2921/version/1

Baram D, Yonath A (2005) From peptide-bond formation to cotranslational folding: dynamic, regulatory and evolutionary aspects. FEBS Lett 579:948–954

Bashan A, Agmon I, Zarivach R, Schluenzen F, Harms J, Berisio R, Bartels H, Franceschi F, Auerbach T, Hansen HAS, et al (2003) Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. Mol Cell 11:91–102

Bieling P, Beringer M, Adio S, Rodnina MV (2006) Peptide bond formation does not involve acid-base catalysis by ribosomal residues. Nat Struct Mol Biol 13:423–428
Belousoff MJ, Tjioe L, Graham B, Spiccia L (2010) Synthesis, X-Ray crystal structures, and phosphate ester cleavage properties of bis(2-Pyridylmethyl)amine copper(II) complexes with guanidinium pendant groups. Inorg Chem 47(19):8641–8651
Bokov K, Steinberg SV (2009) A hierarchial model for evolution of 23S ribsomal RNA. Nature 457:977–980
Davidovich C, Belousoff M, Bashan A, Yonath A (2009) The evolving ribosome: from non-coded peptide bond formation to sophisticated translation machinery. Res Microbiol 160:487–492
Eigen M (1993) The origin of genetic information: viruses as models. Gene 135:37–47
Fani R, Fondia M (2009) Origin and evolution of metabolic pathways. Phys Life Rev 6:23–52
Giel-Pietraszuk M, Barciszewski J (2006) Charging of tRNA with non-natural amino acids at high pressure. FEBS J 273:3014–3023
Gindulyte A, Bashan A, Agmon I, Massa L, Yonath A, Karle J (2006) The transition state for formation of the peptide bond in the ribosome. Proc Natl Acad Sci U S A 103:13327–13332
Gregory ST, Dahlberg AE (2004) Peptide bond formation is all about proximity. Nat Struct Mol Biol 11:586–587
Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. Cell 107:679–688
Illangasekare M, Sanchez G, Nickles T, Yarus M (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. Science 267:643–647
Krzyzaniak A, Barciszewski J, Salanski P, Jurczak J (1994) The non-enzymatic specific amino-acylation of transfer RNA at high pressure. Int J Biol Macromol 16:153–158
Lehmann J, Reichel A, Buguin A, Libchaber A (2007) Efficiency of a self-aminoacylating ribozyme: effect of the length and base-composition of its 3′ extension. RNA 13:1191–1197
Lincoln TA, Joyce GF (2009) Self-sustained replication of an RNA enzyme. Science 323:1229–1232
Pino S, Ciciriello F, Costanzo G, Di Mauro E (2008) Nonenzymatic RNA ligation in water. J Biol Chem 283:36494–36503
Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459:239–242
Sato NS, Hirabayashi N, Agmon I, Yonath A, Suzuki T (2006) Comprehensive genetic selection revealed bases essential for protein synthesis in the peptidyl-transferase center. Proc Natl Acad Sci U S A 193:15386–15391
Szostak JW (2009) Origins of life: systems chemistry on early earth. Nature 459:171–172
Turk RM, Chumachenko NV, Yarus M (2010) Multiple translational products from a five-nucleotide ribozyme. Proc Natl Acad Sci U S A 107:4585–4589
Weinger JS, Strobel SA (2006) Participation of the tRNA A76 hydroxyl groups throughout translation. Biochemistry 45:5939–5948
Woese CR (2001) Translation: in retrospect and prospect. RNA 7:1055–1067
Yarus M (2002) Primordial genetics: phenotype of the ribocyte. Annu Rev Genet 36:125–151
Yonath A (2009) Large facilities and the evolving ribosome, the cellular machine for genetic-code translation. J R Soc Interface 6:575–585

Chapter 5
Covering All the Bases: The Promise of Genome-Wide Sequence Data for Large Population Samples of Bacteria

Santiago Castillo-Ramírez and Edward J. Feil

Introduction and Background

Molecular data comes in waves. The first nucleotide sequence-based population studies on bacteria were published in the late 1980s and early 1990s and typically focused on one or two gene sequences from around a dozen bacterial isolates (Nelson et al. 1991; Nelson and Selander 1992, 1994; Boyd et al. 1994; DuBose et al. 1988; Milkman and Bridges; 1990). Despite the modest size of these datasets by contemporary standards, these early works laid the foundation for a number of inter-related debates: (i) the extent to which bacterial evolution is tree-like or network-like, (ii) the search for a meaningful 'species' concept, (iii) biogeography and the extent to which 'everything is everywhere' and (iv) the roles of selection (ecological adaptation) and neutrality (drift) in shaping population structure. More specifically, data generated during the late 1980s and early 1990s provided unequivocal evidence that at least some species experience high rates of genome-wide homologous recombination. These early data also brought into sharp relief the importance (and difficulties) in assembling representative population samples in bacteria (Smith et al. 1993).

The generation of complete genome sequences for major human pathogens from 1995 onwards (Fleischmann et al. 1995) was of immediate practical benefit for population biology and epidemiological studies, in that it enabled primer design for practically any gene. The additional key advances of capillary-based automated sequencers and the Internet provided the means not only to sequence several gene loci for large population samples (100s of isolates) but also to instantly access, compare, and interrogate these data from anywhere in the world. The development of multi-locus sequence typing (MLST) in 1998 aimed to exploit these advances for pathogenic bacteria (Maiden et al. 1998; Maiden 2006). Although clear sampling biases have meant that the dual questions of epidemiological surveillance and population biology have, at times, been uncomfortable bedfellows (Maiden 2006), these datasets have significantly impacted on both fronts.

S. Castillo-Ramírez · E. J. Feil (✉)
Department of Biology and Biochemistry, University of Bath, Bath, UK
e-mail: E.Feil@bath.ac.uk

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_5, © Springer Science+Business Media New York 2013

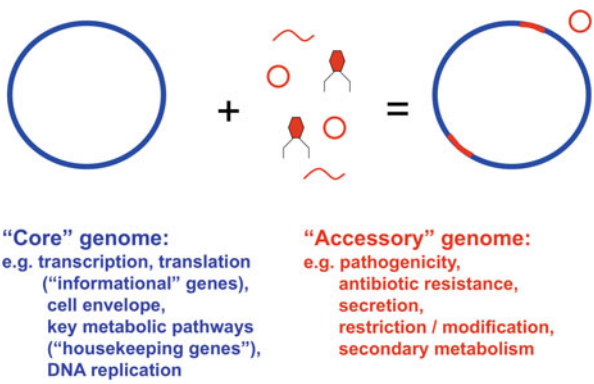


Fig. 5.1 The duality of the bacterial genome into core and accessory genes**.** The *red* 'accessory' genes are acquired by horizontal transfer, a process that usually involves a vector such as a plasmid or phage. Such genes can help the recipient bacterium to exploit new niches; thus, this model supports a saltational view of bacterial evolution

Microarray technology represented another important advance. Combined with the rapidly increasing genomics datasets, this approach confirmed a fundamental duality within bacterial chromosomes: the stable 'core genome' versus the dynamic 'accessory genome' (Feil 2004). For any given taxon (e.g. named species), the core genome is defined simply as all those genes universally present in all strains. In contrast, the accessory genome is composed of genes variably present or absent. Microarray data revealed that the accessory genome is not randomly interspersed throughout the chromosome, but is largely confined to regions of high gene content variability, and that the locations of these variable regions are the same in all strains (Alm and Trust 1999; Dorrell et al. 2001; Fitzgerald et al. 2001). Complete genome sequencing revealed a second key feature of the accessory genome, that of atypical base composition (GC content), GC skew and codon bias, which implicates extra-chromosomal origins and the role of mobile genetic elements (Ochman et al. 2000). A useful analogy is to consider the core genome as the cell's 'operating system' (encoding housekeeping metabolism and DNA processing) and the accessory genome as software modules providing specialised functionality (Fig. 5.1). Thus, the accessory genome can play a major phenotypic role (for example, by conferring antibiotic resistance or virulence traits) (Holden et al. 2004a), but the rapidity by which such genes can be gained or lost may make them poor markers for reconstructing evolutionary history (Turner and Feil 2007).

Molecular typing protocols such as MLST represent pragmatic trade-offs between inter-strain discrimination, the amenability of the data to evolutionary analyses, and the time and expense required to characterise large population samples. The best solution depends on both the taxon being considered and the aims of the study; there is no universally 'correct' typing method. For MLST, a small sample of core genes (typically seven) are sequenced for each isolate, and each strain is defined on the basis of the combined alleles observed over all loci. Freely available databases house MLST data for 1,000s of isolates for many major pathogens (Chan et al. 2001). Although this approach has been very successful for large-scale epidemiological surveillance and for understanding migration rates and evolutionary dynamics, it has two serious drawbacks. First, as MLST is based exclusively on a small sample of

core genes, it is blind to large-scale genome rearrangements and changes in the accessory genome. A striking example of this is provided by a comparison between two related species: *Burkholderia pseudomallei* (an opportunistic pathogen of humans) and *B. mallei* (a pathogen of horses). Although phylogenetic analysis based on the MLST data suggested that *B. mallei* represents a specialised 'clone' nested within the *B. pseudomallei* population (Gevers et al. 2005), comparative genomics revealed that the former had undergone extensive genome degradation (the deletion of around 1 Mb of DNA) and revealed a complete loss of synteny with *B. pseudomallei* (Holden et al. 2004b). These dramatic changes reveal much about the ecological and molecular dynamics underpinning the emergence of *B. mallei*, but would have been completely missed on the basis of MLST data alone.

The second drawback of MLST is that a small sample of seven (or so) core genes provides insufficient discrimination for highly monomorphic species (recently reviewed in Achtman 2008) or, similarly, for single clonal complexes within more diverse species. Cases where a single 'type' predominates on a local scale present a pertinent problem for epidemiological surveillance, as the majority of isolates recovered are indistinguishable by standard procedures. Alternative techniques based on hyper-variable repeat regions have been established for some monomorphic species such as *Mycobacterium* sp.; it is not always clear to what extent the patterns of divergence in the repeat loci are homoplasic or conflict in some other way with the rare single-nucleotide polymorphisms (SNPs) present elsewhere in the genome (Smith et al. 2003, 2006).

Full genome sequencing will of course provide ultimate discriminatory power, but until recently, it has not been possible to generate genome-wide datasets for large population samples. The latest advances in next-generation sequencing technology are set to close this technology gap. Platforms such as the Illumina Genome Analyser (IGA), Roche 454 and SOLiD provide the means to identify genome-wide SNPs for large population samples. Further, we discuss how this technology has been recently applied to the important pathogen *Staphylococcus aureus* and how the high discriminatory power of the data can be used in an epidemiological setting. We will go on to emphasise that next-generation sequence datasets represent far more than just MLST datasets writ large and how these data may also inform on more fundamental evolutionary questions concerning the dynamic interplay between mutation, recombination, drift and selection over very short evolutionary time scales.

The Population Structure of *Staphylococcus aureus*

S. aureus is a low-GC Gram-positive bacterium, which asymptomatically colonises the skin and anterior nares (nostrils) of approximately one-third of the human population at any given point in time (Peacock et al. 2001; Nulens et al. 2005; van Belkum 2006). The species is also commonly recovered from domesticated animals, particularly cows, pigs and chickens (Vanderhaeghen et al. 2010), though the true ecological range in wild animals and the environment is not known. Infection by *S. aureus* can

cause a number of conditions ranging in severity from boils to life-threatening endocarditis. Serious infections are much more common within health care settings, such as hospitals and nursing homes, where disease management is substantially hampered by the spread of resistance to β-lactam antibiotics (initially penicillin-resistant, then methicillin-resistant *S. aureus*; MRSA). Resistance to methicillin is conferred via the horizontal acquisition of a large (20–60 Kb) chromosomal cassette (*SCC*mec), which is thought to have been introduced from naturally resistant commensal staphylococcal species on multiple occasions (Enright et al. 2002). In recent years, sporadic infection by MRSA has become more common in the community, outside of health care settings (Deleo et al. 2010).

Several typing methods have been used for epidemiological surveillance of this species, including pulse-field gel electrophoresis (PFGE), MLST (Cookson et al. 2007) and MLVA (multiple loci VNTR analysis, where VNTR stands for variable number of tandem repeats) (Melles et al. 2009; Schouls et al. 2009). VNTR loci are hyper-variable microsatellites; the most notable example in *S. aureus* being the *spa* gene, which has been used extensively for typing studies in this species (Basset et al. 2009; Mellmann et al. 2008). Although these methods present a range of utility for more detailed evolutionary analyses, they (more or less) consistently delimit the *S. aureus* population into the same discrete clusters or clonal complexes. These can be visualised using the unweighted pair group method with arithmetic mean (UPGMA) dendrograms or a simple clustering algorithm based upon related sequence types (BURST), implemented as the freely available eBURST (Feil et al. 2004) (Fig. 5.2). As it is based on nucleotide sequence data, MLST data is also amenable to phylogenetic analysis. These analyses have also revealed that hospital-acquired (HA)-MRSA isolates are particularly clonal, with a very small number of clusters accounting for almost all the cases of infection worldwide (Aires de Sousa et al. 2005; Conceicao et al. 2007; Crisostomo et al. 2001; Gomes et al. 2006). Methicillin-sensitive *S. aureus* (MSSA) are more diverse, as are community-acquired (CA)-MRSA isolates. This had led to the view that HA-MRSA strains, which are more likely to develop resistance than CA-MRSA strains and are relatively rare outside of health care settings, are specifically adapted to the hospital environment.

S. aureus experiences homologous recombination at relatively low frequency. This is thought to explain in part why the discrete clusters have emerged and been maintained (Feil et al. 2003) and why it is possible to reconstruct reasonably robust phylogenies (Cooper and Feil 2006). However, although there is some evidence for phage specificity in different clusters (Waldron and Lindsay 2006), there is very little evidence that they represent even partial barriers to gene flow; hence, they probably do not equate to biological species. Discounting the recently emerged (and massively over-sampled) HA-MRSA sub-groups, the adaptive relevance of clusters within the broader population remains unclear. Nevertheless, the consistent delineation of the same clonal complexes, even from gene content data generated using microarrays, underline that they are real biological entities and as such are both of evolutionary interest and of epidemiological utility (Turner and Feil 2007).

Taking the species as a whole, MLST, MLVA and PFGE all do an excellent job in assigning isolates to one or other of these groups and in revealing the changing

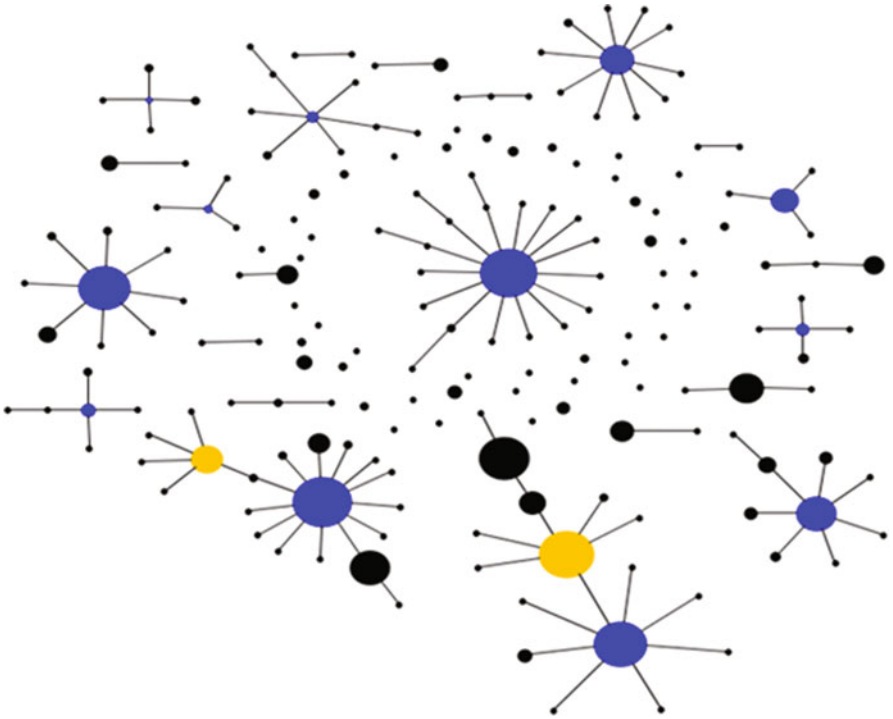


Fig. 5.2 Clonal population structure of *S. aureus* visualised using MLST/eBURST (www.mlst.net/eburst.mlst.net). Each circle represents a unique MLST haplotype. The size of the circle represents the frequency of the haplotype. Haplotypes differing by only one of the seven loci are linked. The vast majority of isolates correspond to a small number of haplotypes and clusters of haplotypes. *Blue* haplotypes are assigned as the most parsimonious founder of each cluster and *yellow* haplotypes as sub-founders, which have experienced subsequent diversification. Distances, angles and positions have no meaning

frequencies of the clusters over time and space (over years, even decades, and on both national and international scales). Typically, a very small number of clones tend to predominate at a given location at any given point in time, a result of sequential waves of infection (De Lencastre et al. 2007). This epidemiological pattern lessens the utility of MLST and other typing schemes to track strain transmission patterns on local scales (within and between hospitals) over short time scales (weeks and months) because strains belonging to a single clonal group are often very difficult to distinguish. However, the typing data clearly reveals minor variation within clonal complexes, usually manifesting as single point mutation on the basis of MLST, occasionally different *spa* types, and minor banding pattern changes by PFGE. Thus, more powerful techniques should be able to dissect out sufficient variation within clonal complexes for tailoring specialised typing protocols. Such data would also provide clues as to the emergence of these clusters to understand evolutionary processes in bacteria over very short time periods.

Zooming in on Single *S. aureus* Clones

ST5

Initial studies aimed at increasing the resolution of MLST by supplementing the data with hyper-variable loci (Robinson et al. 2005; Kuhn et al. 2006) provided evidence in support of low rates of recombination and robust phylogenetic trees, but shed little extra light on the general patterns of diversity within different complexes. A leap forward was taken by Nubel et al. (2008), who used a mutation discovery procedure based on denaturing high-performance liquid chromatography (dHPLC) to identify SNPs in ~45.5 Kb (1.6 % of the genome) in 135 *S. aureus* isolates from 22 countries. All the isolates were identical by MLST (corresponding to the haplotype called ST5). The authors detected sufficient variation to reveal striking geographical structuring, indicating the existence of local variants and raising the possibility of reconstructing transmission chains over short-distances and time-scales. In fact, it is clear that the degree of geographical structuring will be positively correlated with the discriminatory power of the data, as migration rates will always be much slower than local clonal expansion.

The data of Nubel et al. also pointed to multiple independent acquisitions of the *SCC*mec element responsible for methicillin resistance, even within ST5. This argued against the prevailing model of the rapid global dissemination of a few MRSA strains in favour of a model of frequent emergence of 'home-grown' locally restricted MRSA clones, some of which happen to belong to the same 'strain' (meaning MLST-defined haplotype). Although this raises questions concerning the mobility of the *SCC*mec element and its frequency in staphylococcal reservoirs, the clear practical implication is that genome-wide SNP data should resolve cohesive sub-types within ST5, each corresponding to independent acquisitions of *SCC*mec, and which could potentially be easily distinguished by tailor-made typing protocols. A more recent paper using similar methodology focussed on one such sub-cluster within the broader ST5 group, called ST225 (Nubel et al. 2010). The universal presence of a defining deletion within *SCC*mec showed that this cluster corresponds to a single acquisition of this element, and the authors were able to posit that this sub-group was introduced into central Europe from the USA approximately a decade ago, and has subsequently spread rapidly between hospitals.

ST239

One of the earliest HA-MRSA clones, and currently the most dominant globally, is called ST239. This haplotype evolved via an homologous recombination event between unrelated *S. aureus* strains (i.e. belonging to different clonal complexes), which resulted in the replacement of ~20 % of the recipient genome (> 550 Kb) (Robinson and Enright 2004). This event remains the largest single homologous

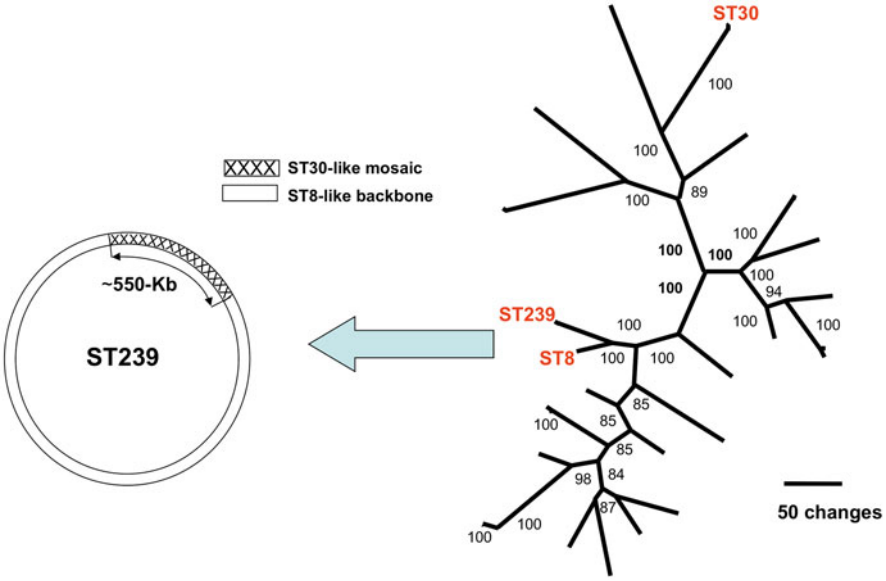


Fig. 5.3 A representative phylogeny of *S. aureus* generated using concatenated data from 40 gene loci (17.8 Kb), reconstructed using MrBayes**.** Values at the nodes are posterior probabilities (see Cooper and Feil 2006). The hybrid ST239 genome is the result of a large-scale recombination event between unrelated lineages, ST8 and ST30 (marked on the tree). Approximately 550 Kb of the ST8 genome has been replaced by the homologous region from an ST30-like donor (see Robinson and Enright 2004)

replacement described in the literature and arose only once via a completely unknown mechanism (Fig. 5.3). This makes this hybrid clade unusual in that it can be defined by a single and unambiguous marker. Any strain exhibiting this replacement can be confidently assumed to have descended from the original recipient. Further, as 'reversals' are implausible, any isolate that does not exhibit this replacement can be excluded (Feil et al. 2008). Although the occasional de novo point mutation means that not all members of this clade are identical by MLST, for ease of terminology, we will use the term ST239 to include all members of this clade as defined by the presence of the large replacement.

All ST239 isolates are resistant to methicillin (MRSA) and all isolates posses the same 'type' (variant) of *SCC*mec cassette (although minor *SCC*mec variation is detected). This again points to common ancestry and is not consistent with multiple independent acquisitions of *SCC*mec within this clade. The specific type common to all ST239 isolates is known as type III. This is the largest *SCC*mec cassette yet described in *S. aureus*, which confers multiple resistance and is restricted to ST239 isolates. ST239 is responsible for ~90 % of hospital-acquired MRSA infection throughout most of mainland Asia (from the Middle East to China) and much of South America (Diekema et al. 2001; Chongtrakool et al. 2006; Xu et al. 2009). However, similar to ST225, ST239 is almost exclusively observed within hospital

environments. This is thought to be because the large *SCC*mec cassette confers a fitness cost that renders it uncompetitive in the community. This implicates direct hospital-to-hospital transmission as playing a key role in its global dissemination.

ST239 is also unusual in that there is both epidemiological and experimental evidence pointing to increased virulence (Amaral et al. 2005; Edgeworth et al. 2007). ST239 was also the predominant MRSA clone in western Europe during the 1980s and 1990s but has subsequently been replaced by other strains (Conceicao et al. 2007). Following an outbreak in a London hospital (Edgeworth et al. 2007), a variant of ST239 called TW20 was sequenced (Holden et al. 2010). The clinical significance, widespread dissemination, availability of a reference sequence and unusual evolution and epidemiology of ST239 makes it an ideal candidate for high-resolution analysis. Smyth et al. (2010) identified all the mutations within ~15 Kb of sequence of 111 ST239 isolates representing 34 years and 29 countries. Again, these authors noted geographical structuring on a continental scale and identified European, South American and Asian sub-clades. These data also pointed to homoplasious deletions within the *SCC*mec element during diversification of the ST239 clade, but (unlike ST5) showed no evidence for multiple acquisitions of completely different *SCC*mec types.

A New Dawn for Next-Generation Sequencing

The ST239 clade was also chosen as the focus for the first population study, which utilised IGA to identify genome-wide SNPs and insertions and deletions (INDELs) compared with TW20 as a reference sequence. Harris et al. (2010) used index adapters to create individually tagged genomic libraries in order to rapidly generate whole-genome DNA sequence data for a large number of isolates. The authors characterised 62 ST239 isolates, 42 of which were globally representative, while the remaining 20 were isolated from a single hospital in northeast Thailand over a 7-month period. The reference TW20 strain was also re-sequenced as a control. The study was thus designed to address both the global diversity of this clone and the utility of next-generation sequencing for very localised epidemiology.

The study identified 6,714 high-quality SNPs, but these were not equally distributed throughout the genome. Regions of high-density SNPs were clearly identified, and these also tended to correspond to regions with relatively low coverage. These diverse regions represented the accessory genome, principally consisting of mobile genetic elements (MGEs) such as *SCC*mec, genomic islands, conjugative elements and prophage. Harris et al. defined core regions simply and conservatively as all regions of > 1 Kb, which were mapped to a high quality in all isolates. This definition provides a subtly different set of genes assigned as 'accessory' than previous microarray and comparative genomic studies on the broader *S. aureus* population. For example, any accessory elements present in the founding ST239 genome that had been stably inherited were assigned as core, whereas any small INDELs that have arisen since the emergence of ST239 will mean the corresponding region and will be assigned as accessory, even though they may be 'native' to the genome. Nevertheless,

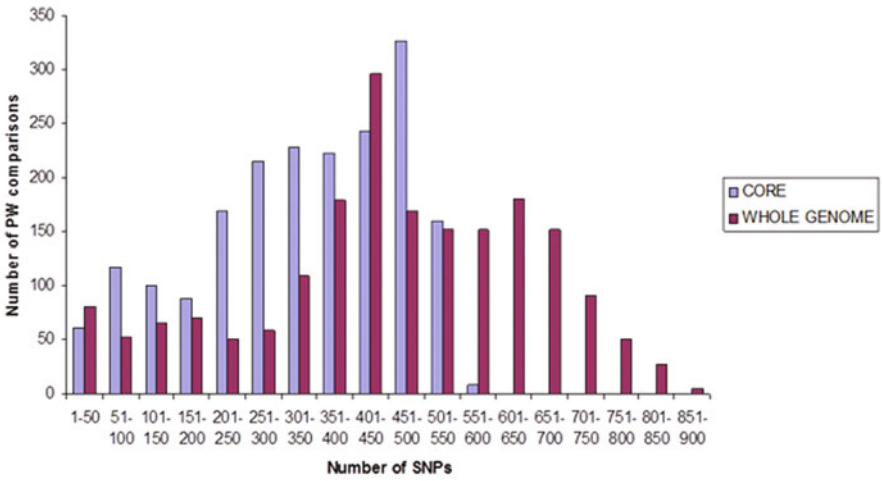


Fig. 5.4 Pairwise comparisons showing the level of diversity in the data of Harris et al. (2010). The maximum number of SNPs between any two isolates is > 850 when the whole genome is considered, but < 601 when only core SNPs are considered

because the vast majority of those SNPs thus assigned as corresponding to accessory regions are likely to have been acquired by horizontal transfer, they were excluded from the phylogenetic analysis. This resulted in the exclusion of 2,404 SNPs, leaving 4,310 variable sites in the core genome; the accessory genome therefore represents a significant fraction of the overall diversity within ST239 (Fig. 5.4).

Similar to the studies mentioned above, Harris et al. noted striking geographical structuring in the data from their tree, which was rooted using ST8, the original recipient of the large replacement (Fig. 5.5). The increased resolution provided by IGA also allowed a number of specific inferences concerning the emergence and spread of ST239. First, it is apparent that European isolates are more diverse than Asian or South American isolates. This observation points to a European origin of ST239, which is consistent with the fact that it was first recorded in this continent, but has subsequently been largely replaced. Second, isolates from South America are extremely homogenous, even more so than the strains from a single hospital in Thailand, despite the fact that they represent Brazil, Chile, Argentina and Uruguay and were recovered from 1993 to 1998. This strongly points to a single introduction of ST239 into South America followed by dramatic spread. Third, no two isolates were identical even when excluding the accessory genome, and this includes isolates recovered days apart from the same ward in Sappasithiprasong Hospital in northeast Thailand. Furthermore, five isolates were differentiated by only 14 SNPs, and these were isolated within a few weeks and from adjacent wards in the same block in the hospital. This observation supports the possibility that this approach could inform on transmission chains even at the scale of a single hospital, which has clear implications for infection control.

More large-scale transmission events were clearly evident in the data of Harris et al. Portugal experienced two waves of infection due to ST239 during the 1990s.

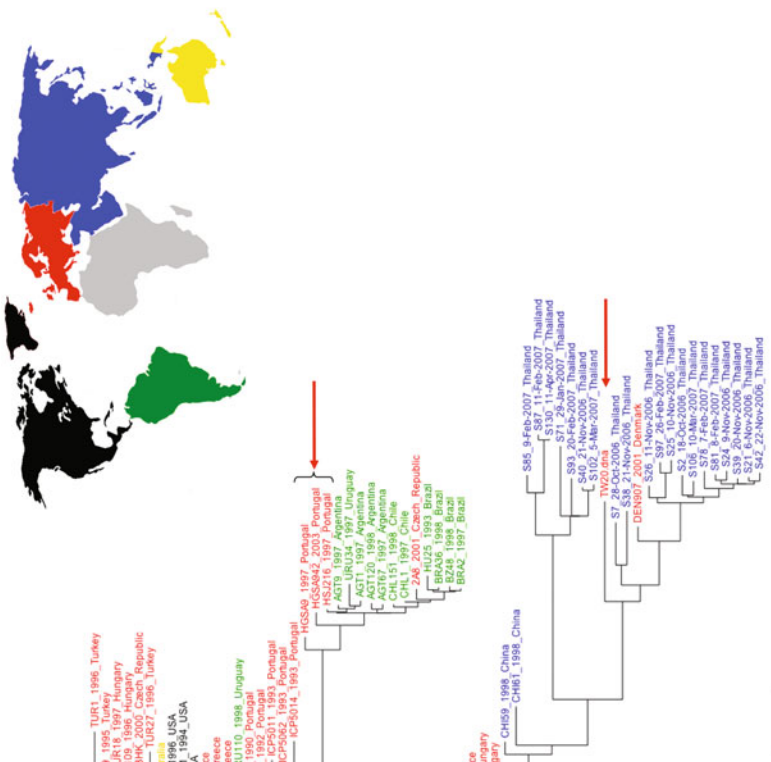


Fig. 5.5 Phylogenetic tree reconstructed using the data of Harris et al. (2010)—only the SNPs corresponding to the core were used. There are few conflicts in the data, and the tree is very robust. The isolates are colour coded according to geographic origin, which reveals striking geographic structure. Discrepancies are highlighted by *red arrows* and represent inter-continental transmission events as discussed in the text

Isolates recovered from the second wave showed subtle band differences by PFGE from the first wave and were more similar to isolates from Brazil than those previously isolated from Portugal. Three Portuguese isolates from this second wave were included in the study of Harris et al., and these clustered with the South American isolates, thus confirming that this second wave of infection in Portugal was seeded from South America. More surprisingly, the TW20 reference genome, which originated from a recent outbreak in an Intensive Care Unit in London (Edgeworth et al. 2007), clustered with the Thai isolates, thus implicating a hitherto unexpected origin—south-east Asian origin—of this outbreak strain.

Selection, Recombination and a Paucity of Homoplasies

In addition to powerful epidemiological surveillance and informing on recent transmission events, next-generation sequence data also sheds light on short-term evolutionary processes. A common feature of all the datasets discussed thus far, and also for those for some monomorphic species such as *Salmonella typhi* (Roumagnac et al. 2006; Holt et al. 2008), is the lack of phylogenetic conflict in the data due to homoplasy. While this is of obvious benefit in enabling easy and robust tree construction, does it betray a deeper significance? Homoplasies are identical character states (in this case, SNPs) that are observed in unrelated lineages, and there are three possible means by which they may be generated. The first is mutational reversal in other sequences, such that homoplasies are in fact identical by descent but have been lost in intervening sequences. In the context of the current discussion, this scenario is so unlikely that it can be effectively discounted. The second possibility is that identical mutational events may arise de novo independently in different lineages. The third possibility is recombination. In this case, SNPs emerge de novo by mutation in one lineage and are then horizontally transferred to other lineages.

Harris et al. noted only 38 homoplasies among the 4,310 SNPs in the core genome ($< 0.1\,\%$). Although recombination is known to be rare in *S. aureus*, it does occasionally happen, the large replacement in ST239 being a notable example. Given the unlikely alternatives, it is therefore reasonable to assume that the majority of the observed homoplasies have arisen through recombination, and this is supported by the physical clustering of these SNPs. Previous statistical and empirical estimates from MLST data have suggested that mutational events impact on the genome $\sim$10–15-fold more frequently than recombination events (Feil et al. 2003; Vos and Didelot 2009). However, the low frequency of homoplasies in the data of Harris et al. is consistent with the view that mutation occurs $>$ 1,000-fold more frequently than recombination. Even with the caveats that homoplasies will only result through recombination between strains present in the dataset, and that recombination events between identical sequences will not be detected, this discrepancy seems too large to ignore.

Harris et al. also considered the selective consequences of the observed homoplasies and noted that almost a third of them correspond to mutations known to confer

antibiotic resistance. This means they will confer a strong selective advantage and be much more likely to be observed than neutral or slightly deleterious changes, which are more likely to be quickly lost through drift. Indeed, an important message from the study of Harris et al. is that examining homoplasies is likely to be a powerful means to identify changes that confer an adaptive advantage, and in particular those conferring drug resistance. What these data also suggest is that the neutral impact of recombination on genome divergence between such extremely closely related *S. aureus* genomes appears to be vanishingly small. Although four synonymous homoplasies clustering within approximately 1 Kb were noted, these are likely to have hitch-hiked with two mutations, conferring trimethoprim resistance, which were located 200–300 bp away.

Speed Dating

Although the figures are to be considered very approximate, the discussion above hints at a striking discrepancy in the dynamics of molecular evolution over extremely short time scales to those between more modestly related lineages within the named species. The relative paucity of homoplasies in the core genome, given the total number of SNPs, in the data of Harris et al. appears to imply that recombination rates in *S. aureus* could be as much as two orders of magnitude lower within ST239 than in the broader population. However, an equivalent interpretation is that recombination rates are similar, or even slightly higher (as may be expected given the high sequence identity), but that mutation rates are approximately two orders of magnitude higher. The estimation of mutation rates (hence dating the emergence of clades) represents a central theme in the high-resolution studies of bacterial populations currently being published, and a clear consensus is emerging. The three intra-clonal *S. aureus* studies discussed above propose strikingly similar figures: 3.3–4.6 $\times$ 10^{-6} per site per year (Smyth et al. 2010), 2.5–4.0 $\times$ 10^{-6}. (Harris et al. 2010) and 1.2 –2.9 $\times$ 10^{-6} (Nubel et al. 2010). It is noteworthy that the estimate of Harris et al., which is likely to be the most accurate as it is based on genome-wide data, falls intermediate between the other two estimates. Harris et al. dated the emergence of ST239 to the mid to late 1960s, whereas the date estimate of Smyth et al. was around a decade earlier. Harris et al. also noted that this rate equates to approximately 1 SNP every 6 weeks (in the core genome), and Smyth et al. similarly opined that this clone is ‘measurably evolving’.

The date estimates given above are approximately 100-fold higher than the standard estimate of 3 $\times$ 10^{-8} for *Escherichia coli* (Achtman et al. 1999) but approximately 10-fold slower than estimates for *Campylobacter jejuni* (Wilson et al. 2009), *Helicobacter pylori* (Falush et al. 2001) and *Neisseria gonorrhoeae* (Perez-Losada et al. 2007). If we cancel out generation times by assuming they are similar between species (although they are in fact notoriously difficult to measure for natural populations), then the faster estimates for these latter species are still easily explained by the fact that they all recombine at a much higher frequency than *S. aureus*. This means that the majority of the mutations observed in these studies will not have arisen de

novo but will have been horizontally acquired. However, this does not explain why all these estimates are so fast relative to the canonical estimate for *E. coli*. The key to understanding this is the fact that these estimates were based on extremely closely related isolates, and there is acceleration of mutation rate as one moves towards the very tips of the trees (Balbi and Feil 2007). This effect has recently been discussed in terms of two categories: a mutation rate and a substitution rate (Achtman 2008, Nubel et al. 2010). The mutation rate is simply the rate at which mutations are generated de novo. The high rate of change within the groups and sub-groups described is assumed to approximate this rate. The substitution rate refers to all the mutations that have subsequently become fixed within the population. This is necessarily much slower than the mutation rate, as most mutations will be lost by selection or drift.

Although classically correct, this interpretation remains something of an oversimplification for species containing substantial sub-structuring. First, as discussed later, there is likely to be a non-trivial proportion of mutations that are so strongly deleterious that they will never be observed; hence, we will always be underestimating the true mutation rate in the natural populations. Second, prokaryotes do not conform to idealised sexual (eukaryotic) populations; what is true for elephants is not always true for *E. coli*. The typical structuring ('clumpiness') of asexual bacterial populations means we must also consider intermediate categories between mutation (polymorphism or standing variation) and substitution, and hence also intermediate rates. For example, what should we call the changes that are fixed (substitutions) on the clonal complex level but are variable (polymorphisms) on the species level? quasisubstitutions perhaps? There may be additional levels to consider; for example, the clonal complexes of *S. aureus* can be robustly delineated into three higher order clades (Cooper and Feil 2006) (Fig. 5.3). This problem basically reflects the arbitrary ring-fencing of 'species' (i.e. populations). As dating estimates are likely to form a central plank of the analyses in re-sequencing studies, this problem deserves careful consideration. Although the relationship between 'mutation' rate and divergence has not yet been examined at the intra-species level for bacteria, one might predict (on the basis of the studies discussed below) a log-linear relationship, such that there is a rapid drop of rate moving back from the most fine-scale clusters, which are known to exhibit high levels of facile polymorphism (Acinas et al. 2004), followed by progressively smaller decreases.

The Purging of Deleterious Mutations Over Time

The decrease in the rate of change moving back from the tips of trees is due to the accumulative purging of slightly deleterious mutations. Purifying selection acts as a sieve in removing these mutations from the population, but this purging does not occur instantaneously; the excess mutations observed at the very tips of the tree represent those slightly deleterious changes destined for subsequent loss. According to the nearly neutral model (Ohta 1973), the efficiency of purifying selection is determined by the selection coefficient (s) and the effective population size (N_e). If $N_e\ s < 1$,

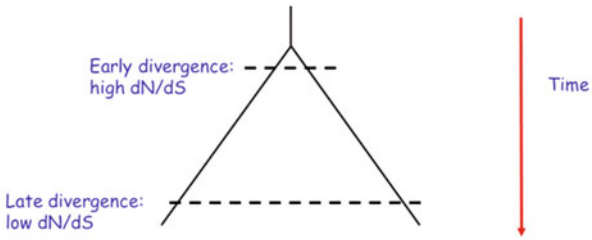


Fig. 5.6 The selective purging of slightly deleterious non-syonymous mutations over time results in a decrease in the dN/dS ratio. (see Rocha et al. 2006)

then the chances of the mutation becoming fixed are predicted to be as if the mutation were neutral, otherwise they are predicted to be slightly deleterious and stand a lower chance of becoming fixed. The progressive purging of slightly deleterious mutations can be easily plotted over time. As de novo non-syonymous mutations are on average more likely to be slightly deleterious than synonymous mutations, they should be enriched between very closely related genomes but preferentially removed over time (Fig. 5.6). Rocha et al. (2006) confirmed this effect by demonstrating that dN/dS ratio between pairs of genomes decreases with increasing divergence, and this effect has subsequently been confirmed by other studies (Hughes et al. 2008; Kryazhimskiy and Plotkin 2008; Garcia Pelayo et al. 2009; Novichkov et al. 2009). Furthermore, Rocha et al. showed by simulation that the trajectory of this decline is dependent upon the both N_e and s, as predicted under the nearly neutral theory.

It is important to note that there is no reason to suppose that this effect is reserved for non-syonymous changes. Although the dN/dS ratio is a very convenient metric, the same principal should apply to any kind of slightly deleterious change relative to a more neutral counterpart. For example, different mutation types arise de novo at different frequencies, and GC→ AT mutations are far more common than the reverse. This bias is so strong that in the absence of selection, the equilibrium base composition of most bacterial genomes would rest at ~20 % GC. This can explain a key characteristic of the genomes of endosymbiotic bacterial species that live inside insects. These species are transmitted from mother to offspring trans-ovarially down the generations (vertically), a lifestyle that imposes repeated bottlenecking, thus lowering the N_e and weakening selection. This is thought to explain why such genomes are characterised by very high AT contents and very high dN/dS ratios, both of which reflect weak purging of slightly deleterious mutations (Moran et al. 2009).

Balbi et al. (2009) examined the selective purging over time of both non-syonymous mutations and GC→ AT mutations (relative to their counterparts) through a comparative genomic analysis of *E. coli* and *Shigella*. The N_e of *E. coli* can be assumed to be very large, as this is a very generalist species, which occupies a wide range of animal guts (usually without harm to the host) and is able to survive in soil and water (Hartl et al. 1994). The four *Shigella* named 'species' are the causative agents of bacillary dysentery and are essentially specialised clones of *E. coli* that have independently acquired a large 'invasion' plasmid, which confers on them the ability to invade host cells and replicate intracellularly (Lan and Reeves 2002). This adaptation to a specialised lifestyle may be accompanied by a reduction in N_e, in a

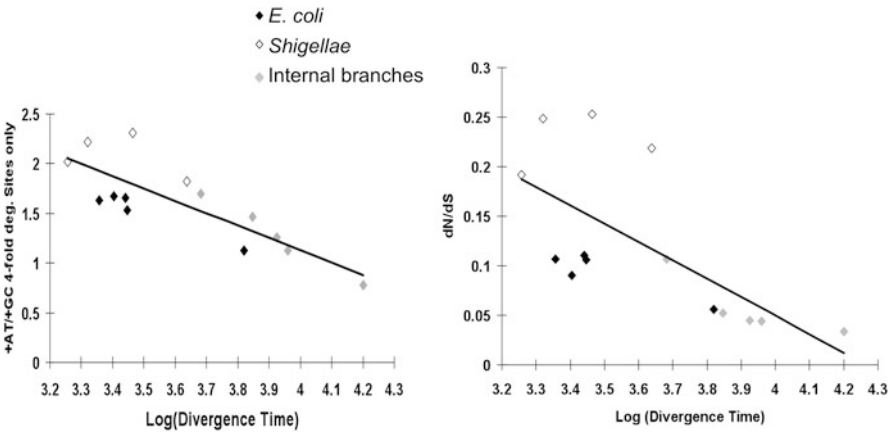


Fig. 5.7 The selective purging of GC→AT mutations, relative to their counterpart, (**a**) and non-syonymous mutations (**b**) over time in *Shigella* and *E. coli*. The effect is weaker in *Shigella* than in *E. coli* because the former has a smaller effective population size. 'Divergence time' is measured as the total number of SNPs, and the relationships are log-linear. (see Balbi et al. 2009)

way somewhat analogous to 'island' populations of eukaryotes (Johnson and Seger 2001; Balbi and Feil 2007).

Balbi et al. examined the molecular evolutionary consequences of this niche restriction by comparing both the proportions of GC→ AT polymorphisms over the reverse and dN/dS in the *E. coli* and *Shigella* genomes. They confirmed two key predictions: (i) slightly deleterious mutations (synonymous or GC→ AT) are selectively purged over time, and (ii) slightly deleterious mutations are enriched in *Shigella* relative to *E. coli* (when divergence time is considered). Figure 5.7 reveals a log-linear purging of two types of slightly deleterious mutations over time (a: GC→ AT mutations; b: non-synonymous mutations). In both cases this purging is less marked in *Shigella* than it is in *E. coli*. Note that only 4-fold degenerate sites are used for the GC→ AT analysis, hence these two effects are independent and the results challenge the widely held view that synonymous changes can be assumed to be neutral. Transversions behave in the same way (relative to transitions), but this is because transversions are more likely to be non-synonymous (Balbi et al. 2009).

Synthesising Selection and Ecology Using Next-Generation Sequence Data

How can genome-wide data help to inform on such analyses? The enrichment of non-synonymous changes between very closely related genomes, with subsequent purification of these SNPs over time, has been repeatedly confirmed for *S. aureus* (Castillo-Ramirez et al. 2011) and a number of other species (Larsson et al. 2009; Holt et al. 2008; He et al. 2010). However, the increased resolution afforded by

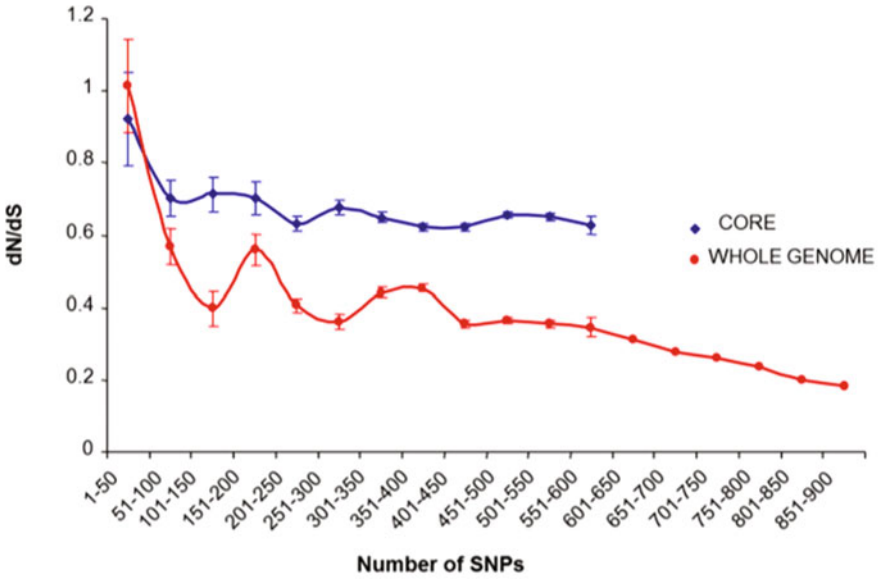


Fig. 5.8 The relationship between dN/dS and divergence in the data of Harris et al. (2010). Considering the whole genome, a decline in dN/dS is noted over time, but for the core SNPs, the ratio remains high. This is probably because core SNPs represent recent de novo mutation and non-core SNPs are horizontally acquired. The initial drop-off of dN/dS implicates a class of highly deleterious mutation. The lack of subsequent decline for the core data may reflect the fact that ST239 is specifically adapted to the hospital environment, which would lower N_e and weaken selection (see text)

next-generation sequencing will reveal the dynamics of this process in greater detail. Revisiting the data of Harris et al., we plotted the pairwise dN/dS of ST239 isolates against the total number of SNPs for the whole genome and for only the core genome (Fig. 5.8). This plot for the whole genome is consistent with a decrease in dN/dS over time but, curiously, the trajectory for just the core genome reveals much higher dN/dS and, after an initial decrease, no obvious downward trend. A likely explanation is that many of the SNPs in the accessory genome have been acquired via horizontal transfer, whereas those in the core genome have arisen by de novo mutation. This means that the latter mutations are, on average, younger and (by the argument outlined earlier) should therefore contain a higher proportion of non-synonymous mutations. This effect is also apparent from Fig. 5.9, where regions of high SNP density (corresponding to various non-core elements) show a relative enrichment of synonymous changes relative to the core genome. Essentially, those SNPs acquired by horizontal transfer are more likely to be synonymous because they have already passed through a selective filter in the wider population. An analysis of the large replacement present in ST239 supports this argument (Castillo-Ramirez et al. 2011).

The dN/dS plot for the core genome only (Fig. 5.8) raises two other points for discussion. First, the rapid initial decrease in dN/dS is indicative of a class of mutation, which is strongly deleterious, hence quickly expunged. Such an approach provides

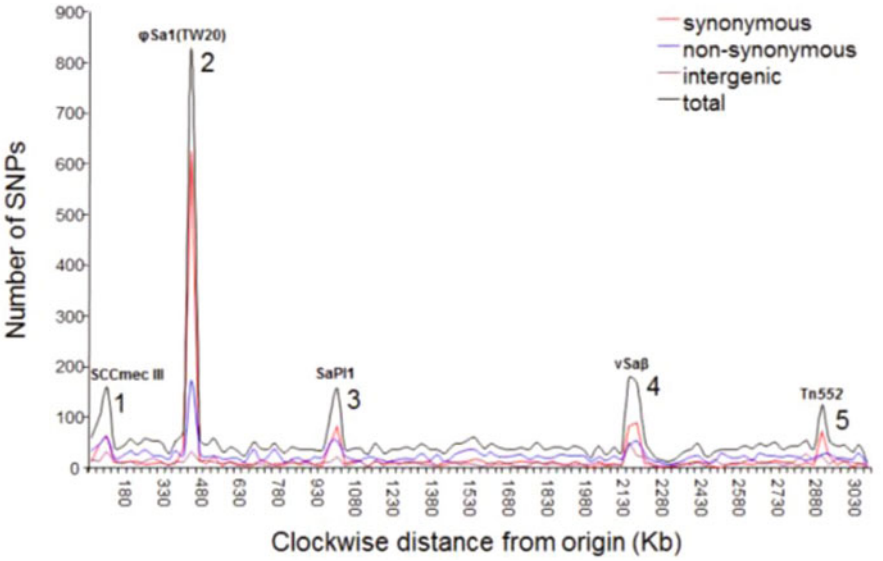


Fig. 5.9 The distribution of SNPs across the TW20 (reference) genome. The five peaks of high SNP density correspond to well characterised components of the accessory genome: *SCC*mec III (a resistance cassette), φSa1(TW20) (a prophage), SaPI1 (a pathogenicity island), νSaβ (a genomic island) and Tn*552* (a transposon). The figure illustrates that for peaks 2–5, synonymous SNPs are more common than non-syonymous SNPs, whereas the reverse is true for regions between the peaks. Other well-characterised elements known to be present in the TW20 genome do not correspond to a clear peak of SNP density because they are less variable between ST239 isolates. However, any region of > 1 Kb, which is not present within all 63 ST239 isolates, regardless of SNP density, is defined as non-core in this analysis

a convenient means to estimate the size of this proportion, and thus goes some way towards estimating the distribution of fitness effects (at least the part corresponding to deleterious mutations). Second, the observation that the dN/dS ratio does not decrease over subsequent divergence bins suggests that only weak purifying selection is operating over this time window. This may solely reflect the short time scale since the emergence of ST239 (∼45 years), or it may, in part, also reflect weakened selection due to the specialised epidemiology of this clone. As ST239 is hospital-adapted, it is possible that it experiences repeated bottlenecks leading to a reduction in N_e. If so, this would have significant implications for health care management, as it would suggest that such clones may be self-limiting, and such a possibility is consistent with the near disappearance of ST239 from western Europe over the last decade or so.

Concluding Remarks

Next-generation sequencing clearly heralds a new dawn in bacterial micro-evolution. As for previous advances, such as MLST, the initial studies have been on pathogenic bacteria, with a view to increased power of epidemiological surveillance and under-

standing patterns of transmission. However, we have shown how these data may also inform on basic evolutionary processes and the molecular consequences of lifestyle changes; advances such as these are certain to equally revolutionise studies on environmental species. Here, we have focused on a single approach to illustrate how comparisons of the trajectories of purifying selection may inform on population level effects relating to bacterial lifestyle. Such analyses may even help to define the limits of drift, thus finally providing a conceptual basis for identifying population boundaries. As the trajectories depend both on N_e and s, comparisons may also be drawn between different mutation types, or between different classes of genes. Many other analytical approaches can be taken with these data, for example, understanding the dynamics of gene loss and gain, the size of the accessory gene 'library' from which any strain of a given population may access and how evolutionary processes vary according to genomic position. An excellent and comprehensive recent study on multiple *E. coli* genomes provides a taste of the many such delights around the corner (Touchon et al. 2009).

Acknowledgments SCR is funded by the TROCAR consortium (FP7-HEALTH #223031). We are grateful to colleagues present at the first international PERMAFROST meeting, Bormio, Italy; March 5–8, 2010, for fruitful discussions.

References

Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu Rev Microbiol 62:53–70

Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 96:14043–14048

Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, et al. (2004) Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430: 551–554.

Aires De Sousa M, Conceicao T, Simas C, De Lencastre H (2005) Comparison of genetic backgrounds of methicillin-resistant and -susceptible *Staphylococcus aureus* isolates from Portuguese hospitals and the community. J Clin Microbiol 43:5150–5157

Alm RA, Trust TJ (1999) Analysis of the genetic diversity of Helicobacter pylori: the tale of two genomes. J Mol Med (Berl) 77: 834–846.

Amaral MM, Coelho LR, Flores RP, Souza RR, Silva-Carvalho MC, et al. (2005) The predominant variant of the Brazilian epidemic clonal complex of methicillin-resistant Staphylococcus aureus has an enhanced ability to produce biofilm and to adhere to and invade airway epithelial cells. J Infect Dis 192: 801–810.

Balbi KJ, Feil EJ (2007) The rise and fall of deleterious mutation. Res Microbiol 158:779–786

Balbi KJ, Rocha EP, Feil EJ (2009) The temporal dynamics of slightly deleterious mutations in *Escherichia coli* and *Shigella* spp. Mol Biol Evol 26:345–355

Basset P, Hammer NB, Kuhn G, Vogel V, Sakwinska O, Blanc DS (2009) *Staphylococcus aureus* clfB and spa alleles of the repeat regions are segregated into major phylogenetic lineages. Infect Genet Evol 9:941–947

Boyd EF, Nelson K, Wang FS, Whittam TS, Selander RK (1994) Molecular genetic basis of allelic polymorphism in malate dehydrogenase (mdh) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc Natl Acad Sci U S A 91:1280–1284

Castillo-Ramirez S, Harris SR, Holden MT, He M, Parkhill J, et al. (2011) The impact of recombination on dN/dS within recently emerged bacterial clones. PLoS Pathog 7: e1002129.

Chan MS, Maiden MC, Spratt BG (2001) Database-driven multi locus sequence typing (MLST) of bacterial pathogens. Bioinformatics 17:1077–1083

Chongtrakool P, Ito T, Ma XX, Kondo Y, Trakulsomboon S, et al. (2006) Staphylococcal cassette chromosome mec (SCCmec) typing of methicillin-resistant Staphylococcus aureus strains isolated in 11 Asian countries: a proposal for a new nomenclature for SCCmec elements. Antimicrob Agents Chemother 50: 1001–1012.

Conceicao T, Aires-De-Sousa M, Fuzi M, Toth A, Paszti J, Ungvari E, Van Leeuwen WB, Van Belkum A, Grundmann H, De Lencastre H (2007) Replacement of methicillin-resistant *Staphylococcus aureus* clones in Hungary over time: a 10-year surveillance study. Clin Microbiol Infect 13:971–979

Cookson BD, Robinson DA, Monk AB, Murchan S, Deplano A, et al. (2007) Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant Staphylococcus aureus strains: the HARMONY collection. J Clin Microbiol 45: 1830–1837.

Cooper JE, Feil EJ (2006) The phylogeny of *Staphylococcus aureus*—which genes make the best intra-species markers? Microbiology 152:1297–1305

Crisostomo MI, Westh H, Tomasz A, Chung M, Oliveira DC, De Lencastre H (2001) The evolution of methicillin resistance in *Staphylococcus aureus*: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. Proc Natl Acad Sci U S A 98:9865–9870

De Lencastre H, Oliveira D, Tomasz A (2007) Antibiotic resistant *Staphylococcus aureus*: a paradigm of adaptive power. Curr Opin Microbiol 10:428–435

Deleo FR, Otto M, Kreiswirth BN, Chambers HF (2010) Community-associated meticillin-resistant *Staphylococcus aureus*. Lancet 375:1557–1568

Diekema DJ, Pfaller MA, Schmitz FJ, Smayevsky J, Bell J, et al. (2001) Survey of infections due to Staphylococcus species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY Antimicrobial Surveillance Program, 1997–1999. Clin Infect Dis 32 Suppl 2: S114–132.

Dorrell N, Mangan JA, Laing KG, Hinds J, Linton D, et al. (2001) Whole genome comparison of Campylobacter jejuni human isolates using a low-cost microarray reveals extensive genetic diversity. Genome Res 11: 1706–1715.

Dubose RF, Dykhuizen DE, Hartl DL (1988) Genetic exchange among natural isolates of bacteria: recombination within the phoA gene of *Escherichia coli*. Proc Natl Acad Sci U S A 85:7036–7040

Edgeworth JD, Yadegarfar G, Pathak S, Batra R, Cockfield J D, Wyncoll D, Beale R, Lindsay JA (2007) An outbreak in an intensive care unit of a strain of methicillin-resistant *Staphylococcus aureus* sequence type 239 associated with an increased rate of vascular access device-related bacteremia. Clin Infect Dis 44:493–501

Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG,(2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc Natl Acad Sci U S A 99:7687–7692

Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, Suerbaum S (2001) Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size, and minimal age. Proc Natl Acad Sci U S A 98:15056–15061

Feil EJ (2004) Small change: keeping pace with microevolution. Nat Rev Microbiol 2:483–495

Feil EJ, Cooper JE, Grundmann H, Robinson DA, Enright MC, Berendt T, Peacock SJ, Smith JM, Murphy M, Spratt BG, Moore CE, Day NP (2003) How clonal is *Staphylococcus aureus*? J Bacteriol 185:3307–3316

Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186:1518–1530

Feil EJ, Nickerson EK, Chantratita N, Wuthiekanun V, Srisomang P, Cousins R, Pan W, Zhang G, Xu B, Day NP, Peacock SJ (2008) Rapid detection of the pandemic methicillin-resistant *Staphylococcus aureus* clone ST 239, a dominant strain in Asian hospitals. J Clin Microbiol 46:1520–1522

Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR, Musser JM (2001) Evolutionary genomics of Staphylococcus aureus: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc Natl Acad Sci U S A 98: 8821–8826.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269: 496–512.

Garcia Pelayo MC, Uplekar S, Keniry A, Mendoza Lopez P, Garnier T, Nunez Garcia J, Boschiroli L, Zhou X, Parkhill J, Smith N, Hewinson RG, Cole ST, Gordon SV (2009) A comprehensive survey of single nucleotide polymorphisms (SNPs) across *Mycobacterium bovis* strains and *M. bovis* BCG vaccine strains refines the genealogy and defines a minimal set of SNPs that separate virulent *M. bovis* strains and *M. bovis* BCG strains. Infect Immun 77:2230–2238

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van De Peer Y, Vandamme P, Thompson FL, Swings J (2005) Opinion: re-evaluating prokaryotic species. Nat Rev Microbiol 3:733–739

Gomes AR, Westh H, De Lencastre H (2006) Origins and evolution of methicillin-resistant *Staphylococcus aureus* clonal lineages. Antimicrob Agents Chemother 50:3237–3244

Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, De Lencastre H, Parkhill J, Peacock SJ, Bentley SD (2010) Evolution of MRSA during hospital transmission and intercontinental spread. Science 327:469–474

Hartl DL, Moriyama EN, Sawyer SA (1994) Selection intensity for codon bias. Genetics 138:227–234

He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, et al. (2010) Evolutionary dynamics of Clostridium difficile over short and long time scales. Proc Natl Acad Sci U S A 107: 7527–7532.

Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, Enright MC, Foster TJ, Moore CE, Hurst L, Atkin R, Barron A, Bason N, Bentley SD, Chillingworth C, Chillingworth T, Churcher C, Clark L, Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S, Jagels K, James KD, Lennard N, Line A, Mayes R, Moule S, Mungall K, Ormond D, Quail MA, Rabbinowitsch E, Rutherford K, Sanders M, Sharp S, Simmonds M, Stevens K, Whitehead S, Barrell BG, Spratt BG, Parkhill J (2004a) Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci U S A 101:9786–9791

Holden MT, Lindsay JA, Corton C, Quail MA, Cockfield JD, Pathak S, Batra R, Parkhill J, Bentley SD, Edgeworth JD (2010) Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). J Bacteriol 192:888–892

Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, Pitt T, Churcher C, Mungall K, Bentley SD, Sebaihia M, Thomson NR, Bason N, Beacham IR, Brooks K, Brown KA, Brown NF, Challis GL, Cherevach I, Chillingworth T, Cronin A, Crossett B, Davis P, Deshazer D, Feltwell T, Fraser A, Hance Z, Hauser H, Holroyd S, Jagels K, Keith KE, Maddison M, Moule S, Price C, Quail MA, Rabbinowitsch E, Rutherford K, Sanders M, Simmonds M, Songsivilai S, Stevens K, Tumapa S, Vesaratchavest M, Whitehead S, Yeats C, Barrell BG, Oyston PC, Parkhill J (2004b) Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. Proc Natl Acad Sci U S A 101:14240–14245

Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M, Dougan G (2008) High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. Nat Genet 40:987–993

Hughes AL, Friedman R, Rivailler P, French JO (2008) Synonymous and nonsynonymous polymorphisms versus divergences in bacterial genomes. Mol Biol Evol 25:2199–2209

Johnson KP, Seger J (2001) Elevated rates of nonsynonymous substitution in island birds. Mol Biol Evol 18:874–881

Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/dS. PLoS Genet 4:e1000304
Kuhn G, Francioli P, Blanc DS (2006) Evidence for clonal evolution among highly polymorphic genes in methicillin-resistant *Staphylococcus aureus*. J Bacteriol 188:169–178
Lan R, Reeves PR (2002) *Escherichia coli* in disguise: molecular origins of *Shigella*. Microbes Infect 4:1125–1132
Larsson P, Elfsmark D, Svensson K, Wikstrom P, Forsman M, Brettin T, Keim P, Johansson A (2009) Molecular evolutionary consequences of niche restriction in *Francisella tularensis*, a facultative intracellular pathogen. PLoS Pathog 5:e1000472
Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95: 3140–3145.
Maiden MC (2006) Multilocus sequence typing of bacteria. Annu Rev Microbiol 60:561–588
Melles DC, Schouls L, Francois P, Herzig S, Verbrugh HA, Van Belkum A, Schrenzel J (2009) High-throughput typing of *Staphylococcus aureus* by amplified fragment length polymorphism (AFLP) or multi-locus variable number of tandem repeat analysis (MLVA) reveals consistent strain relatedness. Eur J Clin Microbiol Infect Dis 28:39–45
Mellmann A, Weniger T, Berssenbrugge C, Keckevoet U, Friedrich AW, Harmsen D, Grundmann H (2008) Characterization of clonal relatedness among the natural population of *Staphylococcus aureus* strains by using spa sequence typing and the BURP (based upon repeat patterns) algorithm. J Clin Microbiol 46:2805–2808
Milkman R, Bridges MM (1990) Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics 126:505–517
Moran NA, Mclaughlin HJ, Sorek R (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 379–382
Nelson K, Selander RK (1992) Evolutionary genetics of the proline permease gene (putP) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J Bacteriol 174:6886–6895
Nelson K, Selander RK (1994) Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (gnd) in enteric bacteria. Proc Natl Acad Sci U S A 91:10227–10231
Nelson K, Whittam TS, Selander RK (1991) Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (gapA) in natural populations of *Salmonella* and *Escherichia coli*. Proc Natl Acad Sci U S A 88:6667–6671
Novichkov PS, Wolf YI, Dubchak I, Koonin EV (2009) Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. J Bacteriol 191:65–73
Nubel U, Dordel J, Kurt K, Strommenger B, Westh H, et al. (2010) A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant Staphylococcus aureus. PLoS Pathog 6: e1000855.
Nubel U, Dordel J, Kurt K, Strommenger B, Westh H, Shukla SK, Zemlickova H, Leblois, R, Wirth T, Jombart T, Balloux F, Witte W (2010) A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. PLoS Pathog 6:e1000855
Nubel U, Roumagnac P, Feldkamp M, Song JH, Ko KS, Huang YC, Coombs G, Ip M, Westh H, Skov R, Struelens MJ, Goering RV, Strommenger B, Weller A, Witte W, Achtman M (2008) Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. Proc Natl Acad Sci U S A 105:14130–14135
Nulens E, Gould I, MacKenzie F, Deplano A, Cookson B, et al. (2005) Staphylococcus aureus carriage among participants at the 13th European Congress of Clinical Microbiology and Infectious Diseases. Eur J Clin Microbiol Infect Dis 24: 145–148.
Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304
Ohta T (1973) Slightly deleterious mutant substitutions in evolution. Nature 246:96–98
Peacock SJ, de Silva I, Lowy FD (2001) What determines nasal carriage of Staphylococcus aureus? Trends Microbiol 9: 605–610.

Perez-Losada M, Crandall KA, Zenilman J, Viscidi RP (2007) Temporal trends in gonococcal population genetics in a high prevalence urban community. Infect Genet Evol 7:271–278

Robinson DA, Enright MC (2004) Evolution of *Staphylococcus aureus* by large chromosomal replacements. J Bacteriol 186:1060–1064

Robinson DA, Monk AB, Cooper JE, Feil EJ, Enright MC (2005) Evolutionary genetics of the accessory gene regulator (agr) locus in *Staphylococcus aureus*. J Bacteriol 187:8312–8321

Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, Smith NH, Feil EJ (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. J Theor Biol 239:226–235

Roumagnac P, Weill FX, Dolecek C, Baker S, Brisse S, Chinh NT, Le TA, Acosta CJ, Farrar J, Dougan G, Achtman M (2006) Evolutionary history of *Salmonella typhi*. Science 314:1301–1304

Schouls LM, Spalburg EC, Van Luit, M, Huijsdens XW, Pluister GN, Van Santen-Verheuvel MG, Van Der Heide HG, Grundmann H, Heck ME, De Neeling AJ (2009) Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with pulsed-field gel electrophoresis and spa-typing. PLoS ONE 4:e5082

Smith JM, Smith NH, O'rourke M, Spratt BG (1993) How clonal are bacteria? Proc Natl Acad Sci U S A 90:4384–4388

Smith NH, Dale J, Inwald J, Palmer S, Gordon SV, Hewinson RG, Smith JM (2003) The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. Proc Natl Acad Sci U S A 100:15271–15275

Smith NH, Gordon SV, De La Rua-Domenech R, Clifton-Hadley RS, Hewinson RG (2006) Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. Nat Rev Microbiol 4:670–681

Smyth DS, Mcdougal LK, Gran FW, Manoharan A, Enright MC, Song JH, De Lencastre H, Robinson DA (2010) Population structure of a hybrid clonal group of methicillin-resistant *Staphylococcus aureus*, ST239-MRSA-III. PLoS ONE 5:e8582

Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguenec C, Lescat M, Mangenot S, Martinez-Jehanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Medigue C, Rocha EP, Denamur E (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet 5:e1000344

Turner KM, Feil EJ (2007) The secret life of the multilocus sequence type. Int J Antimicrob Agents 29:129–135

van Belkum A (2006) Staphylococcal colonization and infection: homeostasis versus disbalance of human (innate) immunity and bacterial virulence. Curr Opin Infect Dis 19: 339–344.

Vanderhaeghen W, Hermans K, Haesebrouck F, Butaye P (2010) Methicillin-resistant *Staphylococcus aureus* (MRSA) in food production animals. Epidemiol Infect 138:606–625

Vos M, Didelot X (2009) A comparison of homologous recombination rates in bacteria and archaea. ISME J 3:199–208

Waldron DE, Lindsay JA (2006) Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. J Bacteriol 188:5578–5585

Wilson DJ, Gabriel E, Leatherbarrow AJ, Cheesbrough J, Gee S, Bolton E, Fox A, Hart CA, Diggle PJ, Fearnhead P (2009) Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. Mol Biol Evol 26:385–397

Xu BL, Zhang G, Ye HF, Feil EJ, Chen GR, et al. (2009) Predominance of the Hungarian clone (ST 239-III) among hospital-acquired meticillin-resistant Staphylococcus aureus isolates recovered throughout mainland China. J Hosp Infect 71: 245–255.

Chapter 6
Role of Symbiosis in Evolution

Amparo Latorre and Andrés Moya

Introduction

The term symbiosis comes from two Greek words sym (together) and bios (life). It was first used by Anton de Bary in 1879 who defined it as "the living together of unlike organisms," although this definition did not consider the harmful/beneficial aspect of the association. The crucial role of symbiogenesis in the generation of biological novelty was proposed by the Russian school embodied by K. Merezhkousky, A. Faminstyn, and Kozo-Polyansky in the late nineteenth and early twentieth century. However, Russian literature was not available in English until 1922. In 1967, Lynn Margulis, salvaging these forgotten works, posed the endosymbiotic theory of cell evolution, currently universally accepted with respect to the origin of mitochondria and chloroplasts. The study of symbiosis has acquired great importance during recent decades, thanks, in particular, to approaches for the isolation and characterization of genes or genomes from microorganisms, which cannot usually be cultured. This has provided the opportunity to test certain theories on the evolution of symbiotic associations put forward within evolutionary biology. Indeed, genomic studies focusing on endosymbiotic alliances have revealed that symbiogenesis (currently defined as the mechanism by which new structures, biochemical or behavioral processes, are established) has far greater impact than anyone could have suspected just a few years ago.

Symbiosis in the broad sense of "living together" is defined as the long-term and close physical contact between two organisms. However, regarding the benefit that such organisms may confer, three types of associations are defined: *parasitism*, when the fitness of one organism is enhanced at the expense of the other; *mutualism*, when

A. Latorre (✉) and A. Moya
Unidad Mixta de Investigación en Genómica y Salud del Centro Superior de Investigación Pública (CSISP) de la Conselleria de Sanitat del Gobierno Valenciano (València, Spain) y el Instituto Cavanilles de Biodiversidad y Biología Evolutiva de la Universitat de València (València, Spain). Centro de Investigación en Red en Epidemiología y Salud Pública (CIBEResp), Spain
e-mail: amparo.latorre@uv.es; andres.moya@uv.es

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_6,

the fitness of both is enhanced; and *commensalism*, when the fitness of one of them is enhanced without affecting the other. According to the location of the symbiont with respect to the host cell, we talk about *endosymbiosis* when the organism, usually a bacterium, is literally confined inside the eukaryote, which often develops specialized cells to house the bacterium, called bacteriocytes. Alternatively, we talk about *ectosymbiosis* when the symbiont lives on the surface of the host, including internal surfaces such as the gastrointestinal tract and glands.

However, the boundaries between such associations are unclear and we cannot exclude the possibility, in fact most likely, that there are evolutionary transitions from one to another. When an organism (usually a microorganism) enters a host (usually a eukaryote), the mechanisms determining whether a mutualistic or parasitic association will arise (with utterly different outcomes for the host) are still unknown. The distinction between a parasitic and a mutualistic symbiont is primarily based upon the impact it has on its host. However, from the bacterium-centered view, both lifestyles involve different molecular solutions to a set of similar problems that must be solved, namely to hold fast, and in the event the endosymbiont enters the host cell, to multiply, get its cells to colonize new cells and/or new hosts, and to trigger evasion mechanisms against the host's immune response. There is heated debate over whether mutualistic endosymbionts arose from parasitic bacteria or vice versa. Phylogenetic studies and comparative genomics seem to support the hypothesis that today's endosymbionts could have evolved from bacterial pathogens, which have somehow minimized this trait, thus becoming beneficial for both members of the partnership (Moran et al. 2008; Moya et al. 2008).

Some Examples of Microbial Symbioses

Today, many examples of symbioses with microorganisms are known (Moya et al. 2008)—each one is fascinating and deserves a separate chapter. Symbiotic associations with prokaryotes are common among protozoa. Many of these unicellular eukaryotes feed on bacteria, which are swallowed during phagocytosis. A fascinating case is that of the ameba *Amoeba proteus*, which was infected by bacteria called X-bacteria. Initially, a parasitic association was established, leading to cell death; however, nowadays, we can find some lineages where they have become mutualistic obligate partners. Leguminous plants have characteristic nodules on their roots, which harbor bacteria belonging to the genus *Rhizobium,* capable of fixing nitrogen from the air and thus making it available to the plant. Consequently, they can grow in nitrogen-poor soils. Also, among eukaryotic organisms, there are many very significant cases of symbiosis, such as associations between fungi and other fungi, protists, animals, plants (such as mycorrhizae) or algae. This is the case of lichens, which are fungi that contain photosynthetically active algae or cyanobacteria. In fact, lichens provide a typical example of symbiogenesis, because the lichen as a whole is more than its two parts separately. It is neither a green alga nor a cyanobacterium, nor is it a fungus. Finally, symbioses between bacteria and protozoa are common in the gut

to digest certain substances; for example, bacteria located in the rumen of ruminants, which enables them to digest cellulose from grasses; in the human intestine, where they form the intestinal microbiota, helping to digest food and absorb nutrients; or finally, in the termite, which harbors bacterial communities in its digestive tract thus enabling it to digest wood.

The First Symbioses

Without doubt, intracellular endosymbionts, which originally descended from free-living prokaryotes, have played an important role in the evolution of eukaryotes. This gains greater significance still with respect to the origin of eukaryotic cells, a chimera of several prokaryotes. Mitochondria (prokaryotes with great oxygen-consuming capacity) were free-living alphaproteobacteria in the past, providing their host with respiration and energy-metabolism efficiency. Chloroplasts, once free-living cyanobacteria, provided their hosts with the ability to photosynthesize (chlorophyll-harboring prokaryotes became photosynthetic cells) and both organelles have made a substantial contribution to the complementary gene set still present in nuclear genomes. Although the nature of the first host cell is a matter of heated debate among cell evolutionists, nowadays, the majority agrees that it was unicellular, possibly belonging to Archaea, and lacked mitochondria. But beyond this, biologists do not agree on the nature of its intracellular organization, its biochemical lifestyle, or how many and which genes it possessed. The ability to breathe oxygen as a result of the acquisition of mitochondria led to the origin of animals, whereas the photosynthetic ability acquired subsequent to chloroplasts gave rise to plants. In both cases, mitochondria and chloroplasts originated from free-living bacteria, whose descendants are still among us.

Symbiosis in Action: The Insect Model

In a study we conducted of well-established stable symbioses between eukaryotes and prokaryotes (Bacteria and Archaea), we found that symbioses are widespread in nature, with symbiotic associations having been documented in virtually every major branch of the tree of life. Indeed, this observation, together with additional data from phylogenetic and genomic studies of different associations, reinforces the view that symbiosis is a key mechanism in the emergence of evolutionary novelty in eukaryotes. Most symbiotic associations have a biochemical basis because, as the heterotrophic metabolism of animals is very limited, they must obtain essential amino acids, vitamins, fatty acids, etc. from their diet; however, if the organism is poor in some of these nutrients, it must obtain them elsewhere, such as from symbiotic bacteria. Insects represent one of the groups in which most examples of mutualistic symbioses have been described and also for which most experimental and genomic studies have been carried out (Gosalbes et al. 2010).

Insects are the group of organisms comprising the greatest number of species, and approximately 15–20 % of them are estimated to have established symbiotic relationships with bacteria, on which they depend for their viability and reproduction. Because of this strict dependence, these bacteria are called primary endosymbionts. Insects emerged as the arthropod lineage some 385 million years ago and diversified rapidly. A key factor in the evolutionary success of this group of organisms could be the establishment of mutualistic associations with bacteria, which has enabled them to explore a wide variety of niches. For this reason, insects provide an ideal model to examine the evolutionary impact of symbiosis. A fundamental feature of mutualistic symbioses with insects concerns the insect's diet. Their diet is usually characterized by being specialized and thus limited in nutrients, which are supplied by the bacteria. For example, aphids, psyllids, whiteflies, and mealybugs feed on plant sap, an unbalanced diet, because it is rich in sugars and deficient in amino acids and vitamins as well as certain essential lipids. To solve this problem, insects have established symbiotic relationships with bacteria (aphids with *Buchnera aphidicola*, psyllids with *Carsonella ruddii,* whiteflies with *Portiera aleyrodidarum*, and mealybugs with *Tremblaya princeps*), which provide the nutrients deficient in their diet. Likewise, the tsetse fly feeds on blood, which is deficient in B vitamins, in some case with the help of other symbiont supplied by its primary endosymbiont, the bacterium *Wigglesworthia glossinidia*. The xylem is the part of the vascular system of plants that transports water and salts from the roots to the rest of the plant and is therefore nutrient deficient. It contains mainly inorganic compounds and mineral salts as well as small amounts of nonessential amino acids, sugars, and organic acids. Despite this nutritional deficiency, the glassy-winged sharpshooter can live on xylem sap thanks, also, to its endosymbiotic bacteria, namely *Baumannia cicadellinicola* and *Sulcia muelleri*. Finally, although cockroaches and ants are omnivorous they also have endosymbionts. The sequencing of the genome of their respective endosymbionts, *Blattabacterium* sp. and *Blochmannia* sp., has revealed that the bacteria in both insects (in a fascinating case of convergent evolution) are involved in nitrogen recycling. In all these cases, the insect–bacterium partnership is one of obligate mutualism. The nutrient-provisioning bacteria supply the insects with what their diets lack, while benefiting from the stable intracellular environment their hosts provide with a permanent supply of resources (for an updated review of these symbioses see Gosalbes et al. 2010).

Sequencing the first genomes of endosymbiotic bacteria and comparing them with related free-living bacteria has revealed that they all share a number of common traits. The phylogeny obtained with the 16S rDNA gene, considered the universal marker, clearly shows that *Buchnera, Wigglesworthia*, and *Blochmannia*, endosymbionts of aphids, tsetse flies, and ants, respectively, are related to free-living bacteria, such as *Escherichia coli*. The fact that the latter have genome sizes of 4–5 megabases (Mb) while bacterial endosymbionts have around 600–700 kilobases (kb) illustrates how they have undergone drastic genome reduction in the transition from a free-living to an intracellular way of life, having lost more than 3 Mb (about 3,000 genes). Furthermore, other molecular, structural, and biochemical changes have taken place, as well as a mutational bias toward AT, high mutation rate, loss of

efficient recombination and DNA repair mechanisms, loss of bias in codon usage, and, in some cases, the presence of plasmids for amino acid biosynthesis.

One of the best-studied associations is that of aphids with *B. aphidicola*, its primary endosymbiont. Aphids have developed specialized cells, the bacteriocytes, to house the bacteria, which are located below the digestive tract in the cross section of an aphid. Transmission is strictly vertical, through infection of eggs or early embryos by a few bacteria; thus, *Buchnera* has coevolved with its hosts. Phylogenetic studies indicate that the initial association between *Buchnera* and aphids was a unique event occurring some 180 million years ago. Before the genomic era, studies of aposymbiotic aphids treated with antibiotics and fed on artificial diets revealed that *Buchnera* supplies the aphid with the essential amino acids that it cannot get from the plant.

So far, genomes of *Buchnera* from five different aphid species have been sequenced: *Acyrthosiphum pisum* (*B. aphidicola* BAp), *Schizaphis graminum* (*B. aphidicola* BSg), *Baizongia pistaceae* (*B. aphidicola* BBp), *Cinara cedri* (*B. aphidicola* BCc) and *Cinara tujafilina* (*B. aphidicola* BCt). The first three have similar genome sizes, between 615 and 641 kb. The gene content of their genomes confirmed the nutritional role postulated for *Buchnera*, as there are the biosynthetic pathways for essential amino acids that are lacking in the aphids' phloem sap diet (Lamelas et al. 2011a).

Natural Minimal Genomes

On sequencing the genome of *B. aphidicola* BCc, endosymbiont of the cedar aphid, we revealed its size to be just 416 kb, some 200 kb smaller than the first three sequenced (Pérez-Brocal et al. 2006). At that time, it was the smallest microbial genome to have been sequenced. Comparison with other *Buchnera* disclosed that genome reduction was exclusively due to the loss of protein-coding genes (362 in BCc compared with 571, 559, and 507 in BAp, BSg, and BBp, respectively). Furthermore, functional genome analysis has revealed that with just 362 protein-coding genes, the BCc genome represents a minimum gene set capable of supporting cell life, because it retains all the genes needed for its own replication, transcription, and translation, as well as a simplified metabolic network for energy production. Therefore, it can be considered an autonomous cell within a particular intracellular environment. However, we cannot say the same concerning its role in symbiosis, as BCc has lost its ability to synthesize tryptophan, an essential amino acid for both the cedar aphid and *Buchnera* itself. In certain *Buchnera* strains, the first two genes of the tryptophan biosynthetic pathway, which code for anthranilate synthetase, are found in plasmids, whereas the remaining genes for the pathway are in the main chromosome. In *C. cedri*, there is a plasmid with both genes, but the genes have been lost from the main chromosome; therefore, BCc cannot synthesize tryptophan.

Electron microscopy studies have shed light on this intriguing puzzle by revealing that *Buchnera* is not alone inside the cedar aphid. Indeed, there is a second kind of

bacteriocyte, where another endosymbiotic bacterium called *Serratia symbiotica* resides, coexisting with *Buchnera* and forming a bacterial consortium. Sequencing of the *Serratia* genome has revealed that it is unable to synthesize tryptophan alone; however, its chromosome holds the genes of the second part of the pathway that *Buchnera* has lost. These results are an example of metabolic complementation. *Buchnera* synthesizes anthranilate synthase, which must then enter *Serratia*, where the remaining genes of the tryptophan synthesis pathway are found, essential to the system as a whole—*Buchnera*, the aphid, and *Serratia* itself Lamelas et al. (2011b).

Other natural minimal genomes have been described, such as the bacteria S*ulcia muelleri* and *Carsonella ruddii*, endosymbionts of cicada and psyllids, respectively, with even smaller genome sizes than that of BCc (246 and 160 kb, respectively). The case of *Sulcia* can be considered similar to that of BCc, as it is not alone in the cicada but forms a bacterial consortium with the bacterium *B. cicadellinicola* MacCutcheon et al. (2009). The analysis of both genomes has also revealed complementary biosynthetic abilities. Therefore, while *Baumannia* provides the system with vitamins, *Sulcia* codes for enzymes involved in the biosynthesis of most amino acids. Undoubtedly, these two bacteria have enabled both insects to adopt a new lifestyle dependent on plant xylem sap.

From Endosymbionts to Organelles

The case of *Carsonella* is intriguing, because researchers have been unable to find a second bacterium in psyllids, as they have in the instances above (Nakabachi 2006). Nonetheless, *Carsonella* cannot be considered a mutualistic endosymbiont (Tamames et al. 2007), having lost its ability to synthesize several essential amino acids (histidine, phenylalanine, and tryptophan). Additionally, it cannot even be considered a living organism, as it has lost the essential life-defining functions, like some genes involved in transcription and replication. The question we should ask is: Is *Carsonella* on the way to becoming an organelle? So far the only organelles to have evolved from free-living ancestors through a process of endosymbiosis are mitochondria and chloroplasts. For this to happen, gene transfer to the host cell nucleus must occur, with the necessary acquisition of sophisticated machinery to import gene products encoded in the nucleus. So far, all attempts to find *Carsonella*'s lost genes in the psyllid nucleus have failed. Sequencing of its genome may solve this riddle and, should it be proven, would be the first known case of a minimal cell becoming organelle.

Another "symbiont versus organelle" debate concerns the chromatophores of the amoeba *Paulinella chromatophora*, which were discovered in 1894 as two kidney-like photosynthetic bodies. Later, they were found to be related to cyanobacteria rather than to chloroplasts in algae and plants and thus may represent a second and independent source of photosynthetic organs, which have come about through a process of symbiogenesis. We had to wait, however, until 2008 for its genome to be sequenced, which disclosed that it is only 1.02 Mb in size and codes 867

proteins (Nowack et al. 2008). Therefore, it represents the minimal cyanobacterial genome, which would suggest that, like other endosymbionts, it has undergone drastic genome reduction. It contains a full repertoire of photosynthetic genes but, like *Carsonella*, has lost genes involved in essential cell functions. The data characterize the chromatophore as a photosynthetic entity that is fully dependent on the ameba for its survival and growth. They are, therefore, the only known descendants of cyanobacteria, in addition to plastids, with a significantly reduced genome that confer photosynthesis to their eukaryotic host. Their comparison with plastids and with other bacterial endosymbionts of invertebrates will shed light on the first steps in the process whereby a photosynthetic prokaryote becomes incorporated into a eukaryotic cell.

What Does the Future Hold?

The advent of bioinformatics and new sequencing techniques has opened new perspectives in the study of symbiosis. Now, thanks to metagenomics, which allows culture-independent analysis of microbial communities growing in their natural environments, we can approach the study of symbiosis in complex communities (McFall-Ngai 2008). This applies to the human microbiota, microorganisms living on and inside humans, which exceed the somatic and germ cells by a factor of 10 (it has been suggested that microorganisms account for approximately 2 kg of our bodyweight).This microbiota resides in different habitats, such as mouth, skin, colon, vagina, stomach, and intestine. Therefore, humans can be considered as "super-organisms" whose metabolism represents the interplay of microbial and human attributes. There are, therefore, two main components of the human body: "the brain and the microbiome," which remain a great mystery compared with other organs, such as the heart or kidney. Take the case of the human intestinal microbiota, which though conceived in our collective imagination as commensals or pathogens, we now know are possibly mutualists in much of their association with the host. Millions of bacteria belonging to hundreds of different species inhabit our gut; having a positive effect on our health as they help digest food, modulate the immune system, and aid in the fight against harmful bacteria by producing antimicrobial factors. By gaining insight into how healthy intestinal microbiota develops, we can produce food and dietary supplements containing the pertinent probiotics, in other words, bacteria that help improve our intestinal health.

Acknowledgments This work has been funded by grants BFU2009-12895-CO2-01; BFU2012-39816-C02-01 and SAF2009-13032-C02-01 from Ministerio de Ciencia e Innovación (MICINN, Spain) and Prometeo/2009/092 from Conselleria de Generalitat Valenciana, Spain.

References

Gosalbes MJ, Latorre A, Lamelas A, Moya A (2010) Genomics of intracellular symbionts in insects. Int J Med Microbiol 300(5):271–278

Lamelas A, Gosalbes MJ, Moya A, Latorre A (2011a) New clues about the evolutionary history of metabolic losses in bacterial endosymbionts, provided by the genome of Buchnera aphidicola from the aphid Cinara tujafilina. Appl Environ Microbiol 77:4446–4454

Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A (2011b) Serratia symbiotica from the Aphid Cinara cedri: a Missing Link from Facultative to Obligate Insect Endosymbiont PLoS Genet 7:e1002357

McCutcheon J, McDonald BR, Moran NA (2009) Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci USA 106:15394–15399

McFall-Ngai M (2008) Are biologists in 'future shock'? Symbiosis integrates biology across domains. Nat Rev Microbiol 6:789–792

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42:165–190

Moya A, Peretó J, Gil R, Latorre A (2008) Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9:218–229

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science 314:267

Nowack EC, Melknonian M, Glöckner G (2008) Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes. Curr Biol 18:410–418

Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A (2006) A small microbial genome: the end of a long symbiotic relationship? Science 314:312–313

Tamames J, Gil R, Latorre A, Peretó J, Silva FJ, Moya A (2007) The frontier between cell and organelle: genome analysis of *Candidatus* Carsonella ruddii. BMC Evol Biol 7:181

Part III
Early Eukaryotes

Chapter 7
The Evolutionary Origin of Animals and Fungi

Sandra Baldauf, Maria Romeralo and Martin Carr

Introduction

Our evolutionary trees are full of holes, that is, they often have many long branches with just a few twigs at the ends. To a certain extent, we expect this since the vast majority of all species that ever lived are extinct (Rhode and Muller 2005). Some of those long unbroken branches are also the result of evolutionary rate acceleration, because some lineages evolve faster than others. For example, parasite genes tend to evolve faster than those of their free-living relatives and isolated populations have faster evolutionary rates (Ohta 1972). However, a shameful number of the holes in our trees are due to missing data from extant species (Purvis and Hector 2000). For example, it is estimated that less than 5 % of all extant fungi have been described and there are similar estimates for many other groups (McLaughlin et al. 2009).

The extent of protist diversity yet to be discovered is a matter of continued, and often heated, debate (O'Malley 2007; Medlin 2007; Telford et al. 2006). Still, regardless of their current diversity, protists account for the vast majority of eukaryotic evolutionary history and roughly two-thirds of the major divisions of eukaryotes consist entirely of single-celled organisms (Fig. 7.1). Progress characterizing these groups is slow. In fact, there are over 200 unique types of protists in culture collections that have never been characterized in molecular terms, not even by cheap and cheerful small subunit ribosomal RNA gene (18S or SSU rDNA) sequencing, the cornerstone of molecular taxonomy. In 1999, Patterson listed 239 unclassified

S. Baldauf (✉) · M. Romeralo
Department of Systematic Biology, Evolutionary Biology Centre,
Uppsala University, Uppsala, Sweden
e-mail: sandra.baldauf@ebc.uu.se

M. Romeralo
e-mail: maria.romeralo@gmail.com

M. Carr
Department of Biology, University of York, Heslington, York, UK
e-mail: mc528@york.ac.uk

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_7, © Springer Science+Business Media New York 2013

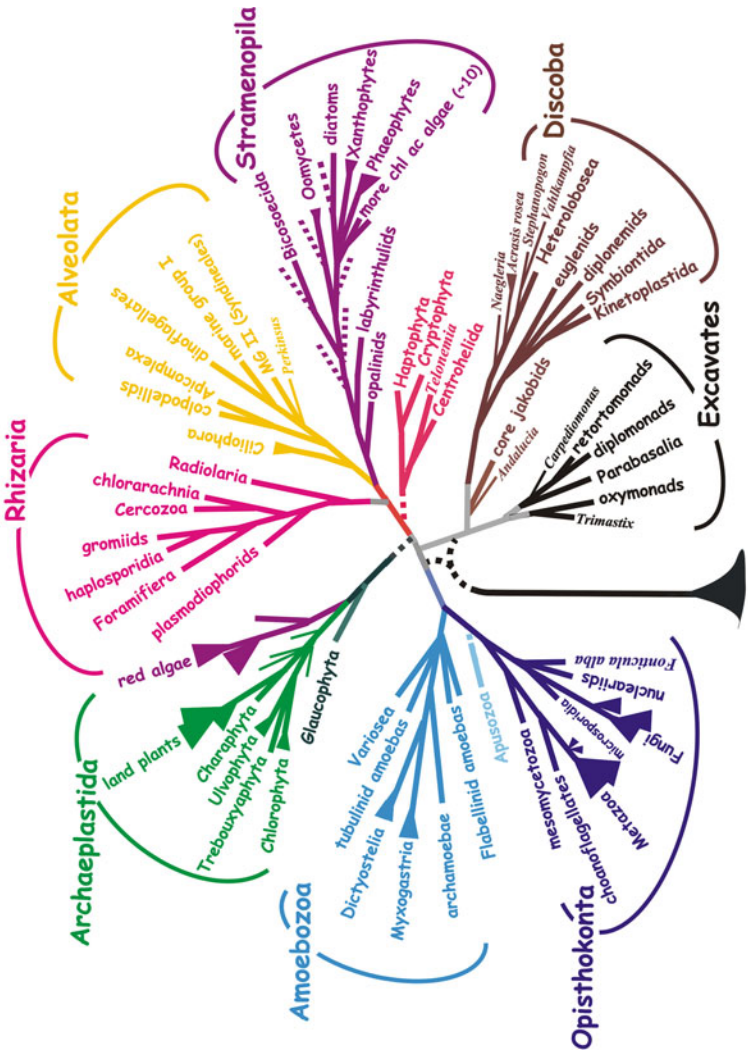


Fig. 7.1 A consensus phylogeny of eukaryotes. The tree shown is an updated version of a previously published consensus phylogeny (Baldauf 2003). Revisions are based on recently published global multigene phylogenies of eukaryotes (Hackett et al. 2007; Burki et al. 2008; Parfreyet al. 2010), and group-specific phylogenies of Alveolata and Stramenopila (Reeb et al. 2009), Amoebozoa and Rhizaria (Pawlowski and Burki 2009), Discoba (Lara et al. 2006), Excavata (Kolisko et al. 2008), and Opisthokonta (see Figs. 7.2 and 7.3 in the following sections). *Thin lines* indicate single species of important taxonomic position. *Filled triangles* indicate multicellular groups. *Dashed lines* indicate environmental clades of Stramenopiles (MAST clades, Massana et al. 2006)

major protist types in culture collections (Patterson 1999), which six years later had only been reduced to 204 (Adl et al. 2005). Of those that have been characterized, some are turning out to represent new major groups such as Apusozoa (Kim et al. 2006) and Centrohelids (Sakaguchi et al. 2007). Meanwhile, metagenomic studies indicate this is only the tip of an iceberg (Not et al. 2009; Marande et al. 2009). Even among the known and "characterized" species or "morphotypes", only a single isolate has been examined for many of them, so in most cases we have not even begun to assess the level of cryptic species.

More unfortunately still, long unbroken branches are major obstacles in building accurate evolutionary trees (Hendy and Penny 1989). Not too surprisingly, long branches tend to be difficult to resolve. Something that is highly derived, simply bears no clear similarity to anything, making it very difficult to determine what is and is not related to. Therefore, long branches tend to be unstable "drifting" branches. This problem of "rogue taxa" is not necessarily confined just to the unstable branch, but can destabilize an entire branch "neighborhood" by alternately invading various nearby clades. This sabotages statistical support values that rely on strictly monophyletic groups, such as the standard nonparametric bootstrapping (Sanderson 1989; Baldauf and Palmer 1993). The problem becomes rapidly worse if there is more than one long branch in a tree, because long branches tend to attract each other (Felsenstein 1978; Felsenstein 2004; Bergsten 2005). This long branch attraction artifact is basically a missing data problem. If we could fill in some of the blanks in evolutionary time, that is, if we could break up long branches with intermediate lineages that bridge the gap between them and their closest relatives, the effect would be to stabilize the branches thereby making them easier to resolve (Hillis 1996; Bergsten 2005).

Not too surprising, the long branch problem becomes worse as we go deeper in evolutionary trees. This is probably due to a combination of many things including less character data, more missing taxa, and increasing amounts of mutational saturation and parallel evolution (homoplasy). However, primarily due to a number of large-scale collaborative sequencing projects, we are beginning to flesh out some of these deeper branches. One group of broad interest is animals and fungi where many of the deepest major branches have only recently been recognized, much less resolved. In the following section, we will try to give a broad overview of the current state of knowledge on the identity and characteristics of the specific protistan relatives of Metazoa and Fungi as well as an introduction to the closest known opisthokont sister groups, the Amoebozoa.

Opisthokonta

A large body of molecular data has clearly shown that animals and fungi are close relatives (Baldauf and Palmer 1993; Wainright et al. 1993), now including all large multigene trees (e.g., Ciccarelli et al. 2006; Burki et al. 2007; Hackett et al. 2007; Hampl et al. 2009). In fact, various single-celled "protist" lineages have recently been added to the evolutionary lines leading directly to animals and fungi. Metazoa is now recognized to be part of a larger lineage, the Holozoa, which also includes

choanoflagellates (Fig. 7.2), ichthyosporeans, *Capsaspora owczarzaki*, *Corallochytrium limacisporum,* and *Ministeria vibrans* (Fig. 7.3) (Carr et al. 2009; Steenkamp et al. 2006). Likewise, Fungi are now seen to be part of a larger group, the Holomycota, which also includes the nucleariid amoebas (Zettler et al. 2001; Steenkamp et al. 2006) and the former slime mold, *Fonticula alba* (Brown et al. 2010).

Nonetheless, acceptance of the integrity of opisthokonts has been something of an uphill battle. This may be, at least in part, because of the almost complete lack of morphological justification for the group. Three defining characters have been proposed: a single basal flagellum on reproductive cells (from which the group gets its name), flattened mitochondrial cristae (Cavalier-Smith 1987), and an insertion in the translation elongation factor protein, EF-1A (Baldauf and Palmer 1993; Steenkamp et al. 2006). Cavalier-Smith noted that most eukaryotes have two or more flagella, while animal and fungal cells, when flagellate, tend to have only a single posteriorly directed flagellum, particularly on their reproductive cells (metazoan sperm and Chytrid zoospores; Stechmann and Cavalier-Smith 2002). Patterson noted that animals and fungi usually have flattened mitochondrial cristae, while most other eukaryotes had tubular or flattened plate-like (discoidal) cristae (Patterson 1988). Baldauf and Palmer identified a 9–17 amino acid insertion in the otherwise highly conserved EF1A protein (Baldauf and Palmer 1993; Steenkamp et al. 2006). However, none of these are universal characters among opisthokonts. Flagella were lost multiple times in Fungi (James et al. 2006a), as well as in nucleariids, ichthyosporeans, *Capsaspora*, *Corallochytrea*, and possibly *Ministeria*. Mitochondrial cristae morphology appears to be plastic in some groups, and lamellar, tubular, and discoidal cristae are all found in nucleariids (Zettler et al. 2001; Ragan et al. 1998). Although, the EF1A insertion is still found only in opisthokont EF1A, EF1A itself is not universally present in holozoan protists (Gile et al. 2009). Nonetheless, these characteristics are widespread across the opisthokonts, especially among the opisthkont protists, suggesting that they represent the ancestral state of the group.

Holozoan Protists

The Holozoa includes Metazoa and their single celled "allies"—choanoflagellates, Ichthyosporea and a number of "enigmatic" unplaced taxa currently represented by only a single examined species: *Ministeria vibrans, Corallochytrium limacisporum,* and *Capsaspora owczarzaki* (Fig. 7.2). There are two described ministeriids, but only *Ministeria vibrans* remains in culture. Corallochytrids are saprotrophs with little morphology to judge them by, and *Corallochytrium limacisporum* is the only species of corallochytrid in culture. *Capsaspora owczarzaki* is a morphologically complex parasite for which a number of conflicting phylogenetic positions have been proposed (e.g., Shalchian-Tabrizi et al. 2008 versus Ruiz-Trillo et al. 2008). For convenience more than anything else, all these taxa except choanoflagellates will be treated together here as "Mesomycetozoa", although this name has had a variety of definitions in the past, none of which are currently entirely valid (Mendoza et al.

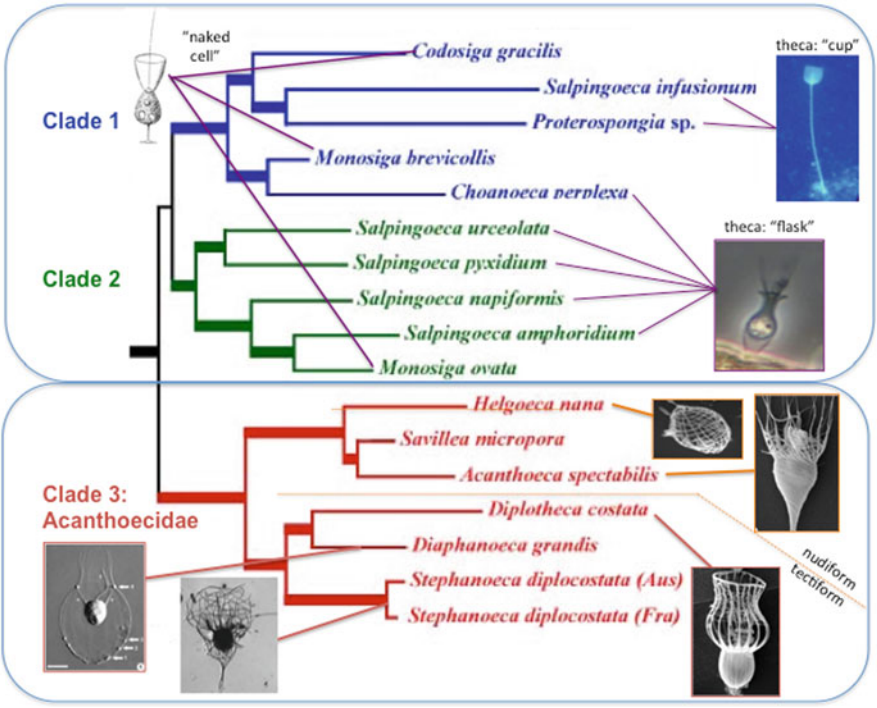


Fig. 7.2 Molecular phylogeny of choanoflagellates and some example morphotypes. The figure shown is adapted from a multigene phylogeny of choanoflagellates (Carr et al. 2009). Micrograph images kindly provided by B.S.C. Leadbeater

2002; Adl et al. 2005). Regardless of the exact composition of Mesomycetozoa, choanoflagellates still consistently appear to be more closely related to Metazoa than any of the proposed Mesomycetozoa. The latter are also mostly parasites with little morphology to distinguish them, while choanoflagellates are free-living cells, exhibiting considerable morphological diversity and with a long history of study. They are also an important group of aquatic microbes (Leadbeater 2008b). However, it is as the sister group to Metazoa that they hold the most evolutionary interest, because they may provide insight into the starting material from which Metazoa evolved.

Choanoflagellates

History, Morphology, Ecology

The choanoflagellates are a widespread group of nano-scale aquatic microbes, possibly the predominant bacterial predators in many aquatic ecosystems (Leadbeater 2008b). There are approximately 300 species in the current choanoflagellate database

(Leadbeater, personal communication), of which roughly 50 occur in fresh water while the rest are found in brackish or marine ecosystems (Leadbeater and Karpov 2000). The majority of species appear to have a global distribution, while others may be very restricted (Thomsen et al. 1991; Nitsche and Arndt 2008). However, molecular sequencing suggests that at least some globally occurring morphotypes may in fact be different species (Carr et al. 2009). Most of the early work on choanoflagellates is that of James-Clark and Saville-Kent. The first unequivocal description of a choanoflagellate was by James-Clark in 1867. Saville-Kent also described many species and was the first to classify them into a new taxonomic order, Choanoflagellata, and to erect the families Codonosigidae and Salpingoecidae (Saville-Kent 1880). These families differ both in morphology and mechanism of cell division. In terms of morphology, salpingoecids have thecae while codonosigids are "naked". In terms of cell division, codonosigids divide laterally while salpingoecids cannot divide in situ because of their rigid theca, so they first become amoeboid and partly exit the theca before cell division (Saville-Kent 1880). The third family, the Acanthoecidae, consisting of species with a silica lorica, was not recognized as such until the work of Norris nearly 100 years later (Norris 1965).

All known choanoflagellates share a uniform and distinctive cell body (protoplast) morphology consisting of a simple ovoid or spherical cell with a single flagellum surrounded by a ring of actin-based microvilli (Karpov and Leadbeater 1997; Leadbeater 2008b). Under light microscopy, the beating microvilli appear to form a single continuous collar-like structure, hence the common name of "collared flagellates". As choanoflagellates are filter feeders, the primary function of the flagellum in trophic cells is to draw water upward and across the outer surface of the microvilli, which capture bacterial cells and other food particles, possibly by a lectin-like mechanism (Zubkov et al. 2006). Food particles are then transported externally along the microvillar surface to the cell body where they are engulfed by pseudopodia that arise at the base of the collar (Leadbeater 1983; Boenigk and Arndt 2000). This process is much more efficient if the cell body is anchored so that the flagellar force is not dissipated in moving the cell (Pettitt et al. 2002). Devising means to anchor the cell during the trophic stage of the life cycle seems to have been a major driving force in choanoflagellate evolution, and this is often achieved by the outer cell covering, or periplast (Leadbeater 2008b).

All choanoflagellate protoplasts have an outer covering consisting of a thin glycocalyx encasing the anterior (non-flagellate) end of the cell in what is essentially a fine extracellular cup (Leadbeater 2008b). This glycocalyx can also extend beyond the posterior of the cell and join with a stalk of carbohydrate microfibrils or peduncle to attach the cell to a substratum (Leadbeater and Morton 1974; Leadbeater 2008b). The glycocalyx is a flexible structure, and, in the absence of further physical restriction, lateral cell division can occur within it. Following division, the daughter cell may dislodge and become a "swarmer" cell, swimming away by means of its flagellum. Alternatively, the juvenile cell may remain attached to the parental cell or stalk, resulting in a cluster of cells or colony (see the following section, Leadbeater 2008b).

The majority of described choanoflagellates possess additional periplast structures external to the glycocalyx. These occur as two main types, either a thick fibrillar organic theca or a basket-like lorica composed of silica strips. Thecae are complex multilayered structures composed of carbohydrate microfibrills. They may take the form of a simple open cup, a flask with a constricted neck region or, rarely, a tube (Leadbeater 2008b). The rigidity of the theca has important consequences for cell division; thecate cells cannot divide in situ due to the physical restriction of the rigid tightly fitting theca. Instead the parent cell first converts to an amoeboid morphology and crawls part way out of the theca before dividing. The parent cell then returns to the theca and regains its flagellum while the daughter cell forms its own flagellum and swims away with only its glycocalyx as covering. These swarmer cells remain naked until they settle down and become attached, at which point they form their own theca (Leadbeater 2008b).

Loricate choanoflagellates surround themselves with a complex inorganic basket-like structure. All loricae are built with a regular number of ribs (costae), each of which in turn is composed of a fixed number of partially overlapping silica strips. These strips are held together end to end and tightly locked in place by an unknown mechanism that survives even harsh cell fixation regimes. All loricae have a similar basic organization consisting of two layers of costae; an external layer of longitudinal ribs is held in position by an internal layer of horizontal ribs, the latter arranged either in rings or helices (Leadbeater 2008b). Lorica construction is a fascinating process dissected and described in detail by Leadbeater and Cheng (2010). First, a complete set of silica strips are presynthesized within the cell in specialized vesicles. The strips are then transported to the cell surface where they are released and stored in firewood-like bundles. Once a full complement of strips has been produced, the lorica is constructed quickly by a single coordinated series of movements in which the microvilli draw the longitudinal ribs up, which in turn drag the horizontal ribs with them. At the same time, the ring of lorica-assembling microvilli rotate between one and four turns (according to species), which causes the internal ribs to spiral (Leadbeater and Chang 2010).

In addition to the costae, loricae frequently contain organic microfibrils arranged between the strips. The result is an enclosed column capable of funneling large volumes of water, which is important in the open water where food particles may be widely dispersed. These loricae can reach macroscopic proportions, sometimes completely dwarfing the single choanoflagellate cell residing deep within it (Thomsen et al. 1990). Some loricate cells are sedentary and attach themselves to the substratum by a peduncle that extends from the lorica. However, the majority of the 150 plus described loricate species are pelagic, and these are the only choanoflagellates to have conquered the open ocean (Leadbeater 2008b).

Loricate choanoflagellates can be separated into two basic types, which also correspond to two different reproductive strategies, nudiform and tectiform (Leadbeater et al. 2008a, 2008b, 2009; Leadbeater 2010). Cell division in nudiform species results in naked swimming dispersal cells, which rapidly settle to produce a new lorica. In tectiform species, the entire set of lorica strips is prefabricated by the parent cell and stored at the top of the collar. When the cell divides, the daughter that will leave the

parent lorica is inverted and takes with it the bundles of strips as it exits the parent lorica and immediately assembles them into a new lorica of its own. Thus, tectiform species are never naked and there is no free-swimming juvenile dispersal stage (Leadbeater 2010). The loricae of nudiform species comprise longitudinal and helical horizontal costae, the latter of which tend to be dense structures with many tightly packed turns. Tectiform species on the other hand have lorica with rings (as well as limited helices in some taxa), requiring fewer costae and thus resulting in a lighter structure. The tectiform plan seems to have been highly successful as there are less than six described nudiform species but over 130 tectiform ones (Leadbeater 2010).

Loricate choanoflagellates are exclusively marine (Leakey et al. 2002), but only the tectiforms are pelagic. This may be related to the lightness of the ringed lorica and the fact that the daugher cell's lorica is presynthesized by the parent cell. Thus, all tectiform cells are never without a lorica at any point in their cell cycle. The function of the lorica seems to be both to aid buoyancy and to provide resistance to the propulsive effects of flagellar motion thereby enhancing the efficiency of water flow for feeding purposes. Nudiform species mainly inhabit specialized ecological niches such as bacterial biofilms. In contrast, the tectiforms, which are the most speciose division of the choanoflagellates, inhabit a broad range of habitats. Tectiforms are also the only choanoflagellates that do not lead sedentary lives; many are free-floating, with some present in the water column while others are benthic and drift in water currents (Leadbeater and Chang 2010; Carr et al. 2009).

Much has been made of the fact that the choanoflagellate protoplast bears a striking resemblance to the collar cells of sponges. The collar cells of sponges line the inner surface of the sponge with their flagellar ends protruding into the luminal space from which they capture food particles via their microvillar collar. James-Clark (1867) and Saville-Kent (1880) both noted this similarity, and Saville-Kent, who described the first colonial choanoflagellates, erected the genus *Protospongia* (later amended to *Proterospongia*) for these forms (Saville-Kent 1880). Choanoflagellate colonies may be uniseriate (e.g., *Desmarella* and *Kentrosiga*) or star-like (e.g., *Astrosiga*) assemblages on a substratum or free-floating, as well as globules of cells (e.g., *Proterospongia* and *Sphaeroeca*). Free-floating colonial cells remain attached to each other after division and are essentially anchored to each other so that once again the flagellum moves the medium and not the cells. Globate colonies are also held together by a gelatinous matrix, which resembles the mesohyl of some sponges (Buss 1987). These similarities have led to a long-held speculation that *Proterospongia* could be an evolutionary intermediate between choanoflagellates and Metazoa.

Molecular Phylogeny

Traditional choanoflagellate classification is based on periplast morphology and recognizes three families. The Codonosigidae, or naked cells, consist of all species with only the thin glycocalyx for a cell covering. The Salpingoecidae include all species with thecae, and can be further divided into those with cup, flask or tube-shaped thecae. The Acanthoecidae include all loricate species, which can be separated

into nudiform and tectiform types (Leadbeater 2008a). The first taxonomically broad molecular phylogeny of choanoflagellates rejected most of this except for the Acanthoecidae and the nudiform and tectiform subgroups within it (Fig. 7.2; Carr et al. 2009). Instead, the molecular phylogeny identifies three major clades, one corresponding to the traditional Acanthoecidae and two major clades consisting of a mixture of naked species, species with cup- and flask-shaped thecae and colony-forming species (Fig. 7.2). Thus, the molecular phylogeny rejects the traditional separation of species with organic periplasts into Salpingoecidae and Codonosigidae as well as the genus *Proterospongia*. In fact, *Proterospongia choanojuncta* is molecularly identical to the thecate species *Choanoeca perplexa* for both SSU and LSU (Carr et al. 2009 Medina et al. 2001). Thus, molecular data confirm morphological observations suggesting that colony formation is a stage in the life cycle of this, and possibly all non-loricate species (Leadbeater 2008b).

One striking example of this mixture of morphotypes is the "genus" *Monosiga.* The two molecular model systems and morphologically nearly indistinguishable species *Monosiga ovata* and *M. brevicollis* are found in molecular clades 1 and 2, respectively (Fig. 7.2). The main difference between them is that *M. ovata* is a freshwater species, while the isolate of *M. brevicollis* that has been sequenced is exclusively marine (King et al. 2008). This is consistent with observations from other groups such as dinoflagellates suggesting that the fresh water-marine transition is rare among protists (Logares et al. 2009). In fact, the current molecular sampling of choanoflagellates indicates only a single invasion of fresh water (Fig. 7.3). However, only a few of the more than 50 described fresh water choanoflagellate species have been examined with molecular data, so this issue is far from resolved for this group. Nonetheless, the current broad molecular sampling of choanoflagellates strongly suggests that this was primitively a marine species that probably invaded fresh water rarely (Carr et al. 2009).

Examination of a matrix of morphological characters shows no simple morphological justification for either choanoflagellate major group 1 or group 2 (Carr et al. 2009). Instead, a mixture of thecate and non-thecate species is found in both clades, including both cup and flask-shaped thecae (there are no sequences yet from species with tube shaped thecae; Fig. 7.2). Since the theca is a complex layered structure, and the combination of periplast structures of Clade 1 and Clade 2 thecae are similar (Leadbeater 2008b), it is most parsimonious to assume that the theca was already present in the last common ancestor of Clades 1 and 2. This theca was then lost or altered on numerous occasions, evolving into a cup-shaped theca at least twice, once during the evolution of both major groups 1 and 2 (Fig. 7.2; Carr et al. 2009). However, the situation is probably even more complex, as many thecate species, including those with tube-shaped thecae are yet to be examined with molecular data.

The taxonomic position of the choanoflagellates as a whole has been the source of much debate and speculation. They have been proposed to be the ancestors of Metazoa (James-Clark 1867; Saville-Kent 1880), highly reduced true Metazoans (Maldonado 2004), algae (Bourrelly 1968; Chadefaud 1960), or paraphyletic ancestors of both Fungi and Metazoa (Cavalier-Smith 1987). Molecular phylogeny now clearly shows that choanoflagellates are monophyletic holozoans (Carr et al. 2009)

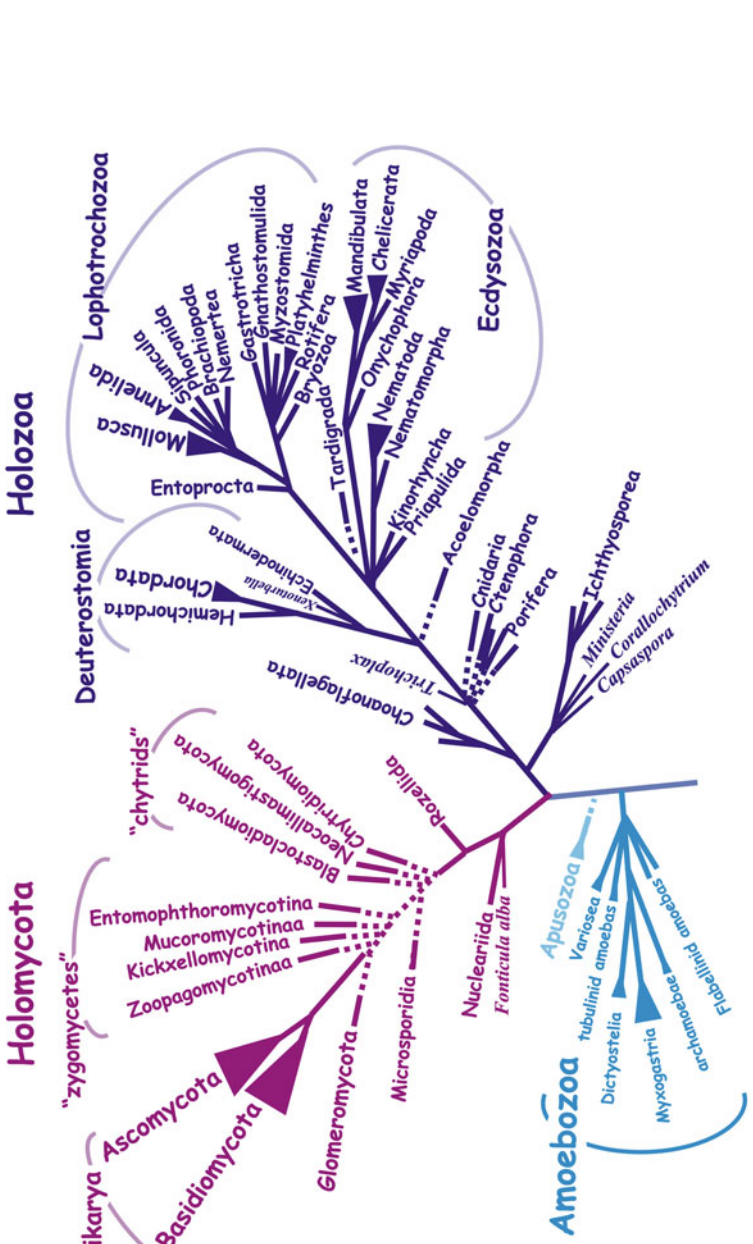


Fig. 7.3 A consensus phylogeny of the major divisions of Opisthokonta. The tree shown is a general consensus of recently published multigene phylogenies of Fungi (James et al. 2006a, b; Hibbett et al. 2007), Metazoa (Dunn et al. 2008; Bourlat et al. 2008; Philippe et al. 2009; Hejnol et al. 2009; Pick et al. 2010; Rota-Stabelli et al. 2010), and Holozoa (Carr et al. 2009; Marshall et al. 2008; Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008) and single gene trees for Rozellida (Lara et al. 2010) and *Fonticula alba* (Brown et al. 2009)

and currently the closest known sister-group to Metazoa (Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008). Thus, there is no suggestion of a specific link between Metazoa and any subset of choanoflagellates, such as "proterospongia", and thus no evidence that Metazoa evolved from a true choanoflagellate ancestor. In fact, it appears that the ability to form colonies may be an ancestral trait of choanoflagellates.

One important implication of the close link between Metazoa and choanoflagellates is the possibility that the latter may help us understand some of the fundamental processes involved in the evolution of Metazoa and their unique and highly complex form of multicellularity. Large scale sampling of messenger RNA sequences (expressed sequence tag, or EST sequencing) from *Monosiga brevicollis*, *M. ovata*, and *Salpingoeca rosetta* (formerly *Proterospongia* sp. 50818), together with genome sequencing of *M. brevicollis* (King et al. 2008) have revealed the presence of a number of molecular functions previously considered unique to Metazoa. These include genes for proteins such as cadherins, receptor tyrosine kinases, and C-type lectins (King and Carroll 2001; King et al. 2003). These are important for cell–cell signaling and adhesion in Metazoa. However the roles of these proteins in choanoflagellates are yet to be determined. *M. ovata* also appears to encode a distant relative of the c-terminal domain of the Hedgehog proteins, which are important in cell–cell signaling during metazoan development (Snell et al. 2006). In choanoflagellates, this "Hoglet" protein is predicted to be autocatalytic, like as true Hedgehog proteins but in this case releasing a unique N-terminal domain with a predicted cellulose binding function that may allow it to temporarily tether the cell to plant matter (Snell et al. 2006). The available molecular data from choanoflagellates are expanding rapidly, especially with the recent completion of the *M. brevicolis* (King et al. 2008) and *Salpingoeca rosetta* (http://www.broadinstitute.org/science/data) genome sequences and plans to sequence additional species (www.genome.gov/10002154).

Mesomycetozoa

Overview

Mesomycetozoa are a diverse group of mostly parasitic protists with little common morphology to judge them by. Many have been known for some time, but were originally or eventually placed with trichomycete fungi because many of them form simple hair-like sporangia on their hosts. Other species have, at one time or another, been placed with green algae, rhizopods, nucleariids, stramenopiles or simply *incertae sedis*, although not necessarily with much confidence. As morphologically 'simple' parasites, these species have been generally classified according to their lifestyles (parasites vs. saprotrophs) or hosts (plants, vertebrates or invertebrates). However, molecular phylogeny indicates that this is not a reliable means of classification for these organisms. As defined here, the Mesomycetozoa consist of one main group, the Ichthyosporea, plus three "oddball" taxa, *Capsaspora owczarzaki*, *Ministeria vibrans*, and *Corallochytrium limacisporum* (Fig. 7.2).

While it is now clear that all members of the Mesomycetozoa as defined here are Holozoa, their precise relationships to each other and to the remaining Holozoa are still uncertain. This is at least partly because there is substantial molecular data for only a handful of them and none at all for most. In fact, only three species have been subjected to extensive molecular sequencing: *Amoebidium parasiticum*, which was the first ichthyosporean to be axenically cultured (Whisler 1960), *Ministeria vibrans*, because early data suggested it could be the sister group to Metazoa (Cavalier-Smith and Chao 2003; Steenkamp et al. 2006), and *Capsaspora owczarzaki*, because appeared to represent a unique early diverging lineage (Ruiz-Trillo et al. 2004). Thus, it should not be too surprising that relationships among these species remain difficult to resolve. Still, the most recent analyses of large molecular data sets with sufficient outgroup sequences indicate that choanoflagellates are a closer sister group to Metazoa than any examined mesomycetozoan so far (Ruiz-Trillo et al. 2008; Shalchian-Tabrizi et al. 2008).

Ichthyosporea

This is an ecologically and morphologically diverse group of predominantly parasitic protists. Spanggaard et al. (1996) and Ragan et al. (1998) were the first to sequence SSU rDNA from them and thus notice that they were related to known opisthokonts and to each other. Ragan named them DRIPs, an acronym of the first known members' names (*Dermocystidium*, Rosette agent of fish, *Ichthyophonus* and *Psorospermium*). To date, there is no known morphological, ultrastructural, or ecological justification for the group, although they all have flattened mitochondrial cristae with the possible exception of *Ichthyophonus hoferi* (Ragan et al. 1998). They are mostly parasites, but there are also symbionts and saprotrophs (Mendoza et al. 2002). The diversity of forms is particularly striking including various mixtures of flagellate, amoeboid, cyst, sporangial, and/or hyphal stages (Mendoza et al. 2002). The Ichthyosporea appear to consist of three major divisions—rhinosporids, ichthyophonids (Mendoza et al. 2002; Cafaro 2005), and aphelids, although there is still no molecular data from aphelids to confirm this.

Rhinosporideacae (formerly Dermocystida) are mainly parasites of fish, mammals, and birds (Mendoza et al. 2002) and have been previously misidentified as fungi, apicomplexans, haplosporideans, and even cyanobacteria (Ahluwalia et al. 1997; Mendoza et al. 2002). Their distinguishing feature within Ichthyosporea is the presence of flagellated cells, which most species use to infect their hosts. Once inside the host, the flagellum is lost and the cell transforms into a walled sporangium, reaching 200–400 μm in diameter and containing thousands of zoospores. These are eventually released into the environment to infect new hosts and the cycle begins again (Herr et al. 1999; Mendoza et al. 2002). The known host range of rhinosporidians is so far restricted to vertebrates, mainly fish but also amphibians (Feldman et al. 2005; Jay and Pohley 1981) and sometimes humans (Herr et al. 1999). In fish, they can infect a variety of tissues including the gills, kidney, and spleen, sometimes causing highly virulent infections of economical significance (Herr et al. 1999).

Ichthyophonids are also mainly parasites, although there are also commensals and possibly a few saprotrophs (Cafaro 2005; Kimura et al. 2002; Mendoza et al. 2002). Their main distinguishing feature within Ichthyosporea is that they possess amoeboid rather than flagellated dispersal cells (Mendoza et al. 2002), although flagellated zoospores have been reported in *Pseudoperkinsus tapetis* (Ordás and Figueras 1998). Ichthyophonids have a wider range of known hosts than rhinosporidians including arthropods, molluscs, fish, amphibians, reptiles, and birds (Baker et al. 1999; Herman 1984; Mendoza et al. 2002; Rand 1994; Spanggaard et al. 1996). However, the best-studied ichthyophonid, *Amoebidium parasiticum*, is a symbiont that attaches itself to the external exoskeleton of insects (Benny and O'Donnell 2000). The two suspected saproptrophic ichthyponids are *Sphaeroforma arctica* and isolate LKM51 (Mendoza et al. 2002).

Many ichthyophonids also appear to lack both flagellate and amoeboid stages (Marshall et al. 2008). *Anurofeca richardsi*, originally thought to be a parasitic alga (Kruger 1894), forms spherical cells that mature into a small number (< 10) of simple rounded endospores in the guts of anuran (frog and toad) larvae (Baker et al. 1999). However, as with many ichthyosporeans, the complete life cycle is not known, so the possible presence of a motile stage still cannot be ruled out. Species of *Eccrinidus* infect arthropod hosts, which they do via simple uninucleate free-floating thick-walled spores that are passively ingested by the host (Cafaro 2005). Other species may be very plastic such as *Ichthyophonus hoferi*, which has dispersal cells that can form hyphae at acidic pH but amoeboid cells at higher pH (Baker et al. 1999).

Aphelidea

The Aphelidea are also entirely parasites and were only recently placed in Ichthyosporea because they possess a single, posteriorly directed flagellum and flat mitochondrial cristae (Cavalier-Smith et al. 2004). Nonetheless, unlike most known ichthyosporeans, aphelids are parasites of algae including multicellular green algae and single-celled stramenopile and alveolate algae such as diatoms and dinoflagellates (Fig. 7.1; Schweikert and Schnepf 1996). Thus, these parasites can circumvent the armor plating of testate algae, as do chytrids. However, unlike chytrids, aphelids inject themselves wholly into the host cell by means of a cyst, forming plasmodia that eventually consume the entire cell contents. Dispersal cells can be either flagellated or amoeboid, which encyst upon attachment to the host. Thus they resemble many other ichthyosporeans in forming plasmodia within the host cell, but their phylogenetic position has yet to be confirmed with molecular data.

Capsaspora owczarzaki

This unusual amoeba was long considered to be a tiny (3–7 μm diameter) nucleariid as it forms similar long unbranching filopodia (Hertel et al. 2002). However, Nucleariids are free-living and mainly engulf prey via phagocytosis. *C. owczarzaki*,

on the other hand, is a parasite and it feeds by puncturing its hosts' cells with a long peduncle of unknown composition through which it sucks out the cell contents (Owczarzak et al. 1980). *C. owczarzaki* also lacks the mucoidal extracellular matrix seen in most nucleariids (Adl et al. 2005). *C. owczarzaki* was first discovered as an endosymbiont of the pulmonate snail *Biomphalaria glabrata* (Stibbs et al. 1979), an intermediate host for the human parasite *Schistosoma mansoni*. The protist enters the snail hemolymph where it attacks and kills the trematode sporocysts. This results in the snail being resistant to the parasite, although *C. owczarzaki* may also attack host cells (Stibbs et al. 1979; Owczarzak et al. 1980).

Early molecular analyses of the phylogenetic position of *C. owczarzaki* caused some excitement, as its sequences tended to appear as long isolated branches arising from among the deepest branches of Holozoa (Cavalier-Smith and Chao 2003; Hertel et al. 2002; Ruiz-Trillo et al. 2004). However, there is still only one examined species and long branches are notoriously difficult to place accurately in molecular trees (Gribaldo and Philippe 2002). Meanwhile, more recent multigene trees seem to have produced robustly supported conflicting results. Two of these studies (Burger et al. 2003; Shalchian-Tabrizi et al. 2008) place *C. owczarzaki* together with the enigmatic *Ministeria vibrans* (see the following section) in a separate clade from Ichthyosporea, while other analyses seem to place it together with Ichthyosporea (Hertel et al. 2002; Ruiz-Trillo et al. 2008). Thus, the exact position of *C. owczarzaki* within Holozoa is still something of an enigma.

Ministeria vibrans

Unlike nearly all known ichthyosporeans, this is a free-living holozoan protist. *Ministeria* protoplasts are small ($< 5\ \mu m$) with a very distinct morphology consisting of ~ 20 evenly spaced microvilli radiating in a symmetrical halo around the simple spherical protoplast (Patterson 1999; Adl et al. 2005). The arms resemble the microvilli of choanoflagellates, at least superficially, in that they possess a non-tubulin-like microfilament core and a sticky surface on which they capture food particles (Cavalier-Smith and Chao 2003). However, phylogeny now clearly places *M. vibrans* outside of choanoflagellates, most likely with the Ichthyosporea (Burger et al. 2003; Shalchian-Tabrizi et al. 2008). A second described species, *M. marisola*, is reported to lack a stalk, but this species has been lost from culture and, unlike *M. vibrans*, it has not been seen again (Patterson et al. 1992). The two species are morphologically very similar, except *M. vibrans* anchors itself to the substratum via a long stalk, whereas *M. marisola* is stalkless and unattached. Although neither species appears to have a flagellum, electron microscopic observations suggest that the stiff vibrating stalk of *M. vibrans* could be a degenerated flagellum (Cavalier-Smith and Chao 2003).

Corallochytrium limacisporum

This marine unicell was first isolated from coral reef lagoons in the Indian Ocean (Raghu-Kumar et al. 1987). Protoplasts are spherical and highly variable in size (4.5–20 μm in diameter), and its mitochondrial cristae are flat (Cavalier-Smith and Allsopp 1996). At maturity, up to 32 daughter cells are released through pores in the cell wall. Daughter cells are elongate and amoeboid; no flagellated cells have been observed at any stage (Mendoza et al. 2002). *C. limacisporum* is a saprotroph and has not been found associated with any host organism (Mendoza et al. 2002). Multigene trees tend to place it with Ichthyosporea as a long basal branch (Carr et al. 2009; Steenkamp et al. 2006). The combination of sporangium-formation followed by endospore release in *C. limacisporum* shows some similarity to the life cycle of the ichthyosporean *Rhinosporidium seeberi*. However, *Corralochytrium* endospores are amoeboid while those of *Rhinosporidium* are flagellated.

In summary, the phylogeny of holozoan protists is still unsettled on a number of major points, but important progress has been made. All examined choanoflagellates form a single monophyletic group and this is the closest known sister group to Metazoa (Carr et al. 2009; Ruiz-Trillo et al. 2008). Molecular phylogeny also identifies a substantial sister group to Metazoa + choanoflagellates, the Ichthyosporea. The latter clearly includes both Ichthyophonidae and Rhinosporideacae (Marshall et al. 2008), and there are compelling morphological reasons to include the Aphelidea in this group as well (Cavalier-Smith et al. 2004). However, it may be some time yet before we understand the phylogenetic position of some of the most intriguing holozoan protists, *Capsaspora*, *Corralochytrium*, and *Ministeria*. While genome sequencing is in progress for *Capsaspora* and *Ministeria*, it would also help tremendously, in both molecular and morphological terms, if we could find and examine additional representatives of these lineages.

Protistan Holomycota

Overview

Fungi are by far the least understood of the classic triumverate of multicellular eukaryotes (i.e., plants, animals and fungi). They are generally defined by the presence of a filamentous growth pattern and absorptive nutrition (Cavalier-Smith 1998; Tehler 1988), and more recently and precisely by possession of the diaminopimelic acid pathway for amino acid biosynthesis and cell walls of chitin and often β-glucan (McLaughlin et al. 2009; Fig. 7.2). Unlike Metazoa, the classic definition of Fungi allows the inclusion of a number of unicellular branches, some ancient and some relatively more recently derived such as various yeasts (James et al. 2006a, b). The ancient unicellular fungi were previously all classified together as chytrids. These are tiny often-parasitic unicells that form acellular filaments and flagellated zoospores. Molecular phylogeny now splits chytrids into at least three lineages that branch off

separately from the main line of fungal descent (Fig. 7.3; James et al. 2006a). Recent molecular data have identified two to three even earlier branches of the Holomycota, Rozellida, which are mostly known from environmental sequences (Lara et al. 2010), Nucleariida, which are strictly amoeboid (Zettler et al. 2001; Steenkamp et al. 2006), and *Fonticula alba*, which is also an amoeba that was previously classified as a slime mold (Brown et al. 2010). Meanwhile, the phylogenetic position of the unicellular and strictly parasitic Microsporidia, remains contentious (see the following section, Fig. 7.3).

The diversity of fungi appears to be extremely large and largely unknown; current estimates suggest over 1.5 million species, of which less than 5 % have been described (Stajich et al. 2009). This is supported by culture-independent sampling surveys (metagenetic analyses), which consistently reveal large numbers of unique but clearly fungal sequences (e.g., Vandenkoornhuyse et al. 2002; McLaughlin et al. 2009). Fungal classification and phylogeny is currently undergoing major revision, primarily due to the large-scale coordinated efforts of the Deep Hypha Research Coordination Network and Assembling the Fungal Tree of Life project (AFTOL1) (Blackwell et al. 2006). These have produced large amounts of sequence data from across the fungal tree. Analyses of these data indicate that many traditional taxa are invalid, particularly in terms of the deepest branches, the chytrids and zygomycetes. Both of these now appear to consist of several independently branching lineages, so that there appear to be at least nine distinct lineages of Holomycota that arose before the Ascomycota-Basidiomycota split (Fig. 7.3; James et al. 2006a, b). This puts fungi in the unique position of having clearly defined ancestors at a number of early stages in their evolution.

Members of the Holomycota appear to be rare in deep marine waters (Bass et al. 2007), with the exception of basal branching lineages such as Rozellida (Lara et al. 2010) and various lineages of "chytrids" (James et al. 2006a). Nonetheless, SSU rDNA sequences from marine waters show a wide diversity of fungi including late-branching ascomycetes and basidiomycetes (Bass et al. 2007). Since these sequences seem to be nested well within terrestrial fungal clades, they probably represent recent reinvasions of the marine environment rather than ancestral species (Bass et al. 2007). Thus, it is still not clear whether the last common ancestor of Fungi was a marine or freshwater organism (James et al. 2006a). Regardless of this, the major radiations of extant fungal lineages seem almost certain to have occurred on land. In fact, the evolution of land plants and the radiation of terrestrial fungi may have been tightly linked; it is now apparent that the two groups have many and highly varied interactions (Finlay 2008). One very significant interaction involves the glomalean (arbuscular mycorrhizal) fungi, which colonize the roots of the vast majority of land plants and increase their capacity to absorb inorganic nutrients (Bofante and Genre 2008). This interaction seems to have arisen early in land plant evolution and probably played a critical role in the successful colonization of land by green plants (Pirozynski and Malloch 1975; James et al. 2006a; Bonfante and Genre 2008).

Rozellida

Species of Rozella were first described over 50 years ago and initially classified as chytrids (Sparrow 1936). Very recently however, it was recognized that rozellids correspond to the deeply branching LKM11 clade, which was previously known only as frequently encountered unclassified 18S rDNA sequences in metagenetic surveys (Lara et al. 2010). All described rozellids are parasites, primarily of chytrids or oomycetes (Stramenopiles, Fig. 7.1), but also possibly of filamentous green algae (Lara et al. 2010). Their life cycle consists of a uniflagellate dispersal stage and a naked intracellular trophic stage, which is an amoeba that feeds by phagocytosis. The amoeba eventually matures into a multinucleate sporangium that gives rise to flagellated dispersal cells. Rozellids can also produce thick-walled resting forms, but no filamentous growth has been observed at any stage (Lara et al. 2010). Most known species are parasites of soil or freshwater hosts, except for one marine species, *Rozella marina* (Lara et al. 2010). Only 26 species have been described from living specimens. However, culture independent 18S rDNA (metagenetic) surveys of both fresh water and marine habitats indicate that this is a large and widely dispersed group that forms a very deep branch of Holomycota. In fact, whether or not they should be considered true fungi is a matter of debate (Lara et al. 2010; James et al. 2006a).

Nucleariida

Nucleariids are a small group of mostly free-living amoebae. The protoplast is a spherical or flattened elongate cell with fine radiating pseudopodia (filopodia; Zettler et al. 2001). Flagellated cells have not been observed at any stage in the life cycle. Although all described species thus far have been isolated from freshwater (Maldonado 2004), even in this situation *N. thermophila* is reported to simply float in the water as a rounded cell, with no evidence of swimming (Yoshida et al. 2009). Most described species are algivorous or bacterivorous (Patterson 1984), with the possible exception of *Nuclearia pattersoni*, which was discovered living in the gills of freshwater fish (Dyková et al. 2003). However, it has not yet been determined whether the association between this amoebae and its host is parasitic or commensal. Most nucleariids feed by phagocytosis, although *N. delicatula* uses its filopodia to penetrate damaged algal cells and remove the cell contents by phagocytosis (Cann 1986).

Nucleariids vary in size and in the ultrastructure of their mitochondria (reviewed in Yoshida et al. 2009); most species have discoidal cristae (Zettler et al. 2001) but *N. pattersoni* appears to have flattened cristae (Dyková et al. 2003). Cells range from 15 to 50 μm in diameter and are largely uninucleate with the exception of *N. delicatula*, which has been observed with 2–12 nuclei (Yoshida et al. 2009). Cyst formation has been observed in *N. pattersoni* and *N. simplex*, but both *N. delicatula* and *N. moebiusi* appear incapable of producing cysts. A mucoidal extracellular matrix, absent from

N. moebiusi, is present in *N. delicatula*, *N. pattersoni*, and *N. simplex* (Dyková et al. 2003; Patterson 1984).

The nucleariids were initially placed within the Filosea, together with the other amoeba with filose pseudopodia (Page 1987). However, Patterson (1999) argued that filopodia are likely to have evolved multiple times in eukaryotes, which was confirmed when the bulk of the Filosea were reclassified into the superkingdom Rhizaria (Fig. 7.1; Nikolaev et al. 2004). Zettler et al. were the first to sequence SSU rDNA from nucleariids and found strong support to include them in Opisthokonta (Zettler et al. 2001), and multi-gene phylogenies eventually placed the nucleariids as the sister-group to Fungi (Moreira et al. 2007; Ruiz-Trillo et al. 2004; Steenkamp et al. 2006). One important exception to this was *Nuclearia* sp. ATCC 30864, whose sequences clearly place it within Holozoa, and this is the organism that has now been reclassified as *Capsaspora owczarzaki* (see the previous section, Fig. 7.3; Hertel et al. 2002).

Fonticula alba

This is a single isolate of a single species with no known close relatives (Worley et al. 1979). It was originally described as a cellular slime mold, because it is an amoeba that aggregates to form a multicellular fruiting body. However, both its trophic and multicellular stages are unlike those of any other slime mold. *Fonticula alba* begins its life cycle as small (7–12 μm) amoebae with disc-shaped mitochondrial cristae (Worley et al. 1979) and sometimes quite-long filose pseudopodia with which it captures bacterial prey (Worley et al. 1979; Brown et al. 2009). When food becomes scarce, the amoebae aggregate by the thousands, initially forming a solitary mound of cells. The mounded cells secrete an extracellular matrix resulting in a pyramid-like structure or "stalk". Within this stalk, roughly two thirds of the amoebae encyst, while the rest remain at the base of the stalk and continue to secrete material. At maturity, the encysted amoebae are forcibly extruded through the stalk apex as a ball of spores (Worley et al. 1979; Brown et al. 2009). Thus, *Fonticula alba* resembles dictyostelids (Amoebozoa) and acrasids (Heterolobosea) in having a multicellular stage formed by aggregation, but it resembles myxogastrids (Amoebozoa) in that the fruiting body itself is secreted rather than the cellular body (Fig. 7.1; Brown et al. 2009; Fiore-Donno et al. 2010). Meanwhile, the morphology of *Fonticula* amoebae is unlike either amoebozoan or discobid amoebae, but rather resembles that of nucleariids in having filose pseudopodia and, in some cases, discoidal mitochondrial cristae (Zettler et al. 2001).

Fonticula alba has always been considered an enigmatic slime mold (Worley et al. 1979) and molecular phylogenies of SSU rDNA and multiple proteins now clearly place it in Holomycota (Brown et al. 2010). In fact, in all molecular phylogenies so far it groups unambiguously with *Nuclearia* (Brown et al. 2010). However, *Fonticula* sequences also tend to form very long branches in molecular trees, especially its SSU rDNA sequence (Brown et al. 2010). Therefore, it may just be an unusual

nucleariid with fast evolving sequences. However, it seems more likely that it is the sole known representative of a distinct and possibly quite extensive clade as it is also quite morphologically and behaviorally unique. In either case, it is clear that aggregative multicellularity has evolved numerous times in eukaryotes—in Dictyostelia (Amoebozoa), Acrasidae (Discoba), Cilliophora (Alveolata, Chaine et al. 2010), Holomycota and Rhizaria (Brown et al. 2012) (Fig. 7.1). In fact, aggregative multicellularity is known in at least one group of bacteria, the Myxobacteria (Velicer and Vos 2009).

Microsporidia

This is quite possibly the most enigmatic of all the opisthkont protist groups. Microsporidia are obligate intracellular parasites with no external morphology except for a highly structured spore stage (Canning and Curry 2004). Their genomes also tend to be extremely reduced in size and gene content (Corradi et al. 2009), and microsporidian molecular sequences tend to form extremely long branches in phylogenetic trees, even for essential housekeeping genes (Kamaishi et al. 1996). This means that they often appear as the deepest branches (e.g., Kamaishi et al. 1996; Vossbrinck et al. 1987). This deep branching position led to speculation that Microsporidia, which also lack many defining eukaryotic cell structures (respiring mitochondria, Golgi stacks), represent an early stage in the evolution of eukaryotic complexity (Cavalier-Smith 1989). However, this turned out instead to be a very striking case of long-branch attraction artifact, a phenomenon that had previously been largely ignored in the field molecular phylogeny (Palmer and Delwiche 1996; Gribaldo and Philippe 2002). It is now clear that Microsporidia belong among Holomycota (Hirt et al. 1999) either as a very deep branch (James et al. 2006b; McLaughlin et al. 2009) or, as has been suggested recently, among zygomycetes (Gill and Fast 2006; Lee et al. 2008).

Amoebozoa

Overview

Amoebozoa is one of three major groups of predominantly amoeboid eukaryotes, the others being Rhizaria and Heterolobosea (Discoba; Fig. 7.1). However, pseudopodial forms are scattered across the eukaryote tree, including among Metazoa, and it is likely that the ability to form pseudopods has arisen multiple times independently (Patterson 1994). Amoebozoa consists mostly of naked (non-testate) amoebae with broad lobose pseudopodia that tend to move in a smooth (non-eruptive) manner (Adl et al. 2005). However, like most major groups of eukaryotes, there are a number of

exceptions, and the taxon also includes a subgroup of testate amoebae (Arcellinida) and at least one major lineage of amoeboflagellates (Pawlowski and Burki 2009).

This is a deep and ancient lineage, at least as old as opisthokonts, and the major divisions of the group are at least hundreds of million years old, if not more (Berney and Pawlowski 2006). However, until recently, there has been little or no molecular data for most of it, except for a few model organisms (*Dictyostelium discoideum*, *Physarum polycephalum*) and pathogens (*Entamoeba histolytica*, *Acanthamoeba castelanii*). The fact that the group exhibits some of the most dramatically varied evolutionary rates for common molecular markers such as SSU rDNA (Tekle et al. 2008) and tubulins (Keeling and Doolittle 1996) has not helped matters. Although a general outline of the group seems to be emerging (Pawlowski and Burki 2009) as well as some understanding of evolutionary relationships within some of the subgroups (e.g., Smirnov et al. 2007; Smirnov et al. 2008; Nikolaev et al. 2006; Fiore-Donno et al. 2010), much of this remains tentative, especially for the deeper branches.

Much of the interest in Amoebozoa has centered on the "slime molds", the Dictyostelia and Myxogastria (Adl et al. 2005). Both possess conspicuous macroscopic stages, which can also, in the case of the myxogastrids be quite large, ornate and colorful, and the unusual developmental cycles of both groups have long fascinated developmental and evolutionary biologists. Both myxogastrids and dictyostelids are predators of bacteria and single-celled fungi living on decaying vegetation, herbivore dung and/or the humus layer of soil. Both also produce macroscopic fruiting bodies essentially consisting of a stalk supporting a "sorus" or ball of living spores. However, these sporophores take quite different forms in the two groups and are formed in fundamentally different ways, that is, by aggregation of free-living cells in dictyostelids or breakdown of a multinucleate plasmodium in myxogastrids (Olive 1975). Therefore, the relatedness of these two groups has been a matter of debate for a long time. The situation was exacerbated at first by molecular phylogeny (Johansen et al. 1988), since sequences for the two model organisms, *Dictyostelium discoideum* (Dictyostelia) and *Physarum polycephalum* (Myxogastria) were quite widely separated in early SSU rDNA trees (Gribaldo and Philippe 2002). This now appears to have been the result of radical differences in their nucleotide composition and relative evolutionary rates (Baldauf and Doolittle 1997; Gribaldo and Philippe 2002), and subsequent trees based on protein coding genes have firmly placed them together (Baldauf and Doolittle 1997; Baldauf et al. 2000; Hackett et al. 2007; Palfrey et al. 2010).

The most recent revision of eukaryote taxonomy places the Myxogastria and Dictyostelia together in a so-called "Eumycetozoa" along with protostelid slime molds (Adl et al. 2005). The latter are amoebae or amoeboflagellates that form simple fruiting bodies consisting of a minute stalk supporting one or a few spores. Some protostelids seem more similar to myxogastrids in being amoeboflagellates and/or forming tiny plasmodia while others are more like dictyostelids in lacking a flagellate or plasmodial stage (Olive 1975). This led to the hypothesis that various protostelids might represent early branches along the evolutionary lineages leading to dictyostelids or myxogastrids (Olive 1975). However, molecular phylogeny now shows that the "protostelids" are dispersed throughout much of the Amoebozoa

(Fiore-Donno et al. 2010; Shadwick et al. 2009). This indicates that Protostelia is not a valid taxon and, perhaps more importantly, the ability to raise one's spores above the substrate has evolved multiple times in Amoebozoa (2009). Thus, the grouping of Myxogastria and Dictyostelia is probably best referred to as the Macromycetozoa (Fiore-Donno et al. 2010).

Myxogastria

These are also called Myxomycetes or plasmodial slime molds and they represent the most numerous group of Amoebozoa with over 600 described species (Lado 2001). They are also the oldest known group of Macromycetozoa, and there are recorded observations from as early as the middle of the seventeenth century (Olive 1975). Traditionally, myxogastrids are often studied by mycologists because their sporophores superficially resemble the fruiting bodies of Fungi and overlap them in size. The group also includes the model organism and favorite of introductory biology labs, *Physarum polycephalum*. Traditionally, six taxonomic orders of Myxogastria have been recognized based primarily on macro- and microscopic characteristics of the sporophores (Lado 2001). These are Ceratiomyxales, Echinosteliales, Liceales, Physarales, Stemonitales, and Trichiales (Martin and Alexopoulos 1969). Recent molecular phylogeny separates the latter five into two major clades based on the pigmentation of the spores (dark-spored and light-spored clades, Fiore-Donno et al. 2005). Meanwhile, Ceratiomyxales may be the sister group to all Myxogastria or even Myxogastria plus Dictyostelia (Fiore-Donno et al. 2010).

The life cycle of myxogastrids has two very different trophic stages. One consists of uninucleate amoebae, which can be flagellated or not, and the other consists of a sometimes large and highly multinucleate structure, the plasmodium. Haploid spores germinate as uninucleate amoeba or flagellates of different mating types. Cells of opposite mating type may fuse and undergo karygamy. The resulting diploid cell may then give rise to a mobile plasmodium that can grow by feeding or fusing with other plasmodia (Olive 1975). When conditions are appropriate, the plasmodium gives rise to fruiting bodies. When conditions are adverse, the myxomycetes may form a multinucleate resistant structure called a sclerotium or individual nuclei may form dormant microcysts (Olive 1975).

Myxogastrids are particularly striking because of the size of the plasmodium, which can be a meter or more in diameter, containing tens or even hundreds of thousands of nuclei and maybe brightly colored. The fruiting bodies of myxogastrids are usually tiny but still macroscopic and often quite ornate, colorful, and abundant. The latter is due to the fact that the entire plasmodium may break down at once, resulting in a field of uniform sporophores. Although sporophore structure can appear quite complex, seemingly with different tissue layers of quite different structure, composition, and pigmentation, these are noncellular structures built by a plasmodium of undifferentiated nuclei (Olive 1975). Thus, myxogastrids are perhaps unique

among eukaryotes in having developed the ability to be large and conspicuous without cellular differentiation. Nonetheless, the cytoplasm of myxogastrids is probably an organized structure capable of rapid communication among the nuclei (Glöckner et al. 2008; Barrantes et al. 2010; Marwan 2010).

Dictyostelia

The phylogenetic position of dictyostelids was controversial for many years due to their life cycle characteristics. Some of these seem to be shared with Fungi, others with plants and still others with various eukaryotic microbes. Dictyostelia were considered to be Fungi based on morphological similarities in their fruiting bodies (Cappuccinelli and Ashworth 1977). However, others argued that since dictyostelids lack hyphae, they should be classified as protists and placed together with acrasids which are also amoeba with a multicellular stage (Olive 1975). However, Olive also noted fundamental differences in dicytostelid and acrasid amoebae in terms of their pseudopodial morphology and style movement (Olive 1975). This has been confirmed by molecular data, which now clearly show that acrasids are related to the heterolobosean amoebae such as *Naegleria* (supergroup Discoba) while Dictyostelids are clearly amoebozoans (Fig. 7.1; Roger et al. 1996; Baldauf et al. 2000; Palfrey et al. 2010). Thus, although dictyostelids and acrasids form a multicellular fruiting body by aggregation of formerly free-living cells, they are probably as far apart in the eukaryote tree as any two species can get (Fig. 7.1).

One of the most striking things about dictyostelids is their life cycle (Raper 1984; Kessin 2001). They have both sexual and asexual cycles, although the second is the most well known because it is the one most commonly observed in the laboratory. Both cycles begin with a trophic stage consisting of independent amoebas feeding on bacteria and multiplying by binary fission. When food becomes scarce, the amoebae signal each other using small diffusible molecules (acrasins) causing the cells to aggregate and form a multicellular organism. The first multicellular structure is the pseudoplasmodia or slug, which moves about as a single unit seeking optimal conditions for fruiting. At this point, cellular differentiation takes place and in the majority of species, approximately 80 % of the cells will be fated to become spore cells and 20 % to become stalk cells. When a favorable location is found, the pseudoplasmodium transforms into a multicellular fruiting body consisting of a cellular stalk composed of dead cells (except for *Acytostelium*-type species which have acellular stalks). These cells have essentially sacrificed themselves in order to elevate the sorus, the ball of live encysted cells (spores) above the substrate. Once these spores have dispersed, a new cycle begins. As in other groups of amoebae, individual cells can also form cysts to survive unfavorable conditions such as cold or dessication.

Traditionally dictyostelids have been classified based on fruiting body morphology and this classification is being retained until the radical revision indicated by molecular phylogeny can be undertaken (Schaap et al. 2006; Romeralo et al. 2010). Thus, there are three recognized "genera": the *Acytostelium* type includes species without cellular fruiting body stalks or streaming aggregation, the *Polysphondylium*

type are species with highly ordered multiheaded fruiting bodies with regularly spaced whorls of side branches, and the *Dictyostelium* type includes all species with cellular fruiting stalks and single or irregularly arranged multiheaded fruiting bodies (Olive 1975; Raper 1984). However, cladistic analysis of morphological data first suggested that this system was flawed (Swanson et al. 2002) followed by molecular data that clearly divide the group into four major clades, none of which correspond to the three traditional morphologically defined genera (Schaap et al. 2006; Romeralo et al. 2007).

Moreover, detailed examination of Dictyostelia using fine level molecular markers (ITS rDNA) indicates that many of the currently described species are in fact species complexes (Romeralo et al. 2010). That is, more often than not, multiple isolates of the same morphotype are not directly related to each other (Mehdiabadi et al. 2009; 2010). In fact, multiple species complexes have been found in all four major groups, and members of these "complexes" can sometimes even be quite distantly related to each other (2010). The most striking example is probably the *Polysphondylium* type, which is found in the two widely separated major groups 2 and 4 (Schaap et al. 2006). In fact, one of these, *P. violaceum*, appears to represent a new major group in its own right (Romeralo et al. 2011). Since most dictyostelid species have been described based on only one or just a few isolates, the number of species complex is probably much higher.

Five years ago, there were around 75 described species of Dictyostelia (Schaap et al. 2006). This seems surprising given the tremendous molecular depth of the group for SSU rDNA, which is roughly equivalent to that of the Metazoa (Schaap et al. 2006). While SSU rDNA is clearly not a very simple molecular chronometer (Gribaldo and Philippe 2002), the difference in the number of described species between Dictyostelia and Metazoa is still orders of magnitude (Purvis and Hector 2000). Thus, there is good reason to expect that there may be a substantial hidden diversity of Dictyostelia, which could in fact be the case with many if not most groups of eukaryotic microbes (e.g., Nolte et al. 2010). Thus, it should not be too surprising that the first large scale global effort to sample the diversity of Dictyostelia has more than doubled the number of described species (Romeralo et al. 2011).

In 2002, the National Science Foundation (USA) funded an international consortium to conduct a 5-year global survey of Mycetozoa, the only protist group to be funded in the first round of this global biodiversity survey effort. As a result, in the last 4 years alone the number of known dictyostelid species has more than doubled with species isolated from locations around the world including the USA, Argentina, Australia, South Africa, and Madagascar (Cavender et al. 2005; Vadell and Cavender 2007; Romeralo et al. 2011). Most of these new species are tiny acytostelid- and dictyostelid-types and are good candidates to fill in some of the more blank parts of the tree, which seem to be largely populated by small delicate species (Schaap et al. 2006; Romeralo et al. 2011). In fact, SSU rDNA phylogeny including the first 50 of these new species and isolates indicates that there may be as many as four more major divisions of dictyostelid, or a total of eight (Romeralo et al. 2011). The new species have also expanded the known morphological diversity of the original four major groups, so that there is no longer any discernible pattern of morphological evolution

across the group as a whole and no clear morphological justification for any of the original four major groups. Instead, it is now clear that many morphological patterns have arisen several times independently during the evolution of the dictyostelids.

Model Organism

Dictyostelids have been used as model organisms since the 1960s, when the first axenic strain of *Dictyostelium discoideum* was developed (Raper 1984). Bonner used this and other species to conduct a series of elegant experiments. These led to various discoveries such as that of the small molecule cyclic AMP (cAMP), which dictyostelids use for cell–cell signaling during development, and which Group 4 species also use for signaling aggregation (Bonner 2003). Although some dictyostelids use other small molecules to signal aggregation, all appear to use cAMP to control cell differentiation during development (Alvarez-Curto et al. 2007). Cell–cell recognition is also very important in dictyostelids, as in all forms except *Acytostelium* a proportion of the aggregating cells must sacrifice themselves in order to form the fruiting body stalk (Gilbert et al. 2007). Strassmann, Queller, and coworkers have used this to develop dictyostelids as a model system to study the evolution of social behavior (e.g., Khare et al. 2009).

Dictyostelids have a number of important advantages that have led to their widespread use as model organisms. They are generally easy to culture, non-pathogenic, tend to have small genomes and most known species readily complete their life cycle under laboratory conditions. Strains of *D. discoideum* have been developed that can be grown axenically, and these cells can be transformed with foreign DNA using a variety of vectors and selectable markers (Escalante and Vicente 2000). In fact, *D. discoideum* has a number of important advantages as a microbial model system for Metazoa (Williams et al. 2006), such as the fact that it seems to have retained more genes related to metazoan development than the other widely used model system for this purpose, *Saccharomyces cerevisiae* (Eichinger et al. 2005). The *D. discoideum* genome is small (34 megabases) with small introns and an estimated 12,500 proteins (Eichinger et al. 2005). One of the most fascinating features of the genome is that it is extremely AT-rich (70–80 %), and contains large tracts of triplet repeats in many of its protein coding genes, in fact more than any other eukaryote genome sequenced to date. These tracts appear to be translated and retained in the mature proteins (Eichinger et al. 2005).

This first *Dictyostelium* genome offers a first glimpse of how multicellularity evolved at the molecular level in the amoebozoan lineage. In fact, a broad survey of proteins known to be required for multicellular development in other lineages shows a number that have been retained in *Dictyostelium* but are missing in *Saccharomyces*. Since Fungi are more closely related to Metazoa than are the Amoebozoa, these proteins have presumably been lost at some point during fungal evolution (Eichenger et al. 2005). These include proteins involved in cell adhesion and signaling modules normally associated exclusively with Metazoa. At the same time, the *D. discoideum*

genome encodes more than 40 proteins involved in cellulose metabolism that are probably involved in fruiting body formation and maybe homologous to similar genes in plants (Eichinger et al. 2005).

Availability of the *D. discoideum* genome sequence greatly enhances many areas of molecular research, including large-scale post-genomic studies (Torija et al. 2006). This model system is also increasingly used to investigate basic aspects of important human diseases (Barth et al. 2007; Williams et al. 2006). Comparative genomics of *Dictyostelium* and related pathogens may also be important in defining amoebozoa-specific genes that could open new avenues of research aimed at controlling amoebic diseases. As with many important taxa, the pace of sequencing in Dictyostelia is accelerating and there are now three complete genome sequences publicly available: *D. discoideum* (Eichenger et al. 2005), *D. purpureum* (http://genome.jgi-psf.org/Dicpu1/Dicpu1.home.html) and *P. pallidum* and *D. fasciculatum* (www.dictybase.org) and many more to come soon. Thus, the era of comparative genomics in Dictyostelia is just beginning. In contrast, only a single genome-sequencing project is in progress for a myxogastrid, which is the best known myxogastrid model organism, *Physarum polycephalum* (http://www.ncbi.nlm.nih.gov/genomeprj/12700; Glöckner et al. 2008), while the only sequenced genomes for other amoebozoan are the human pathogens *Entamoeba histolytica* (Clark et al. 2007) and *Acanthamoeba castelanii* (Clarke et al. 2013).

Conclusions

Abundant evidence now supports the integrity of the eukaryotic supergroup Opisthkonta, and a number of single celled sister lineages are now known for both Metazoa and Fungi (Fig. 7.3). Much work remains to be done in order to develop a deeper understanding of the biology of these sister lineages and their relationships to each other and to their important multicellular relatives. This should involve more data from the few taxa we currently know, including full genome sequencing, but also discovery and characterization of more relatives of some of the most rare and highly distinct of these isolates. At the same time, it will be important to learn more about their sister taxa, the Amoebozoa and the Apusozoa (Kim et al. 2006; Palfrey et al. 2010) and the enigmatic *Breviata* (Minge et al. 2009), to give us a baseline for these comparisons but also because these are fascinating organisms in their own right and represent alternative evolutionary trajectories to multicellular complexity. Together these data will likely reshape our basic understanding of early events and alternative pathways in the evolution of eukaryotic complexity.

Acknowledgments We would like to thank Barry Leadbeater for numerous helpful discussions and for careful reading of the manuscript. We are also very grateful to Andreas Walberg and David Hibbett for advice on deep phylogeny of Metazoa and Fungi, respectively, and critical evaluation of Fig. 7.3. Maria Romeralo was supported for part of this work by a postdoctoral fellowship from the Systematics Society of Spain and by EU Marie Curie grant PIEF-GA-2009-236501. The work on Choanoflagellates was supported by a grant from the BBSRC (UK).

References

Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52:399–451

Ahluwalia KBN, Maheshwari N, Deka RC (1997) Rhinosporidiosis: a study that resolves etiologic controversies. Am J Rhinol 11:479–483

Alvarez-Curto E, Rozen DE, Ritchie AV, Fouquet C, Baldauf SL, Schaap P (2007) Evolutionary origin of cAMP-based chemoattraction in the social amoebae. Proc Natl Acad Sci U S A 102:6385–6390

Baker GC, Beebee TJ, Ragan MA (1999) *Prototheca richardsi*, a pathogen of anuran larvae, is related to a clade of protistan parasites near the animal-fungal divergence. Microbiology 145:1777–1784

Baldauf SL (2003) The deep roots of eukaryotes. Science 300:1703–1706

Baldauf SL, Doolittle WF (1997) Origin and evolution of the slime molds (Mycetozoa). Proc Natl Acad Sci U S A 94:12007–12012

Baldauf SL, Palmer JD (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci U S A 90:11558–11562

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972–977

Barrantes I, Glockner G, Meyer S, Marwan W (2010) Transcriptomic changes arising during light-induced sporulation in Physarum polycephalum. BMC Genomics 11:115

Barth C, Le P, Fisher PR (2007) Mitochondrial biology and disease in Dictyostelium. Int Rev Cytol 263:207–252

Bass D, Howe A, Brown N, Barton H, Demidova M, Michelle H, Li L, Sanders H, Watkinson SC, Willcock S, Richards TA (2007) Yeast forms dominate fungal diversity in the deep oceans. Proc Biol Sci 274:3069–3077

Benny GL, O'Donnell K (2000) *Amoebidium parasiticum* is a protozoan, not a Trichomycete. Mycologica 92:1133–1137

Bergsten J (2005) A review of long-branch attraction. Cladistics 21(2):163–193

Berney C, Pawlowski J (2006) A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc Biol Sci 273:1867–1872

Blackwell M, Hibbett DS, Taylor JW, Spatafora JW (2006) Research Coordination Networks: a phylogeny for kingdom Fungi (Deep Hypha). Mycologia 98:829–837

Boenigk J, Arndt H (2000) Comparative studies on the feeding behavior of two heterotrophic nanoflagellates: the filter-feeding choanoflagellate *Monosiga ovata* and the raptorial-feeding kinetoplastid *Rhynchomonas nasuta*. Aquat Microb Ecol 22:243–249

Bonfante P, Genre A (2008) Plants and arbuscular mycorrhizal fungi: an evolutionary-developmental perspective. Trends Plant Sci 13:492–498

Bonner JT (2003) On the origin of differentiation. J. Biosci. 28: 523-528

Bourlat SJ, Nielsen C, Economou AD, Telford MJ (2008) Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. Mol Phylogenet Evol 49:23–31

Bourrelly P (1968) Les algues d'eau douce. Tome II: Les algues jaunes et brunes. Boubée et Cie, Paris

Bozarth RF (1972) Mycoviruses: a new dimension in microbiology. Environ Health Perspect 2:23–39

Brown MW, Spiegel FW, Silberman JD (2009) Phylogeny of the "forgotten" cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. Mol Biol Evol 26:2699–2709

Brown MW, Silberman JD, Spiegel FW (2010) A morphologically simple species of Acrasis (Heterolobosea, Excavata), Acrasis helenhemmesae n. sp. J Eukaryot Microbiol 57:346–353
Brown MW, Kolisko M, Silberman JD, Roger AJ (2012) Aggregative multicellularity evolved independently in the eukaryotic supergroup Rhizaria. Curr Biol 22:1123–1127
Burger G, Forget L, Zhu Y, Gray MW, Lang BF (2003) Unique mitochondrial genome architecture in unicellular relatives of animals. Proc Natl Acad Sci U S A 100:892–897
Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J (2007) Phylogenomics reshuffles the eukaryotic supergroups. PLoS One 2(8):e790
Burki F, Shalchian-Tabrizi K, Pawlowski J (2008) Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett 4:366–369
Buss KW (1987) The evolution of individuality. Princeton University Press, Princeton, p 203
Cafaro MJ (2005) Eccrinales (Trichomycetes) are not fungi, but a clade of protists at the early divergence of animals and fungi. Mol Phylogenet Evol 35:21–34
Cann JP (1986) The feeding behavior and structure of *Nuclearia delicatula* (Filosea: Aconchulinida). J Eukaryot Microbiol 33:392–396
Canning EU, Curry A (2004) Further observations on the ultrastructure of Cystosporogenes operophterae (Canning, 1960) (phylum Microsporidia) parasitic in Operophtera brumata L. (Lepidoptera, Geometridae). J Invertebr Pathol 87:1–7
Cappuccinelli P, Ashworth JM (1977) Development and differentiation in the cellular slime moulds. Elsevier/North-Holland Biomedical Press, Amsterdam
Carr M, Leadbeater BS, Hassan R, Nelson M, Baldauf SL (2009) Molecular phylogeny of choanoflagellates, the sister group to Metazoa. Proc Natl Acad Sci U S A 105:16641–16646
Cavalier-Smith T (1987) The origin of fungi and pseudofungi. In: Rayner ADM, Brasier CM, Moore D (eds) Evolutionary biology of the fungi. Cambridge University Press, Cambridge, pp 339–353
Cavalier-Smith T (1989) Molecular phylogeny. Archaebacteria and Archezoa. Nature 339:100–101
Cavalier-Smith T (1998) A revised six kingdom system of life. Biol Rev Camb Philos Soc 73:203–266
Cavalier-Smith T, Allsopp MTEP (1996) *Corallochytrium*, an enigmatic non-flagellate protozoan related to choanoflagellates. Eur J Protistol 32: 306–310
Cavalier-Smith T, Chao EEY (2003) Phylogeny of Choanozoa, Apusozoa, and other protozoa and early eukaryotic megaevolution. J Mol Evol 56:540–563
Cavalier-Smith T, Chao EE (2010) Phylogeny and evolution of Apusomonadida (Protozoa: Apusozoa): new genera and species. Protist 161:549–576
Cavalier-Smith T, Chao EEY, Oates B (2004) Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont *Phalansterium*. Eur J Protistol 40: 21–48
Cavender JC, Vadell EM, Landolt JC, Stephenson SL (2005) New species of small dictyostelids from the Great Smoky Mountains National Park. Mycologia 97:493–512
Chadefaud M (1960) Les végétaux non vasulaires (Cryptogamie). In: Chadefaud M, Emberger L (eds) Traité de botanique systématique, vol. 3. Massor et Cie, Paris, p 1018
Chaine AS, Schtickzelle N, Polard T, Huet M, Clobert J (2010) Kin-based recognition and social aggregation in a ciliate. Evolution 64:1290–1300
Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 311:1283–1287
Clark CG, Alsmark UC, Tazreiter M, Saito-Nakano Y, Ali V, Marion S, Weber C, Mukherjee C, Bruchhaus I, Tannich E, Leippe M, Sicheritz-Ponten T, Foster PG, Samuelson J, Noël CJ, Hirt RP, Embley TM, Gilchrist CA, Mann BJ, Singh U, Ackers JP, Bhattacharya S, Bhattacharya A, Lohia A, Guillén N, Duchê ne M, Nozaki T, Hall N (2007) Structure and content of the Entamoeba histolytica genome. Adv Parasitol 65:51–190
Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, Bertelli C, Schilde C, Kianianmomeni A, Burglin TR, Frech C, Turcotte B, Kopec KO, Synnott JM, Choo C, Paponov I, Finkler A, Soon Heng Tan C, Hutchins AP, Weinmeier T, Rattei T, Chu JS, Gimenez G, Irimia M, Rigden DJ, Fitzpatrick DA, Lorenzo-Morales J, Bateman A, Chiu CH, Tang P, Hegemann P, Fromm H, Raoult D, Greub G, Miranda-Saavedra D, Chen N, Nash P, Ginger ML, Horn M, Schaap P, Caler L, Loftus B (2013) Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome Biol 14:R11

Corradi N, Haag KL, Pombert JF, Ebert D, Keeling PJ (2009) Draft genome sequence of the Daphnia pathogen Octosporea bayeri: insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. Genome Biol 10:R106

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sørensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745–749

Dyková I, Veverková M, Fiala I, Macháčková B, Pecková H (2003) *Nuclearia pattersoni* sp. n. (Filosea), a new species of amphizoic amoeba isolated from gills of roach (*Rutilus rutilus*), and its rickettsial endosymbiont. Folia Parasitol 50:161–170

Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q (2005) The genome of the social amoeba Dictyostelium discoideum. Nature 435:43–57

Escalante R, Vicente JJ (2000) Dictyostelium discoideum: a model system for differentiation and patterning. Int J Dev Biol 44:819–835

Feldman SH, Wimsatt JH, Green DE (2005) Phylogenetic classification of the frog pathogen *Amphibiothecum* (*Dermosporidium*) *penneri* based on small ribosomal subunit sequencing. J Wildl Dis 41:701–706

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410

Felsenstein J (2004) Inferring phylogenies. Sinauer Associates, Inc., Sunderland

Finlay RD (2008) Ecological aspects of mycorrhizal symbiosis: with special emphasis on the functional diversity of interactions involving the extraradical mycelium. J Exp Bot 59:1115–1126

Fiore-Donno AM, Berney C, Pawlowski J, Baldauf SL (2005) Higher-order phylogeny of plasmodial slime molds (Myxogastria) based on elongation factor 1-A and small subunit rRNA gene sequences. J Eukaryot Microbiol 52:201–210

Fiore-Donno AM, Nikolaev SI, Nelson M, Pawlowski J, Cavalier-Smith T, Baldauf SL (2010) Deep phylogeny and evolution of slime moulds (mycetozoa). Protist 161:55–70

Gilbert OM, Foster KR, Mehdiabadi NJ, Strassmann JE, Queller DC (2007) High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. Proc Natl Acad Sci U S A 104:8913–8917

Gile GH, Faktorová D, Castlejohn CA, Burger G, Lang BF, Farmer MA, Lukes J, Keeling PJ (2009) Distribution and phylogeny of EFL and EF-1alpha in Euglenozoa suggest ancestral co-occurrence followed by differential loss. PLoS One 4:e5162

Gill EE, Fast NM (2006) Assessing the microsporidia-fungi relationship: combined phylogenetic analysis of eight genes. Gene 375:103–109

Glöckner G, Golderer G, Werner-Felmayer G, Meyer S, Marwan W (2008) A first glimpse at the transcriptome of Physarum polycephalum. BMC Genomics 9:6

Gribaldo S, Philippe H (2002) Ancient phylogenetic relationships. Theor Popul Biol 61:391–408

Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D (2007) Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. Mol Biol Evol 24:1702–1713

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci U S A 106:3859–3864

Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, Martinez P, Baguà J, Bailly X, Jondelius U, Wiens M, Müller WE, Seaver E, Wheeler WC, Martindale MQ, Giribet G, Dunn CW (2009) Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc Biol Sci 276:4261–4270

Hendy MD, Penny D (1989) A framework for the quantitative study of evolutionary trees. Syst Zool 38:297–309

Herman RL (1984) Ichthyophonus-like infection in newts (*Notophthalmus viridescens* Rafinesque). J Wildl Dis 20:55–56

Herr RA, Ajello L, Taylor JW, Arseculeratne SN, Mendoza L (1999) Phylogenetic analysis of *Rhinosporidium seeberis* 18S small-subunit ribosomal DNA groups this pathogen among members of the protoctistan Mesomycetozoa clade. J Clin Microbiol 37:2750–2754

Hertel LA, Bayne C, Loker ES (2002) The symbiont *Capsaspora owczarzaki*, nov. gen. sp., isolated from three strains of the pulmonate snail *Biomphalaria glabrata* is related to members of the Mesomycetozoa. Int J Parasitol 32:1183–1191

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, Thorsten Lumbsch H, Lutzoni F, Matheny PB, McLaughlin DJ, Powell MJ, Redhead S, Schoch CL, Spatafora JW, Stalpers JA, Vilgalys R, Aime MC, Aptroot A, Bauer R, Begerow D, Benny GL, Castlebury LA, Crous PW, Dai YC, Gams W, Geiser DM, Griffith GW, Gueidan C, Hawksworth DL, Hestmark G, Hosaka K, Humber RA, Hyde KD, Ironside JE, Kõljalg U, Kurtzman CP, Larsson KH, Lichtwardt R, Longcore J, Miadlikowska J, Miller A, Moncalvo JM, Mozley-Standridge S, Oberwinkler F, Parmasto E, Reeb V, Rogers JD, Roux C, Ryvarden L, Sampaio JP, Schüssler A, Sugiyama J, Thorn RG, Tibell L, Untereiner WA, Walker C, Wang Z, Weir A, Weiss M, White MM, Winka K, Yao YJ, Zhang N (2007) A higher-level phylogenetic classification of the Fungi. Mycol Res 111:509–547

Hillis DM (1996) Inferring complex phylogenies. Nature 383: 130–131

Hirt RP, Logsdon JM Jr, Healy B, Dorey MW, Doolittle WF, Embley TM (1999) Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. Proc Natl Acad Sci U S A 96:580–585

Holterman M, van der Wurff A, van den Elsen S, van Megen H, Bongers T, Holovachov O, Bakker J, Helder J (2006) Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among Nematodes and accelerated evolution toward crown clades. Mol Biol Evol 23:1792–1800

James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung GH, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schüssler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeyer B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lücking R, Büdel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R (2006a) Reconstructing the early evolution of fungi using a six-gene phylogeny. Nature 443:818–822

James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D, Powell MJ, Grffith GW, Vilgalys R (2006b) A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). Mycologia 98:860–871

James-Clark H (1866) Note on the infusoria flagellata and the spongiae ciliatae. Am J Sci 1:113–114

James-Clark H (1867) On the Spongiae Ciliatae as Infusoria Flagellata: or observations on the structure, animality and relationship of *Leucosolenia botryoides* Bowerbank. Mem Boston Soc Nat Hist 1:305–340

Jay JM, Pohley WJ (1981) *Dermosporidium penneri* sp. n. from the skin of the American toad, *Bufo americanus* (Amphibia: Bufonidae). J Parasitol 67:108–110

Johansen T, Johansen S, Haugli FB (1988) Nucleotide sequence of the Physarum polycephalum small subunit ribosomal RNA as inferred from the gene sequence: secondary structure and evolutionary implications. Curr Genet 14:265–273

Kamaishi T, Hashimoto T, Nakamura Y, Nakamura F, Murata S, Okada N, Okamoto K, Shimizu M, Hasegawa M (1996) Protein phylogeny of translation elongation factor EF-1 alpha suggests microsporidians are extremely ancient eukaryotes. J Mol Evol 42: 257–263

Karpov SA, Leadbeater BSC (1997) Cytoskeleton structure and composition in choanoflagellates. Eur J Protistol 33:323–334

Keeling PJ, Doolittle WF (1996) Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. Mol Biol Evol 13:1297–1305

Kessin RH (2001) Dictyostelium: evolution, cell biology, and the development of multicellularity. Cambridge University Press, Cambridge

Khare A, Santorelli LA, Strassmann JE, Queller DC, Kuspa A, Shaulsky G (2009) Cheater-resistance is not futile. Nature 461:980–982

Kim E, Simpson AG, Graham LE (2006) Evolutionary relationships of apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. Mol Biol Evol 23:2455–2466

Kimura H, Harada K, Hara K, Tamaki A (2002) Enzymatic approach to fungal association with arthropod guts: a case study for the crustacean host, *Nihonotrypaea harmandi*, and its foregut fungus, *Enteromyces callianassae*. Mar Ecol 23:157–183

King N, Carroll SB (2001) A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. Proc Natl Acad Sci U S A 98:15032–15037

King N, Hittinger CT, Carroll SB (2003) Evolution of key cell signaling and adhesion protein families predates animal origins. Science 301:361–363

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, Marr M, Pincus D, Putnam N, Rokas A, Wright KJ, Zuzow R, Dirks W, Good M, Goodstein D, Lemons D, Li W, Lyons JB, Morris A, Nichols S, Richter DJ, Salamov A, Sequencing JG, Bork P, Lim WA, Manning G, Miller WT, McGinnis W, Shapiro H, Tjian R, Grigoriev IV, Rokhsar D (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature 451:783–788

King N, Young SL, Abedin M, Carr M, Leadbeater BS (2009) Separation of choanoflagellate and bacterial genomic DNA. Cold Spring Harb Protoc 2009(2):pdb.prot5154

Kolisko M, Cepicka I, Hampl V, Leigh J, Roger AJ, Kulda J, Simpson AG, Flegr J (2008) Molecular phylogeny of diplomonads and enteromonads based on SSU rRNA, alpha-tubulin and HSP90 genes: implications for the evolutionary history of the double karyomastigont of diplomonads. BMC Evol Biol 8:205

Kruger W (1894) Kurze Charakteristik einiget niederer Organismen in Saftlusse der Laubbaume. Hedwigia 33:241–266

Lado C (2001) NOMENMYX. A nomenclaturaltaxabase of Myxomycetes. Cuadernos de Trabajo de Flora Mycologica Iberica 16:1–224

Lara E, Chatzinotas A, Simpson AG (2006) Andalucia (n. gen.)–the deepest branch within jakobids (Jakobida; Excavata), based on morphological and molecular study of a new flagellate from soil. J Eukaryot Microbiol 53:112–120

Lara E, Moreira D, López-García P (2010) The environmental clade LKM11 and *Rozella* form the deepest branching clade of fungi. Protist 161:116–121

Leadbeater BS, Karpov SA (2000) Cyst formation in a freshwater strain of the choanoflagellate Desmarella moniliformis Kent. J Eukaryot Microbiol 47:433–439

Leadbeater BSC (1983) Distribution and chemistry of microfilaments in choanoflagellates, with special reference to the collar and other tentacle systems. Protistologica 19:157–166

Leadbeater BSC (2008a) Choanoflagellate lorica construction and assembly: the nudiform condition. I. *Savillea* species. Protist 159:259–268

Leadbeater BSC (2008b) Choanoflagellate evolution: the morphological perspective. Protistology 5:256–267

Leadbeater BSC (2010) Choanoflagellate lorica construction and assembly: the tectiform condition. Volkanus costatus (= Diplotheca costata). Protist 161:160–176

Leadbeater BSC, Cheng R (2010) Costal strip production and lorica assembly in the large tectiform choanoflagellate Diaphanoeca grandis Ellis. Eur J Protistol 46: 96–110

Leadbeater BSC, Morton C (1974) A microscopical study of a marine species of *Codosiga* James-Clark (Choanoflagellata) with special reference to the ingestion of bacteria. Biol J Linn Soc 6:337–347

Leadbeater BSC, Thomsen H (2000) "Order Choanoflagellida". An illustrated guide to the Protozoa, vol. 451, 2nd edn. Society of Protozoologists, Lawrence, pp 14–38

Leadbeater BSC, Hassan R, Nelson M, Carr M, Baldauf SL (2008b) A new genus, Helgoeca gen. nov., for a nudiform choanoflagellate. Eur J Protistol 44:227–237

Leadbeater BSC, Yu Q, Kent J, Stekel DJ (2009) Three-dimensional images of choanoflagellate loricae. Proc R Soc B 276:3–11

Leakey RJG, Leadbeater BSC, Mitchell E, McCready SMM, Murray AWA (2002) The abundance and biomass of choanoflagellates and other nanoflagellates in waters of contrasting temperature to the north-west of South Georgia in the Southern Ocean. Eur J Protistol 38:333–350

Lee SC, Corradi N, Byrnes EJ 3rd, Torres-Martinez S, Dietrich FS, Keeling PJ, Heitman J (2008) Microsporidia evolved from ancestral sexual fungi. Curr Biol 18:1675–1679

Lichtwardt RW (1986) The Trichomycetes: fungal associates of arthropods. Springer-Verlag, New York, p 343

Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K (2009) Infrequent marine-freshwater transitions in the microbial world. Trends Microbiol 17:414–422

Maldonado M (2004) Choanoflagellates, choanocytes, and animal multicellularity. Invertebr Biol 123:1–22

Marande W, López-García P, Moreira D (2009) Eukaryotic diversity and phylogeny using small- and large-subunit ribosomal RNA genes from environmental samples. Environ Microbiol 11:3179–3188

Marshall WL, Celio G, McLaughlin DJ, Berbee ML (2008) Multiple isolations of a culturable, motile Ichthyosporean (Mesomycetozoa, Opisthokonta), Creolimax fragrantissima n. gen., n. sp., from marine invertebrate digestive tracts. Protist 159:415–433

Martin GW, Alexopoulos CJ (1969) The Myxomycetes. University of Iowa Press, Iowa City

Marwan W (2010) Systems biology. Amoeba-inspired network design. Science 327:419–420

Massana R, Terrado R, Forn I, Lovejoy C, Pedrós-Alió C (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. Environ Microbiol 8:1515–1522

McLaughlin DJ, Hibbett DS, Lutzoni F, Spatafora JW, Vilgalys R (2009) The search for the fungal tree of life. Trends Microbiol 17:488–497

Medina M, Collins AG, Silberman JD, Sogin ML (2001) Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. Proc Natl Acad Sci U S A 98:9707–9712

Medlin LK (2007) If everything is everywhere, do they share a common gene pool? Gene 406:180–183

Mehdiabadi NJ, Kronforst MR, Queller DC, Strassmann JE (2009) Phylogeny, reproductive isolation and kin recognition in the social amoeba Dictyostelium purpureum. Evolution 63:542–548

Mendoza L, Taylor J, Ajello L (2002) The class Mesomycetozoea: a heterogeneous group of microorganisms at the animal-fungal boundary. Ann Rev Microbiol 56:315–344

Minge MA, Silberman JD, Orr RJ, Cavalier-Smith T, Shalchian-Tabrizi K, Burki F, Skjaeveland A, Jakobsen KS (2009) Evolutionary position of breviate amoebae and the primary eukaryote divergence. Proc Biol Sci 276:597–604

Moreira D, von der Heyden S, Bass D, López-García P, Chao EEY, Cavalier-Smith T (2007) Global eukaryote phylogeny: combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. Mol Phylogenet Evol 44:255–266

Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, Aleshin VV, Petrov NB, Pawlowski J (2004) The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci U S A 101:8066–8071

Nikolaev SI, Berney C, Petrov NB, Mylnikov AP, Fahrni JF, Pawlowski J (2006) Phylogenetic position of Multicilia marina and the evolution of Amoebozoa. Int J Syst Evol Microbiol 56:1449–1458

Nitsche F, Arndt H (2008) A new choanoflagellate species from Taiwan: morphological and molecular biological studies of *Diplotheca elongata* nov. spec. and *D. costata*. Eur J Protistol 44:220–226

Nolte V, Pandey RV, Jost S, Medinger R, Ottenwälder B, Boenigk J, Schlötterer C (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. Mol Ecol 19:2908–2915

Norris RE (1965) Neustonic marine Craspedomonadales (Choanoflagellata) from Washington and California. J Protozool 12:589–612
Not F, del Campo J, Balagué V, de Vargas C, Massana R (2009) New insights into the diversity of marine picoeukaryotes. PLoS One 4:e7143
Ohta T (1972) Population size and rate of evolution. J Mol Evol 1:305–314
Olive LS (1975) The Mycetozoans. Academic, New York
O'Malley MA (2007) The nineteenth century roots of 'everything is everywhere'. Nat Rev Microbiol 5:647–651
Ordás MC, Figueras A (1998) *In vitro* culture of *Perkinsus atlanticus*, a parasite of the carpet shell clam *Ruditapes decussates*. Dis Aquat Org 33:129–136
Owczarzak A, Stibbs HH, Bayne CJ (1980) The destruction of *Schistosoma mansoni* mother sporocysts in vitro by amoebae isolated from *Biomphalaria glabrata*: an ultrastructural study. J Invertebr Pathol 35:26–33
Page FC (1987) The classification of 'naked' amoebae (Phylum Rhizopoda). Arch Protistenkd 133:199–217
Palmer JD, Delwiche CF (1996) Second-hand chloroplasts and the case of the disappearing nucleus. Proc Natl Acad Sci U S A 93:7432–7435
Parfrey LW, Grant J, Tekle YI, Lasek-Nesselquist E, Morrison HG, Sogin ML, Patterson DJ, Katz LA (2010) Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. Syst Biol 59:518–533
Parikh A, Miranda ER, Katoh-Kurasawa M, Fuller D, Rot G, Zagar L, Curk T, Sucgang R, Chen R, Zupan B, Loomis WF, Kuspa A, Shaulsky G (2010) Conserved developmental transcriptomes in evolutionarily divergent species. Genome Biol 11:R35
Patterson DJ (1984) The genus *Nuclearia* (Sarcodina, Filosea): species composition and characteristics of the taxa. Arch Protistenkd 128:127–139
Patterson DJ (1988) The evolution of protozoa. Mem Inst Oswaldo Cruz 83(Suppl 1):580–600
Patterson DJ (1994) Protozoa, evolution and classification. In: Progress in protozoology. Proceedings of the IX International Congress of Protozoology. Fischer, Stuttgart, pp 1–14
Patterson DJ (1999) The diversity of eukaryotes. Am Nat 154:96–124
Patterson DJ, Nygaard K, Steinberg G, Turley CM (1992) Heterotrophic flagellates and other protists associated with detritus in the mid North Atlantic. J Mar Biol Assoc U K 73:67–95
Pawlowski J, Burki F (2009) Untangling the phylogeny of amoeboid protists. J Eukaryot Microbiol 56:16–25
Pettitt ME, Orme BAA, Blake JR, Leadbeater BSC (2002) The hydrodynamics of filter feeding in choanoflagellates. Eur J Protistol 38:313–332
Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Quéinnec E, Da Silva C, Wincker P, Le Guyader H, Leys S, Jackson DJ, Schreiber F, Erpenbeck D, Morgenstern B, Wörheide G, Manuel M (2009) Phylogenomics revives traditional views on deep animal relationships. Curr Biol 19:706–712
Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, Wiens M, Alié A, Morgenstern B, Manuel M, Wörheide G (2010) Improved phylogenomic taxon sampling noticeably affects non-bilaterian relationships. Mol Biol Evol 27:1983–1987
Pirozynski KA, Malloch DW (1975) The origin of land plants: a matter of mycotrophism. Biosystems 5:153–164
Purvis A, Hector A (2000) Getting the measure of biodiversity. Nature 405: 212–219
Ragan MA, Goggins CL, Cawthorn RJ, Cerenius L, Jamienson AV, Plourde SM, Rand TG, Söderhäll K, Gutell RR (1998) A novel clade of protistan parasites near the animal-fungal divergence. Proc Natl Acad Sci U S A 93:11907–11912
Raghu-Kumar S, Chandramohan D, Ramaiah N (1987) Contribution of the thraustochytrid *Corallochytrium limacisporum* Raghu-kumar to microbial biomass in coral reef lagoons. Indian J Mar Sci 16:122–125
Rand TG (1994) An unusual form of *Ichthyophonus hoferi* from yellowtail flounder *Limanda Ferruginea* from the Nova Scotia shelf. Dis Aquat Org 18:21–28
Raper KB (1984) The Dictyostelids. Princeton University Press, Princeton

Reeb VC, Peglar MT, Yoon HS, Bai JR, Wu M, Shiu P, Grafenberg JL, Reyes-Prieto A, Rümmele SE, Gross J, Bhattacharya D (2009) Interrelationships of chromalveolates within a broadly sampled tree of photosynthetic protists. Mol Phylogenet Evol 53:202–211

Rohde RA, Muller RA (2005) Cycles in fossil diversity. Nature 434:208–210

Roger AJ, Smith MW, Doolittle RF, Doolittle WF (1996) Evidence for the Heterolobosea from phylogenetic analysis of genes encoding glyceraldehyde-3-phosphate dehydrogenase. J Eukaryot Microbiol 43:475–485

Romeralo M, Cavender J, Landolt J, Stepehnson S, Baldauf S (2011) The expanding phylogeny of social amoebae defines new major lineages and emerging patterns in morphological evolution. (BMC Evolutionary Biology 11:84)

Romeralo M, Escalante R, Sastre L, Lado C (2007) Molecular systematics of dictyostelids: 5.8S ribosomal DNA and internal transcribed spacer region analyses. Eukaryot Cell 6:110–116

Romeralo M, Spiegel FW, Baldauf SL (2010) A fully resolved phylogeny of the social amoebas (Dictyostelia) based on combined SSU and ITS rDNA sequences. Protist 161:539–548

Rota-Stabelli O, Kayal E, Gleeson D, Daub J, Boore JL, Telford MJ, Pisani D, Blaxter M, Lavrov DV (2010) Ecdysozoan mitogenomics: evidence for a common origin of the legged invertebrates, the Panarthropoda. Genome Biol Evol 2:425–440

Ruiz-Trillo I, Burger G, Holland PWH, King N, Lang BF, Roger AJ, Gray MW (2007) The origins of multicellularity: a multi-taxon genome initiative. Trends Genet 23:113–118

Ruiz-Trillo I, Inagaki Y, Davis LA, Sperstad S, Landfald B, Roger AJ (2004) *Capsaspora owczarzaki* is an independent opisthokont lineage. Curr Biol 14:R946

Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF (2008) A phylogenomic investigation into the origin of Metazoa. Mol Biol Evol 25:664–672

Sakaguchi M, Inagaki Y, Hashimoto T (2007) Centrohelida is still searching for a phylogenetic home: analyses of seven Raphidiophrys contractilis genes. Gene 405:47–54

Sanderson MJ (1989) Confidence limits on phylogenies: the bootstrap revisited. Cladistics 5:113–129

Saville-Kent W (1880) Manual of the infusoria, including a description of all known flagellate, ciliate, and tentaculiferous protozoa, British and foreign and an account of the organization and affinities of the sponges, vol 1. pp 135–142

Schaap P, Winckler T, Nelson M, Alvarez-Curto E, Elgie B, Hagiwara H, Cavender J, Milano-Curto A, Rozen DE, Dingermann T, Mutzel R, Baldauf SL (2006) Molecular phylogeny and evolution of morphology in the social amoebas. Science 314:661–663

Schweikert M, Schnepf E (1996) Pseudaphelidium drebesii, gen et spec nov (incerta sedis), a parasite of the marine centric diatom Thalassiosira punctigera. Arch Protistenkd 147:11–17

Shadwick LL, Spiegel FW, Shadwick JD, Brown MW, Silberman JD (2009) Eumycetozoa = Amoebozoa?: SSUrDNA phylogeny of protosteloid slime molds and its significance for the amoebozoan supergroup. PLoS One 4:e6754

Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T (2008) Multigene phylogeny of Choanozoa and the origin of animals. PLoS One 3:e2098

Smirnov AV, Nassonova ES, Chao E, Cavalier-Smith T (2007) Phylogeny, evolution, and taxonomy of vannellid amoebae. Protist 158:295–324

Smirnov AV, Nassonova ES, Cavalier-Smith T (2008) Correct identification of species makes the amoebozoan rRNA tree congruent with morphology for the order Leptomyxida Page 1987; with description of Acramoeba dendroida n. g., n. sp., originally misidentified as 'Gephyramoeba sp.'. Eur J Protistol 44:35–44

Snell EA, Brooke NM, Taylor WR, Casane D, Philippe H, Holland PW (2006) An unusual choanoflagellate protein released by Hedgehog autocatalytic processing. Proc Biol Sci 273:401–407

Spanggaard B, Skouboe P, Rossen L, Taylor JW (1996) Phylogenetic relationships of the intercellular fish pathogen *Ichthyophonus hoferi*, and fungi, choanoflagellates and the rosette agent. Mar Biol 126:109–115

Sparrow FK (1936) Biological observations on the marine fungi of Woods Hole waters. Biol Bull 70:236–263

Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, Taylor JW (2009) The fungi. Curr Biol 19:R840–R845

Stechmann A, Cavalier-Smith T (2002) Rooting the eukaryote tree by using a derived gene fusion. Science 297:89–91

Steenkamp ET, Wright J, Baldauf SL (2006) The protistan origins of animals and fungi. Mol Biol Evol 23:93–106

Stibbs HH, Owczarzak A, Bayne CJ, DeWan P (1979) Schistosome sporocyst-killing amoebae isolated from *Biomphalaria glabrata*. J Invertebr Pathol 33:159–170

Swanson AR, Spiegel FW, Cavender JC (2002) Taxonomy, slime molds, and the questions we ask. Mycologia 94:968–979

Tehler A (1988) A cladistic outline of the Eumycota. Cladistics 4:227–277

Tekle YI, Grant J, Anderson OR, Nerad TA, Cole JC, Patterson DJ, Katz LA (2008) Phylogenetic placement of diverse amoebae inferred from multigene analyses and assessment of clade stability within 'Amoebozoa' upon removal of varying rate classes of SSU-rDNA. Mol Phylogenet Evol 47:339–352

Telford MJ (2009) Animal evolution: once upon a time. Curr Biol 19:R339–R341

Telford RJ, Vandvik V, Birks HJ 2006) Dispersal limitations matter for microbial morphospecies. Science 312:1015

Thomsen HA, Buck KB, Coale SL, Garrison DL, Gowing MM (1990) Loricate choanoflagellates (Acanthoecidae Choanoflagellida) from the Weddell Sea, Antarctica. Zoologica Scripta 19:367–387

Thomsen HA, Buck K, Chavez F (1991) Choanoflagellates of the central California waters: taxonomy, morphology and species assemblages. Ophelia 33:131–164

Torija P, Robles A, Escalante R (2006) Optimization of a large-scale gene disruption protocol in Dictyostelium and analysis of conserved genes of unknown function. BMC Microbiol 6:75

Vadell E, Cavender JC (2007) Dictyostelids living in the soils of the Atlantic Forest. Iguazú region, Misiones, Argentina. Description of new species. Mycologia 99:112–124

van Etten JL, Lane LC, Meints RH (1991) Viruses and viruslike particles of eukaryotic algae. Microbiol Rev 55:586–620

Vandenkoornhuyse P, Baldauf SL, Leyval C, Straczek J, Young JP (2002) Extensive fungal diversity in plant roots. Science 295:2051

Velicer GJ, Vos M (2009) Sociobiology of the myxobacteria. Annu Rev Microbiol 63:599–623

Vossbrinck CR, Maddox JV, Friedman S, Debrunner-Vossbrinck BA, Woese CR (1987) Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature 326:411–414

Williams RS, Boeckeler K, Graf R, Muller-Taubenberger A, Li Z, Isberg RR, Wessels D, Soll DR, Alexander H, Alexander S (2006) Towards a molecular understanding of human diseases using Dictyostelium discoideum. Trends Mol Med 12:415–424

Worley AC, Raper KB, Hohl M (1979) Fonticula alba: a new cellular slime mold (Acrasiomycetes). Mycologia 71:746–760

Wainright PO, Hinkle G, Sogin ML, Stickel SK (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. Science 260:340–342

Whisler HC (1960) Pure culture of the Trichomycete, Amoebidium parasiticum. Nature 186:732–733

Yoshida M, Nakayama T, Inouye I (2009) *Nuclearia thermophila* sp. nov. (Nucleariidae), a new nucleariid species isolated from Yunoko Lake in Nikko (Japan). Eur J Protistol 45:147–155

Zettler LAA, Nerad TA, O'Kelly CJ, Sogin ML (2001) The nucleariid amoebae: more protists at the animal-fungal boundary. J Eukaryot Microbiol 48:293–297

Zubkov MV, Martin-Cereceda M, Novarino G, Wootton EC, Roberts EC (2006) Cell surface lectin-binding glycoconjugates on marine planktonic protists. FEMS Microbiol Lett 265:202–207

Chapter 8
Written in Stone: The Fossil Record of Early Eukaryotes

Shuhai Xiao

Introduction: Eukaryotic Phylogenetic Tree and Geological Time Scale

Eukaryotes, archaeobacteria, and eubacteria constitute the three domains of living organisms. Most life forms that we can see with naked eyes are multicellular eukaryotes, such as animals, fungi, plants, and seaweeds. However, many eukaryotes are single-celled and microscopic, and some of these microscopic eukaryotes (e.g., foraminifers, coccolithophorans, and diatoms) are important players in regulating the Earth's systems. A eukaryotic cell is cytologically distinct from archaeobacterial and eubacterial cells. It typically contains membrane-bound intracellular structures such as nucleus, mitochondrion, and, for photosynthetic eukaryotes (such as algae and plants), chloroplast. Eukaryotes are also distinctively characterized by DNA-associated histone, eukaryotic gene regulation, tubulin- and actin-based cytoskeletons, and the capability to carry out endocytosis and exocytosis.

The phylogenetic relationships among different eukaryotic groups have been a matter of debate. However, in recent years, the big picture of eukaryote phylogeny has emerged (see Chap. 7 in this volume). Based on molecular data, six major eukaryote clades have been recognized (Fig. 8.1), although many details of the eukaryotic tree remain to be resolved. These clades are the opisthokonts, amoebozoans, plants (red and green algae), chromalveolates, rhizarians, and excavates (Baldauf 2003; Keeling et al. 2005; Fehling et al. 2007). The rooting of the eukaryotic tree is still controversial (Brinkmann and Philippe 2007; Cavalier-Smith 2010a; Roger and Simpson 2009), and several possibilities have been proposed: the deepest eukaryote root is between the unikonts (opisthokonts + amoebozoans) and bikonts (plants + chroma-

S. Xiao (✉)
Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060, USA
e-mail: xiao@vt.edu

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,* Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_8,

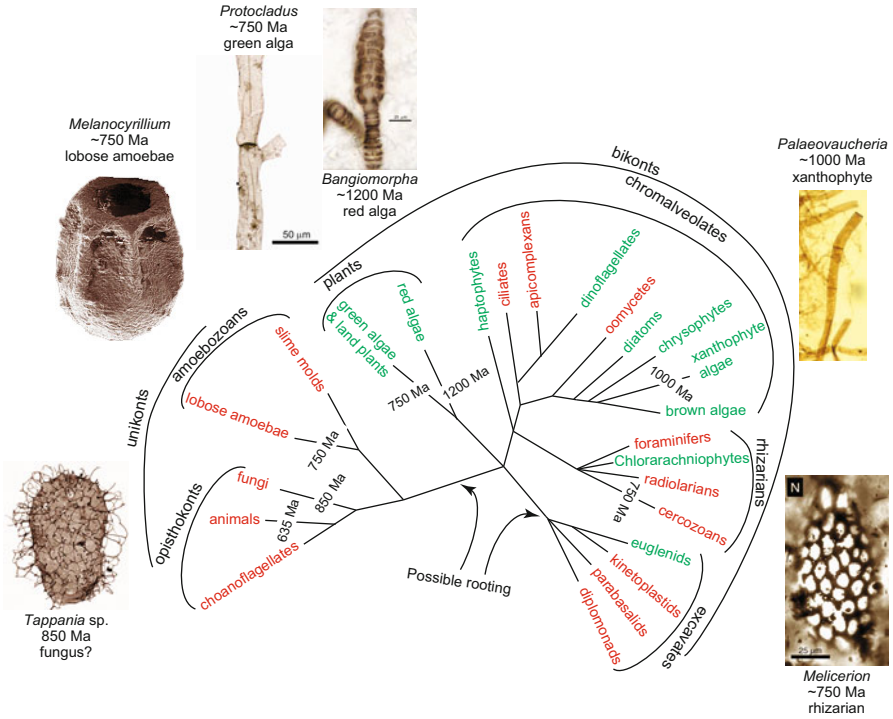


Fig. 8.1 Phylogenetic relationships among major eukaryotic clades based on molecular data. Modified after (Baldauf 2003; Keeling et al. 2005; Porter 2006; Roger and Simpson 2009). Taxa labeled in *green* are photosynthetic groups. Representative fossils (all from Stage II) are shown and their interpreted phylogenetic affinities and ages are marked on the phylogenetic tree

lveolates + rhizarians + excavates) (Stechmann and Cavalier-Smith 2002), between the Euglenozoa and all other eukaryotes, or within the Euglenozoa (Cavalier-Smith 2010a). Regardless, the deep divergence of the major eukaryotic groups predicts that they may have diverged during a eukaryote "big bang" (Philippe et al. 2000).

For those who are not familiar with the rock record, it is necessary to introduce the geological time scale (Fig. 8.2). The geological time is organized in a hierarchical system (just like a hierarchical taxonomic system), with the highest rank being the eons. The geological time is divided into four eons: the Hadean (> 4,000 Ma, million years ago), Archean (4,000–2,500 Ma), Proterozoic (2,500–542 Ma), and Phanerozoic (542–0 Ma). The Archean Eon is further divided into four eras, with youngest being the Neoarchean Era (2,800–2,500 Ma). The Proterozoic Eon includes the Paleoproterozoic, Mesoproterozoic, and Neoproterozoic eras, the latter of which is further divided into the Tonian, Cryogenian, and Ediacaran periods. The focus in the paper is on the Neoarchean Era and the Proterozoic Eon.

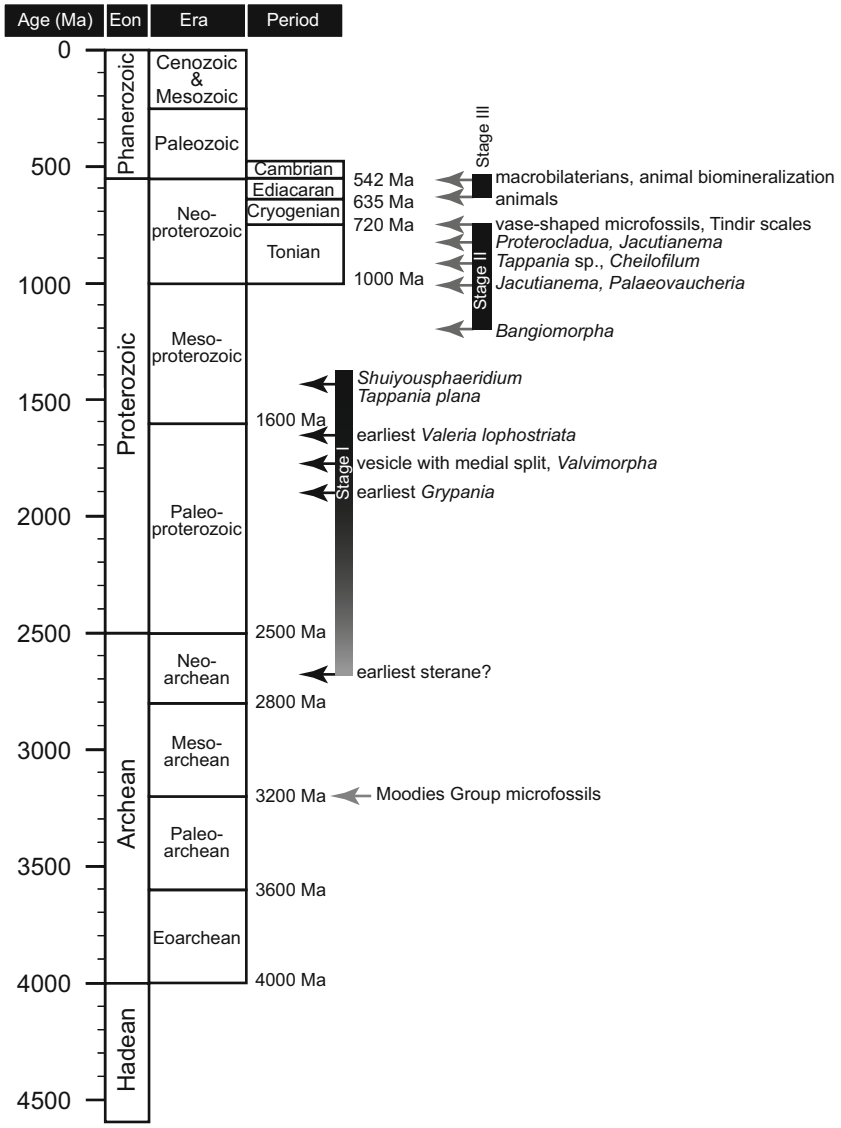


Fig. 8.2 The geological time scale. Modified from (Ogg et al. 2008). Ages of important fossil groups discussed in the text are marked against the geological time scale

Stage I (Neoarchean–Mesoproterozoic): Stem Group Eukaryotes

The earliest eukaryotes were probably morphologically simple, single-celled organisms. The identification of these simple eukaryotic fossils can be challenging. Important cytological features, such as nucleus, mitochondrion, chloroplast, histone,

and cytoskeleton are rarely preserved in the fossil record. It has been suggested that cell size may be a useful guidance; eukaryotes tend to have larger cells than prokaryotes. However, the size range of eukaryotic cells overlaps with bacterial cells, which can reach several hundred micrometers in size (Schulz and Jørgensen 2001). Thus, cell size is an inconclusive evidence of eukaryotic affinity. Morphological features can sometimes provide more definitive evidence for eukaryotic affinity. For example, some eukaryotic cells are capable of maintaining distinct cell morphologies, such as spiny processes, cell wall ornamentations, and biologically programmed excystment structures, which can be readily preserved in the fossil record. For multicellular fossils, eukaryotic morphologies extend to distinct cell organization and differentiation. Finally, eukaryotic biochemistry may leave fingerprints in ancient rocks, in the form of molecular fossils or biomarkers. For example, the sterol biosynthesis is predominantly a eukaryotic characteristic. Thus, abundant steranes (geologically derived from sterols) preserved in the rock record can be regarded as molecular evidence (or biomarkers) for eukaryotes.

Molecular and morphological evidence for the earliest eukaryote is controversial. Steranes from the ~2,600 Ma, Marra Mamba Formation and the ~2,715 Ma Maddina Formation in Western Australia were interpreted as evidence for Neoarchean eukaryotes (Brocks et al. 1999, 2003). However, contamination can be a major issue in biomarker studies (Kirschvink and Kopp 2008) and recent reexamination of the Australian samples indicates that the steranes reported in the original studies may be contaminants and thus may not provide evidence for the existence of eukaryotes in the Neoarchean (Rasmussen et al. 2008). On the other hand, several other studies, some with improved laboratory protocols to detect postdepositional contamination, reported additional evidence of Neoarchean steranes from Western Australia and Canada (Sherman et al. 2007; Ventura et al. 2007). Furthermore, fluid inclusions trapped in quartz and feldspar minerals in the ~2,450 Ma Matinenda Formation at Elliot Lake, Canada, contain steranes (Dutkiewicz et al. 2006); these biomarkers provide good evidence for eukaryotic life in the early Paleoproterozoic. Thus, although there is some tantalizing molecular evidence for the existence of eukaryotic life in the Neoarchean, such evidence should be viewed with caution and awaits further verification.

Even if the existence of Neoarchean sterane biomarkers were confirmed, the evolutionary significance of such biomarkers can be ambiguous. The presence of steranes in Archean rocks does not necessarily imply that all features (e.g., mitochondrion, cytoskeleton, and histone) that collectively define extant eukaryotes have evolved in the Archean. Sterol biosynthesis may be one of the eukaryotic features that appeared early in the eukaryote history, and the Archean sterol-producing organisms represent ancient and extinct evolutionary offshoots that branched after the divergence of the eukaryotes from the other two domains but before the divergence of the six major eukaryote clades that are still extant today. In other words, they represent stem group eukaryotes. It is also possible that there may have been extinct biosynthetic pathways that were responsible for the production of Archean sterols (Kirschvink and Kopp 2008).

Because of the phylogenetic ambiguity of molecular fossils, it is important that they are substantiated by morphological fossils. Unfortunately, there are no Archean morphological fossils that can be unequivocally interpreted as eukaryotes. Microfossils from Archean rocks, which are often metamorphosed through geological ages, are often graphitized and their morphologies are obscured (Schiffbauer et al. 2007). Some exceptionally preserved organic-walled microfossils have been reported from the 3,200 Ma Moodies Group in South Africa (Javaux et al. 2010). The Moodies Group microfossils are ~ 120 μm in diameter and they could be eukaryotic microfossils, although a bacterium interpretation cannot be excluded.

The earliest known morphological fossils that can be reasonably interpreted as eukaryotes come from Paleoproterozoic rocks. Coiled carbonaceous ribbons of size in centimeters from the ca. 1,900–1,800 Ma Negaunee Iron-Formation of Michigan have been described as *Grypania* and interpreted as eukaryotic fossils (Han and Runnegar 1992). Broadly similar fossils have also been found in Mesoproterozoic rocks in North China (Fig. 8.3a), central India, and Montana (Walter et al. 1990; Sharma and Shukla 2009), although some paleontologists interpret the coiled ribbons as cyanobacterial fossils (Sharma and Shukla 2009). Other macroscopic fossils, including discoidal to elliptical carbonaceous compressions that are often described as *Chuaria* and *Tawuia* are known from the ca. 1,800 Ma Changzhougou and Chuanlinggou Formations in northern China (Hofmann and Chen 1981; Zhu et al. 2000), and they could also represent early eukaryotic fossils. The slightly younger (ca. 1,700–1,630 Ma) Tuanshanzi Formation in northern China contains millimeter- to centimeter-size carbonaceous ribbons and blades (Yan 1995; Zhu and Chen 1995). Although these fossils are not preserved in cellular detail, their macroscopic and stable morphologies are consistent with a eukaryotic interpretation.

More convincing Paleoproterozoic eukaryotes, however, come from the micropaleontological record (Knoll et al. 2006). The Changzhougou and Chuanlinggou Formations also contain organic-walled microfossils (or acritarchs) and multicellular microfossils (Yan and Liu 1993). Most of these acritarchs can be described as *Leiosphaeridia.* They are flattened vesicles with simple morphologies (typically smooth-walled) and a diameter of ~ 100 μm. Their affinities are difficult to determine because of the lack of diagnostic features. However, there are several ornamented forms that show morphological complexities indicative of a eukaryotic affinity. For example, *Thecatovalvia annulata* and *Valvimorpha annulata* from the Chuanlinggou Formation are characterized by concentric striations (Yan and Liu 1993), indicating the presence of a eukaryotic cytoskeleton system. Also, vesicles with regular medial splits (Fig. 8.3b, c) from the Changzhougou and Chuanlinggou Formations are consistent with biologically programmed excystment structures of eukaryotic resting cysts, an interpretation that is further supported by their vesicle wall ultrastructures (Lamb et al. 2009; Peng et al. 2009). Another fossil, *Qingshania magnifica*, also from the Chuanlinggou Formation, shows evidence of cellular differentiation—it is characterized by a clavate filament with an expanded terminal cell that was interpreted as a differentiated reproductive cell (Yan and Liu 1993). *Qingshania magnifica* may well be a multicellular eukaryote.

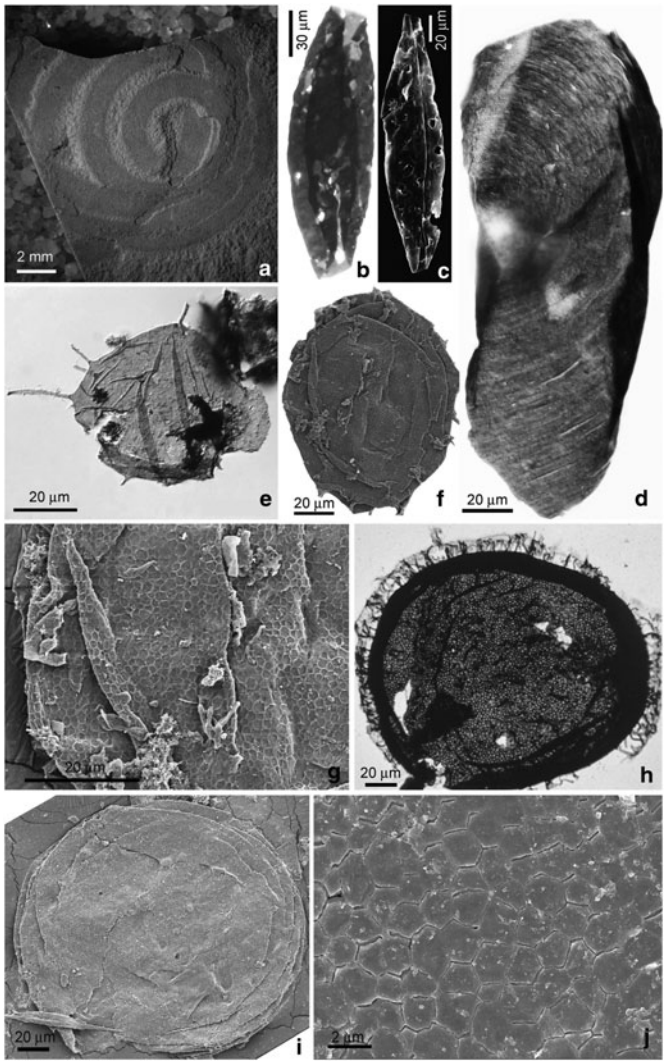


Fig. 8.3 Representative Stage I microfossils. **a** Reflected light photomicrograph of *Grypania spiralis* from the Mesoproterozoic Gaoyuzhuang Formation (ca. 1,500–1,400 Ma) in North China. Photo courtesy of Lifu Tian. **b**, **c** Transmitted light photomicrograph (**b**) and scanning electron photomicrograph (**c**) of medially split and enrolled half vesicles from the Paleoproterozoic Chuanlinggou Formation (~ 1,800 Ma) in North China. Photos courtesy of Yongbo Peng. **d**, **e** Transmitted light photomicrographs of *Valeria lophostriata* (**d**) and *Tappania plana* (**e**) from the Mesoproterozoic Kamo Group in Siberia. Note tubular protrusion (**e**, *right*) and spiny processes (**e**, *left*). Photos courtesy of Konstantin Nagovitsin. **f**–**h** Scanning electron photomicrographs (**f**, **g**) and transmitted light photomicrograph (**h**) of *Shuiyousphaeridium macroreticulatum* from the Paleo-Mesoproterozoic Ruyang Group in North China. **g** Magnified view of (**f**) to show polygonal ornamentation on vesicle surface. **i**, **j** Scanning electron photomicrographs of *Dictyosphaera delicata* from the Ruyang Group in North China. **j** Magnified view to show polygonal pattern on vesicle inner surface

Eukaryotic microfossils are more diverse and less controversial in the Mesoproterozoic. One example is the acritarch *Valeria lophostriata* (Fig. 8.3d), which ranges from ~1,650 Ma (Javaux et al. 2004) to ~700 Ma (Hofmann 1999) but has been recorded from multiple Mesoproterozoic successions in Australia, North China, Siberia, Greenland, and North America (Hofmann and Jackson 1994, 1996; Xiao et al. 1997; Javaux et al. 2001, 2004; Nagovitsin 2009). *Valeria lophostriata* is characterized by vesicle walls with concentric striation. Like *Thecatovalvia annulata* and *Valvimorpha annulata*, *Valeria lophostriata* is best interpreted as a eukaryotic microfossil. Other convincing examples of Mesoproterozoic eukaryotes include the acritarch fossils *Tappania plana* (Fig. 8.3e), *Shuiyousphaeridium macroreticulatum* (Fig. 8.3f–h), *Dictyosphaera delicata* (Fig. 8.3i, j), and *Spiromorpha segmentatus* (Xiao et al. 1997; Prasad and Asher 2001; Javaux et al. 2004; Prasad et al. 2005; Yin et al. 2005; Nagovitsin 2009). Their eukaryotic affinity is supported by their complex vesicle wall microstructures (e.g., tubular protrusions in *T. plana*, spiny processes in *T. plana* and *S. macroreticulatum*, polygonal fields in the vesicle walls of *S. macroreticulatum* and *D. delicata*, and oval to spindle vesicles with transverse stripes and grooves in *S. segmentatus*).

To summarize, although molecular fossils suggest that eukaryotes may have diverged from archaeobacteria and eubacteria in the Neoarchean, it is not until the Paleoproterozoic when morphologically diagnostic eukaryotes enter the fossil record. By the Mesoproterozoic, positively identified eukaryotic fossils are widespread and moderately diverse. Some scientists have contended that eukaryotes did not evolve until 850 Ma (Cavalier-Smith 2010b). However, the above analysis suggests that even a conservative reading of the fossil record implies that the eukaryotic clade must have diverged from the other two domains in the Mesoproterozoic or earlier. It should be pointed out that this represents a minimum estimate, because of the incompleteness of the fossil record and also because the earliest eukaryotes may have been morphologically simple and unrecognizable even if they are preserved in the fossil record. Many modern eukaryotes, particularly single-celled ones, are morphologically simple and volumetrically small. If preserved as fossils, they would be morphologically indistinguishable from bacterial fossils. The majority of Archean and Paleoproterozoic microfossils are morphologically simple. Many of them are featureless filamentous or spheroidal structures, less than a few tens of micrometers in size, preserved in cherts or shales. The default interpretation of these microfossils is prokaryotes. For many, this is not an unreasonable interpretation, particularly if they can be shown as microbial mat builders. For others, however, a eukaryote interpretation is equally plausible, but there is simply not enough morphological information to allow a positive identification. Excluding these morphologically simple fossils from our inventory of eukaryotic fossils means that the divergence time based on unambiguously identified eukaryotes preserved in the fossil record is necessarily a minimum estimate.

What is the evolutionary significance of Stage I eukaryotic fossils? Where is their phylogenetic home in the eukaryotic tree (Fig. 8.1)? Most of these fossils, even if they have some eukaryotic features, do not have distinct features that place them in

one of the six major extant clades. Possibly, many if not all Stage I eukaryotic fossils may represent evolutionary offshoots (or stem group eukaryotes) that branched after the divergence of the eukaryotic domain but before the divergence of the six major extant clades.

Stage II (Late Mesoproterozoic–Early Neoproterozoic): Deep Divergence of Major Eukaryotic Clades

Early Neoproterozoic successions contain a number of fossils that are held over from Stage I. For example, *Valeria lophostriata* is present in the early Neoproterozoic (Hofmann 1999). *Chuaria*, *Tawuia*, and similar fossils are abundant in early Neoproterozoic successions, including the Little Dal Group (850–780 Ma) in northwestern Canada, Chuar Group (~ 750 Ma) in Grand Canyon, and Huainan and Huaibei Groups (ca. 900–740 Ma) in North China (Xiao and Dong 2006).

The Huainan and Huaibei Groups also host macroscopic carbonaceous fossils such as *Sinosabellidites huainanensis*, *Pararenicola huaiyuanensis*, and *Protoarenicola baiguashanensis* (Fig. 8.4a). These are compressed tubular fossils of millimetric diameter and centimetric length. They have closely spaced transverse annulations that superficially resemble metameric segmentation in animals. A few specimens bear poorly defined terminal structures that were interpreted as proboscis-like structures. These and similar fossils (*Parmia anastassiae*) from early Neoproterozoic rocks in the Timan region of Russia (Gnilovskaya et al. 2000) have been interpreted as worm-like animals (Sun et al. 1986), but recent study has shown that they are probably benthic tubular algae with holdfast structures (Dong et al. 2008). Another Neoproterozoic benthic algal fossil with holdfast structures is *Longfengshania stipitata*, which has been described from the Little Dal Group and the Changlongshan Formation (ca. 900–800 Ma) in northern China (Hofmann1985; Du and Tian 1986). *L. stipitata* is characterized by centimeter-size thalli with an ellipsoidal head, a stipe, and a simple holdfast, indicating a degree of morphological differentiation that is consistent with eukaryotic algae, but finds few analogs among modern prokaryotic organisms.

What distinguishes Stage II from Stage I, however, is the appearance of crown group eukaryotes, particularly a number of fossils that can be phylogenetically resolved to one of the six major eukaryotic clades. One of the earliest eukaryotic fossils that have been confidently resolved to an extant algal group is perhaps *Bangiomorpha pubescens* (Fig. 8.4b, c) from the ~ 1,200 Ma Hunting Formation and equivalent rocks in Arctic Canada (Butterfield 2000). This is a multicellular filamentous fossil with differentiated holdfasts and reproductive cells that indicate an affinity with modern bangiophyte red algae (Butterfield 2000). Another multicellular fossil that has been reasonably resolved into the clade of plants is the genus *Proterocladus* from the Svanbergfjellet Formation (ca. 800–700 Ma) in Spitsbergen. *Proterocladus* is characterized by uniseriate and occasionally branching filaments with sparse septa (Fig. 8.4d), and it is interpreted as a coenocytic green alga similar to the modern genus *Cladophora* (Butterfield et al. 1994).

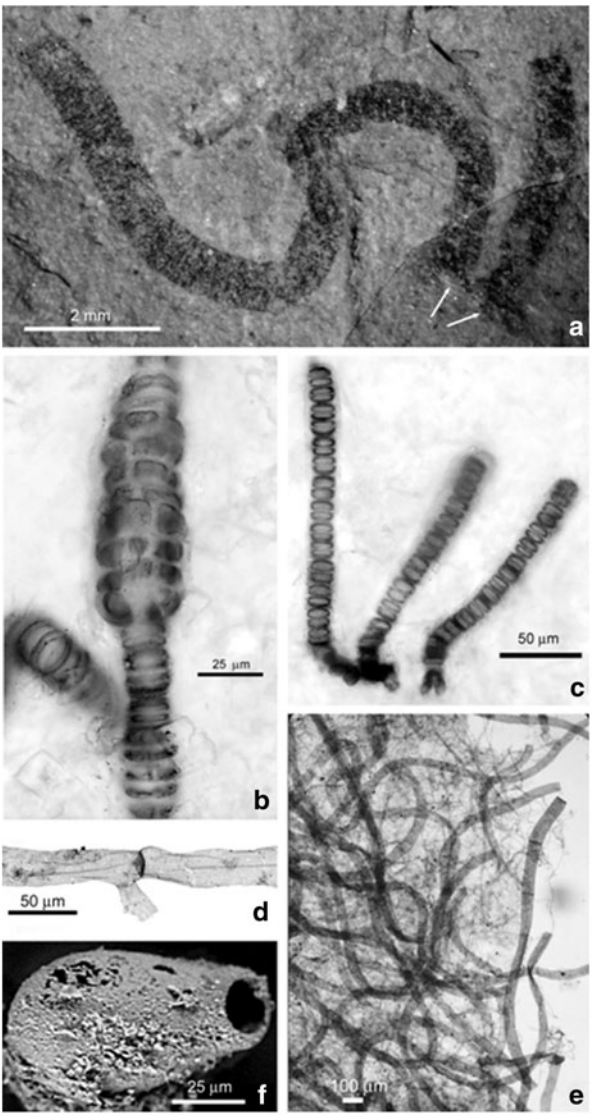


Fig. 8.4 Representative Stage II fossils. **a** Reflected light photomicrograph of *Protoarenicola baiguashanensis*, probably a benthic alga, from the early Neoproterozoic Huaibei Group (900–740 Ma) in North China. *Arrows* point to holdfast structures. Photo courtesy of Xunlai Yuan. **b**, **c** Transmitted light photomicrographs of *Bangiomorpha pubescens*, interpreted as a bangiophyte red alga, from the late Mesoproterozoic Hunting Formation (~ 1,200 Ma) in Arctic Canada. Specimen in (b) shows transition from uniseriate to multiseriate growth, suggesting multiple ontogenetic phases and probably sexual reproduction. Note basal holdfast structures in (**c**). Photos courtesy of Nicholas J. Butterfield. **d** Transmitted light photomicrograph of *Proterocladus major*, interpreted as a coenocytic green alga, from the early Neoproterozoic Svanbergfjellet Formation (800–700 Ma) in Spitsbergen. Photo courtesy of Nicholas J. Butterfield. **e** Transmitted light photomicrograph of *Palaeovaucheria clavata*, interpreted as a xanthophyte alga, from the late Mesoproterozoic Lakhanda Group (~ 1,000 Ma) in southeastern Siberia. Photo courtesy of Andrew H. Knoll. **f** Scanning electron photomicrograph of *Cycliocyrillium simplex*, a vase-shaped microfossil, interpreted as a possible lobose testate amoeba (amoebozoan), from the early Neoproterozoic Chuar Group (~ 750 Ma) in Grand Canyon, Arizona. Photo courtesy of Susannah M. Porter

Chromalveolates may have also diverged by the late Mesoproterozoic. *Palaeovaucheria clavata* (Fig. 8.4e), *Jacutianema solubila*, and related fossils from the ~1,000 Ma Lakhanda Group in southeastern Siberia and the 800–700 Ma Svanbergfjellet Formation in Spitsbergen are characterized by a suite of reproductive features that relate them to xanthophyte algae (Hermann 1990; Butterfield 2004), which are photosynthetic chromalveolates whose plastids were derived from a red alga through secondary endosymbiosis (Hackett et al. 2007).

The Wynniatt Formation (ca. 900–800 Ma) of arctic Canada contains some of the best preserved microfossils in the early Neoproterozoic. One of these microfossils, *Tappania* sp., has been interpreted as a probable fungus, on the basis of secondary cell fusion among hypha-like filaments (Butterfield 2005a). Butterfield also drew morphological comparisons between the Wynniatt microfossil *Cheilofilum hysteriopsis* and hyphomycetous fungi (Butterfield 2005b) and proposed that even the Mesoproterozoic *Tappania plana* could be a possible hyphal fungus (Butterfield 2005a). These interpretations need further substantiation, but if any of them are correct, the opisthokont clade must have diverged by the early Neoproterozoic, implying that its sister group—the amoebozoans—must have also diverged by then.

Indeed, vase-shaped microfossils (Fig. 8.4f), which are widely distributed in early Neoproterozoic successions, including the Chuar Group in Arizona, Visingsö Formation in Sweden, Draken Conglomerate Formation in Spitsbergen, and Eleonore Bay Group in East Greenland (Mus and Moczydłowska 2000; Porter and Knoll 2000), have been interpreted as testate amoebae, possibly representing both euglyphid amoebae (members of the rhizarian clade) and lobose testate amoebae (members of the amoebozoan clade) (Porter et al. 2003; Porter 2006). Almost all vase-shaped microfossils are preserved three-dimensionally (i.e., they are not compressed to two dimensions), suggesting that their tests had some rigidity and some may have had mineralized skeletons (Porter et al. 2003). Mineralized tests of vase-shaped microfossils, together with mineralized scales from the lower Tindir Group (~810–720 Ma) in Alaska and northwestern Canada (Macdonald et al. 2010a, b), indicate that biologically controlled mineralization evolved in some early Neoproterozoic eukaryotes.

To summarize, five of the six major extant eukaryote clades may be represented in the early Neoproterozoic fossil record. Considering the topology of the eukaryotic tree (Fig. 8.1), excavates may have also diverged by the early Neoproterozoic but left no recognizable fossils. Thus, Stage II represents the deep divergence of the six major eukaryotic clades that have survived to the present.

Stage III (Ediacaran): Radiation of Multicellular Eukaryotes

Stage II and Stage III are separated by major environmental crises in the Cryogenian (~720–635 Ma). During the Cryogenian, the Earth experienced at least two global glaciations (known as "snowball Earth events"), when glaciers reached tropical oceans (Hoffman 2009). The documented diversity in Cryogenian successions is

extremely low, possibly because of diversity loss related to extreme glaciations, or poor fossil preservation in Cryogenian rocks, or a combination of both (Xiao 2004b). Regardless, members of the major eukaryote clades that had diverged in Stage II must have survived the Cryogenian glaciations.

Eukaryote diversity rose sharply in the succeeding Ediacaran Period. Both acritarchs and macroalgae reached unprecedented morphological complexity and taxonomic diversity in the Ediacaran Period (Huntley et al. 2006; Knoll et al. 2006; Xiao and Dong 2006). More importantly, the earliest animals emerged in the Ediacaran Period. The late Ediacaran Period is also characterized by a group of enigmatic macroscopic fossils (Ediacara fossils), some of which may well be macrometazoans and macrobilaterians. Toward the end of the Ediacaran Period, animal biomineralization evolved, paving the road for the Cambrian explosion of skeletonized animals.

The Doushantuo Formation (635–551 Ma) in South China provides several taphonomic windows onto the early Ediacaran biosphere. Cherts, phosphorites, and black shales of the Doushantuo Formation preserve some of most extraordinary eukaryote fossils (Yuan et al. 2002). Chert nodules from the lower Doushantuo Formation in the Yangtze Gorges area contain multicellular algae (Xiao 2004a) and numerous spiny acritarchs (Zhang et al. 1998). Taxonomically similar spiny acritarchs have also been reported from other early Ediacaran successions in Australia and Russia (Grey 2005; Vorob'eva et al. 2008, 2009). One of the Doushantuo acritarchs, *Tianzhushania spinosa* (Fig. 8.5a, b), first appears within meters above Cryogenian glacial deposits and has been interpreted as diapause eggs of early animals (Yin et al. 2007), suggesting animal emergence shortly after the Cryogenian glaciations. Phosphorites of the upper Doushantuo Formation at Weng'an, Guizhou Province, have yielded numerous cellularly preserved microfossils (Xiao et al. 1998; 2004; Chen 2005; Hagadorn et al. 2006), including spiny acritarchs (Fig. 8.5c), possible red algae with reproductive structures and thallus differentiation (Fig. 8.5d, e), and probable animal embryos (Fig. 8.5f–h). Although there have been reports of microbilaterians from Doushantuo phosphorites (Chen et al. 2004, 2006, 2009), the interpretation of these fossils is complicated by the taphonomic complexity of phosphatization (Bengtson and Budd 2004; Hagadorn et al. 2006). Black shales of the uppermost Doushantuo Formation preserve more than 20 taxa of macroscopic carbonaceous compression fossils. Most of these compressions can be unambiguously interpreted as multicellular algae (Fig. 8.5i), that show clear evidence of holdfast anchoring, dichotomous branching, apical meristematic growth, and specialized reproductive structures (Xiao et al. 2002), but others are possible or probable macrometazoans (Fig. 8.5j, k) (Zhu et al. 2008).

The late Ediacaran Period is characterized by the Ediacara biota (580–542 Ma), which includes numerous enigmatic macroscopic fossils, possible stem group animals, macrobilaterians, and biomineralizing animals (Narbonne 2005; Fedonkin et al. 2007a; Xiao and Laflamme 2009). Among the best-known Ediacara assemblages, the Avalon assemblage (ca. 580–565 Ma) is probably the oldest, followed by the White Sea (ca. 565–550 Ma) and Nama assemblages (ca. 550–542 Ma). The Avalon assemblage in Newfoundland is dominated by the rangeomorphs (Narbonne 2004). Rangeomorphs are a group of modular organisms (Fig. 8.6a, b)

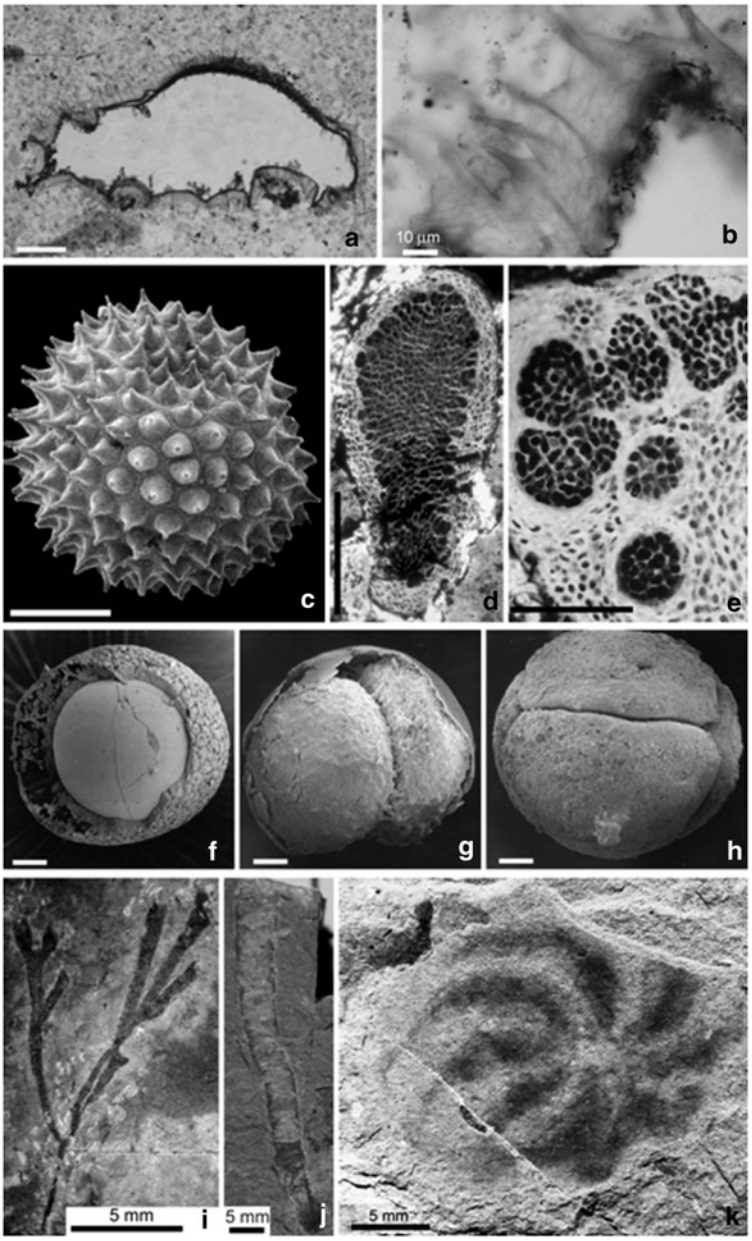


Fig. 8.5 Representative Stage III fossils from the Doushantuo Formation (635–551 Ma) in South China. **a**, **b** Transmitted light photomicrographs of *Tianzhushania spinosa*, a spiny acritarch interpreted as a diapause egg, from lower Doushantuo chert nodules. **b** Magnified view of (**a**) showing spiny processes. **c** Scanning electron photomicrograph of a spiny acritarch, *Meghystrichosphaeridium reticulatum*, from upper Doushantuo phosphorite. **d**, **e** Transmitted light photomicrographs of multicellular algae (possibly florideophyte red algae) with medulla–cortex thallus differentiation

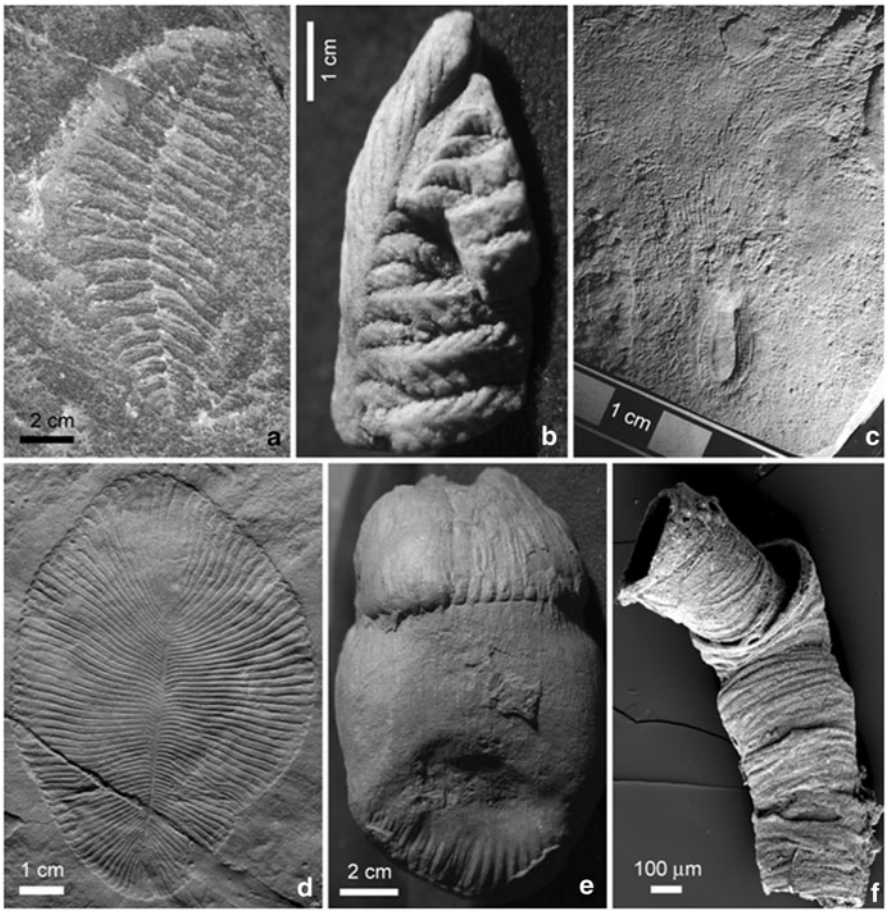


Fig. 8.6 Representative Stage III fossils from the late Ediacaran Period (575–542 Ma). All except (**f**) are reflected light photomicrographs. **a**, **b** Rangeomorphs from Newfoundland (**a**, *Fractofusus*, Avalon assemblage) and Namibia (**b**, *Rangea*, Nama assemblage). **c** *Kimberella quadrata*, a macrobilaterian animal from the White Sea assemblage. Photo courtesy of Andrey Ivantsov. **d** *Dickinsonia costata* from the Flinders Ranges of South Australia. **e** An erniettomorph from the Nama assemblage in Namibia. **f** Scanning electron photomicrograph of *Cloudina riemkeae*, a biomineralizing tubular animal fossil, from the Dengying Formation in South China. Photo courtesy of Hong Hua

---

(**d**) and cell islands interpreted as differentiated reproductive cells (**e**), from upper Doushantuo phosphorite. **f**–**h** Scanning electron photomicrographs of animal egg and blastulas at one- (**f**), two- (**g**), and four-cell (**h**) stages, from upper Doushantuo phosphorite. **i** Reflected light photomicrograph of a dichotomously branching alga, *Konglingiphyton erecta*, from upper Doushantuo black shale. **j** Reflected light photomicrograph of *Sinospongia typica*, possibly a benthic tubular alga or animal, from upper Doushantuo black shale. **k** Reflected light photomicrograph of *Eoandromeda octobrachiata*, a possible diploblastic animal, from upper Doushantuo black shale. Photo courtesy of Maoyan Zhu. Scale bars are 100 µm unless otherwise noted

that are characterized by a self-similar branching system probably adapted for osmotrophic feeding strategies (Laflamme et al. 2009). No macrobilaterian body fossils have been known from the Avalon assemblage. A putative trace fossil has been reported from the Avalon assemblage, but the identity of the trace maker is elusive and it may or may not be a bilaterian animal (Liu et al. 2010). The White Sea assemblage has the highest taxonomic diversity of Ediacara fossils that include both trace and body fossils of convincing macrobilaterians such as *Kimberella quadrata* (Fig. 8.6c) (Fedonkin et al. 2007b; Ivantsov 2009), as well as classic Ediacara fossils such as *Dickinsonia* (Fig. 8.6d) whose phylogenetic affinity remains uncertain (Retallack 2007; Sperling and Vinther 2010). The Nama assemblage includes rangeomorphs (Fig. 8.6b), erniettomorphs (Fig. 8.6e), and biomineralizing animals such as *Cloudina* (Fig. 8.6f), *Namapoikia*, and *Namacalathus* (Grant 1990; Grotzinger et al. 2000; Wood et al. 2002; Hua et al. 2005).

Thus, the fossil record of the Ediacaran Period records a radiation of multicellular eukaryotes, including both algae and animals. The three Ediacara assemblages document a succession of evolutionary events that led to the emergence of macrobilaterians in the late Ediacaran Period and the rise of animal biomineralization toward the end of the Ediacaran Period.

Summary

The fossil record of early eukaryotes is murky, particularly in the Neoarchean and Paleoproterozoic, but becomes increasingly clear in the Mesoproterozoic and Neoproterozoic. Three stages of early eukaryote evolution can be recognized. Stage I (?Neoarchean–Mesoproterozoic) records the origin of eukaryotes and the evolution of stem group eukaryotes. Stage II (late Mesoproterozoic–early Neoproterozoic) represents the divergence of major eukaryote clades that have survived to the present, including the red algae, green algae, chromophyte algae, rhizarians, amoebozoans, and possibly fungi. Stage III (Ediacaran Period) is characterized by the radiation of multicellular eukaryotes, including early animals, macrobilaterians in the late Ediacaran Period, and animal biomineralization toward the end of the Ediacaran Period. This sequence of eukaryote evolution is broadly consistent with predictions from the molecular phylogeny of eukaryotes.

Acknowledgments National Science Foundation, NASA Exobiology and Evolutionary Biology Program, and National Natural Science Foundation of China provided support for my research. Nicholas J. Butterfield, Mikhail Fedonkin, Hong Hua, Andrew H. Knoll, Konstantin Nagovitsin, Yongbo Peng, Susannah M. Porter, Lifu Tian, Xunlai Yuan, and Maoyan Zhu kindly provided photographs used in this chapter.

References

Baldauf SL (2003) The deep roots of eukaryotes. Science 300:1703–1706

Bengtson S, Budd G (2004) Comment on small bilaterian fossils from 40 to 55 million years before the Cambrian. Science 306:1290a-1291a

Brinkmann H, Philippe H (2007) The diversity of eukaryotes and the root of the eukaryotic tree. In: Jékely G (ed) Eukaryotic membranes and cytoskeleton: origins and evolution. Springer, New York, pp 20–37

Brocks JJ, Logan GA, Buick R, Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. Science 285:1033–1036

Brocks JJ, Buick R, Logan GA, Summons RE (2003) Composition and syngeneity of molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Pilbara Craton, Western Australia. Geochim Cosmochim Acta 67:4289–4319

Butterfield NJ (2000) *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. Paleobiology 26:386–404

Butterfield NJ (2004) A vaucheriacean alga from the middle Neoproterozoic of Spitsbergen: implications for the evolution of Proterozoic eukaryotes and the Cambrian explosion. Paleobiology 30:231–252

Butterfield NJ (2005a) Probable proterozoic fungi. Paleobiology 31:165–182

Butterfield NJ (2005b) Reconstructing a complex early Neoproterozoic eukaryote, Wynniatt Formation, arctic Canada. Lethaia 38:155–169

Butterfield NJ, Knoll AH, Swett K (1994) Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. Fossils Strata 34:1–84

Cavalier-Smith T (2010a) Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. Biol Lett 6(3):342–345

Cavalier-Smith T (2010b) Deep phylogeny, ancestral groups and the four ages of life. Philos Trans R Soc Lond B Biol Sci 365:111–132

Chen J (2005) The dawn of animal world. Jiangsu Science and Technology Press, Nanjing

Chen J-Y, Bottjer DJ, Oliveri P, Dornbos SQ, Gao F, Ruffins S, Chi H, Li C-W, Davidson EH (2004) Small bilaterian fossils from 40 to 55 million years before the Cambrian. Science 305:218–222

Chen J-Y, Bottjer DJ, Davidson EH, Dornbos SQ, Gao X, Yang Y-H, Li C-W, Li G, Wang X-Q, Xian D-C, Wu H-J, Hwu Y-K, Tafforeau P (2006) Phosphatized polar lobe-forming embryos from the Precambrian of southwest China. Science 312:1644–1646

Chen J-Y, Bottjer DJ, Davidson EH, Li G, Gao F, Cameron RA, Hadfield MG, Xian D-C, Tafforeau P, Jia Q-J, Sugiyama H, Tang R (2009) Phase contrast synchrotron X-ray microtomography of Ediacaran (Doushantuo) metazoan microfossils: phylogenetic diversity and evolutionary implications. Precambrian Res 173:191–200

Dong L, Xiao S, Shen B, Yuan X, Yan X, Peng Y (2008) Restudy of the worm-like carbonaceous compression fossils *Protoarenicola*, *Pararenicola*, and *Sinosabellidites* from early Neoproterozoic successions in North China. Palaeogeogr Palaeoclimatol Palaeoecol 258:138–161

Du R, Tian L (1986) The macroalgal fossils of the Qingbaikou period in the Yanshan range. Hebei Science and Technology Press, Shijiazhuang

Dutkiewicz A, Volk H, George SC, Ridley J, Buick R (2006) Biomarkers from Huronian oil-bearing fluid inclusions: an uncontaminated record of life before the Great Oxidation Event. Geology 34:437–440

Fedonkin MA, Gehling JG, Grey K, Narbonne GM, Vickers-Rich P (2007a) The rise of animals: evolution and diversification of the kingdom Animalia. Johns Hopkins University Press, Baltimore

Fedonkin MA, Simonetta A, Ivantsov AY (2007b) New data on *Kimberella*, the Vendian mollusc-like organism (White Sea region, Russia): palaeoecological and evolutionary implications. In: Vickers-Rich P, Komarower P (eds) The rise and fall of the Ediacaran Biota. Geological Society of London Special Publications 286, London, pp 157–179

Fehling J, Stoecker D, Baldauf SL (2007) Photosynthesis and the eukaryote tree of life. In: Falkowski PG, Knoll AH (eds) Evolution of primary producers in the sea. Elsevier Academic Press, Burlington, pp 76–107

Gnilovskaya MB, Veis AF, Bekker AY, Olovyanishnikov VG, Raaben ME (2000) Pre-Ediacarian fauna from Timan (Annelidomorphs of the late Riphean). Stratigr Geol Correl 8:327–352

Grant SWF (1990) Shell structure and distribution of *Cloudina*, a potential index fossil for the terminal Proterozoic. Am J Sci 290-A:261–294
Grey K (2005) Ediacaran palynology of Australia. Memo Assoc Austral Palaeontol 31:1–439
Grotzinger JP, Watters WA, Knoll AH (2000) Calcified metazoans in thrombolite-stromatolite reefs of the terminal Proterozoic Nama Group, Namibia. Paleobiology 26:334–359
Hackett JD, Yoon HS, Butterfield NJ, Sanderson MJ, Bhattacharya D (2007) Plastid endosymbiosis: sources and timing of the major events. In: Falkowski PG, Knoll AH (eds) Evolution of primary producers in the sea. Elsevier Academic Press, Burlington, pp 109–132
Hagadorn JW, Xiao S, Donoghue PCJ, Bengtson S, Gostling NJ, Pawlowska M, Raff EC, Raff RA, Turner FR, Yin C, Zhou C, Yuan X, McFeely MB, Stampanoni M, Nealson KH (2006) Cellular and subcellular structure of Neoproterozoic embryos. Science 314:291–294
Han T-M, Runnegar B (1992) Megascopic eukaryotic algae from the 2.1 billion-year-old Negaunee Iron-Formation, Michigan. Science 257:232–235
Hermann TN (1990) Organic world billion year ago. Nauka, Leningrad
Hoffman PF (2009) Pan-glacial—a third state in the climate system. Geol Today 25:107–114
Hofmann HJ (1985) The mid-Proterozoic Little Dal macrobiota, Mackenzie Mountains, north-west Canada. Palaeontology 28:331–354
Hofmann HJ (1999) Global distribution of the Proterozoic sphaeromorph acritarch *Valeria lophostriata* (Jankauskas). Acta Micropalaeontol Sin 16:215–224
Hofmann H, Chen J (1981) Carbonaceous megafossils from the Precambrian (1800 Ma) near Jixian, northern China. Can J Earth Sci 18:443–447
Hofmann HJ, Jackson GD (1994) Shale-facies microfossils from the Proterozoic Bylot Supergroup, Baffin Island, Canada. Paleontol Soc Memoir 37:1–35
Hofmann HJ, Jackson GD (1996) Notes on the geology and micropaleontology of the Proterozoic Thule Group, Ellesmere Island, Canada and North-West Greenland. Geol Surv Canada Bull 495:1–26
Hua H, Chen Z, Yuan X, Zhang L, Xiao S (2005) Skeletogenesis and asexual reproduction in the earliest biomineralizing animal *Cloudina*. Geology 33:277–280
Huntley JW, Xiao S, Kowalewski M (2006) 1.3 billion years of acritarch history: an empirical morphospace approach. Precambrian Res 144:52–68
Ivantsov AY (2009) New reconstruction of *Kimberella*, problematic Vendian metazoan. Paleontol J 43:601–611
Javaux EJ, Knoll AH, Walter MR (2001) Morphological and ecological complexity in early eukaryotic ecosystems. Nature 412:66–69
Javaux EJ, Knoll AH, Walter MR (2004) TEM evidence for eukaryotic diversity in mid-Proterozoic oceans. Geobiology 2:121–132
Javaux EJ, Marshall CP, Bekker A (2010) Organic-walled microfossils in 3.2-billion-year-old shallow-marine siliciclastic deposits. Nature 463:934–938
Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW (2005) The tree of eukaryotes. Trends Ecol Evol 22:670–676
Kirschvink JL, Kopp RE (2008) Palaeoproterozoic ice houses and the evolution of oxygen-mediating enzymes: the case for a late origin of photosystem. Philos T Roy Soc Lond B Biol Sci 363:2755–2765
Knoll AH, Javaux EJ, Hewitt D, Cohen P (2006) Eukaryotic organisms in Proterozoic oceans. Philos T Roy Soc Lond B Biol Sci 361:1023–1038
Laflamme M, Xiao S, Kowalewski M (2009) Osmotrophy in modular Ediacara organisms. Proc Natl Acad Sci U S A 106:14438–14443
Lamb DM, Awramik SM, Chapman DJ, Zhu S (2009) Evidence for eukaryotic diversification in the ~1800 million-year-old Changzhougou Formation, North China. Precambrian Res 173:93–104
Liu AG, McIlroy D, Brasier MD (2010) First evidence for locomotion in the Ediacara biota from the 565 Ma Mistaken Point Formation, Newfoundland. Geology 38:123–126
Macdonald FA, Cohen PA, Dudás FÖ, Schrag DP (2010a) Early Neoproterozoic scale microfossils in the Lower Tindir Group of Alaska and the Yukon Territory. Geology 38:143–146

Macdonald FA, Schmitz MD, Crowley JL, Roots CF, Jones DS, Maloof AC, Strauss JV, Cohen PA, Johnston DT, Schrag DP (2010b) Calibrating the cryogenian. Science 327:1241–1243

Mus MM, Moczydłowska M (2000) Internal morphology and taphonomic history of the Neoproterozoic vase-shaped microfossils from the Visingsö Group, Sweden. Norsk Geologisk Tidsskrift 80:213–228

Nagovitsin K (2009) *Tappania*-bearing association of the Siberian platform: biodiversity, stratigraphic position and geochronological constraints. Precambrian Res 173:137–145

Narbonne GM (2004) Modular construction of early Ediacaran complex life forms. Science 305:1141–1144

Narbonne GM (2005) The Ediacara Biota: Neoproterozoic origin of animals and their ecosystems. Annu Rev Earth Plan Sci 33:421–442

Ogg JG, Ogg G, Goldstein FM (2008) The concise geological time scale. Cambridge University Press, Cambridge

Peng Y, Bao H, Yuan X (2009) New morphological observations for Paleoproterozoic acritarchs from the Chuanlinggou Formation, North China. Precambrian Res 168:223–232

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Müller M, Le Guyader L (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc Biol Sci 267:1213–1221

Porter SM (2006) The Proterozoic fossil record of heterotrophic eukaryotes. In: Xiao S, Kaufman AJ (eds) Neoproterozoic geobiology and paleobiology. Springer, Dordrecht, pp 1–21

Porter SM, Knoll AH (2000) Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. Paleobiology 26:360–385

Porter SM, Meisterfeld R, Knoll AH (2003) Vase-shaped microfossils from the Neoproterozoic Chuar Group, Grand Canyon: a classification guided by modern testate amoebae. J Paleontol 77:409–429

Prasad B, Asher R (2001) Acritarch biostratigraphy and lithostratigraphic classification of Proterozoic and lower Paleozoic sediments (pre-unconformity sequence) of Ganga Basin. India Paleontogr Ind 5:1–151

Prasad B, Uniyal SN, Asher R (2005) Organic-walled microfossils from the Proterozoic Vindhyan Supergroup of Son Valley, Madhya Pradesh, India. Palaeobotanist 54:13–60

Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. Nature 455:1101–1104

Retallack GJ (2007) Growth, decay and burial compaction of *Dickinsonia*, an iconic Ediacaran fossil. Alcheringa 31:215–240

Roger AJ, Simpson AGB (2009) Evolution: revisiting the root of the eukaryote tree. Curr Biol 19:R165–R167, doi:10.1016/j.cub.2008.12.032

Schiffbauer J, Yin L, Bodnar RJ, Kaufman AJ, Meng F, Hu J, Shen B, Yuan X, Bao H, Xiao S (2007) Ultrastructural and geochemical characterization of Archean-Paleoproterozoic graphite particles: implications for recognizing traces of life in highly metamorphosed rocks. Astrobiology 7:684–704

Schulz HN, Jørgensen BB (2001) Big bacteria. Annu Rev Microbiol 55:105–137

Sharma M, Shukla Y (2009) Taxonomy and affinity of Early Mesoproterozoic megascopic helically coiled and related fossils from the Rohtas Formation, the Vindhyan Supergroup, India. Precambrian Res 193:105–122

Sherman LS, Waldbauer J, Summons RE (2007) Methods for biomarker analyses of high maturity Precambrian rocks. Org Geochem 38:1987–2000

Sperling EA, Vinther J (2010) A placozoan affinity for *Dickinsonia* and the evolution of late Proterozoic metazoan feeding modes. Evol Dev 12:201–209

Stechmann A, Cavalier-Smith T (2002) Rooting the eukaryote tree by using a derived gene fusion. Science 297:89–91

Sun W, Wang G, Zhou B (1986) Macroscopic worm-like body fossils from the Upper Precambrian (900–700 Ma), Huainan district, Anhui, China and their stratigraphic and evolutionary significance. Precambrian Res 31:377–403

Ventura GT, Kenig F, Reddy CM, Schieber J, Frysinger GS, Nelson RK, Dinel E, Gaines RB, Schaeffer P (2007) Molecular evidence of Late Archean archaea and the presence of a subsurface hydrothermal biosphere. Proc Natl Acad Sci U S A 104:14260–14265

Vorob'eva NG, Sergeev VN, Chumakov NM (2008) New finds of early Vendian microfossils in the Ura Formation: revision of the Patom Supergroup Age, middle Siberia. Dokl Earth Sci 419A:411–416

Vorob'eva NG, Sergeev VN, Knoll AH (2009) Neoproterozoic microfossils from the northeastern margin of the East European Platform. J Paleontol 83:161–196

Walter MR, Du R, Horodyski RJ (1990) Coiled carbonaceous megafossils from the middle Proterozoic of Jixian (Tianjin) and Montana. Am J Sci 290-A:133–148

Wood RA, Grotzinger JP, Dickson JAD (2002) Proterozoic modular biomineralized metazoan from the Nama Group, Namibia. Science 296:2383–2386

Xiao S (2004a) New multicellular algal fossils and acritarchs in Doushantuo chert nodules (Neoproterozoic, Yangtze Gorges, South China). J Paleontol 78:393–401

Xiao S (2004b) Neoproterozoic glaciations and the fossil record. In: Jenkins GS, McMenamin M, Sohl LE, McKay CP (eds) The extreme proterozoic: geology, geochemistry, and climate. American Geophysical Union (AGU), Washington, pp 199–214

Xiao S, Dong L (2006) On the morphological and ecological history of Proterozoic macroalgae. In: Xiao S, Kaufman AJ (eds) Neoproterozoic Geobiology and Paleobiology. Springer, Dordrecht, pp 57–90

Xiao S, Laflamme M (2009) On the eve of animal radiation: phylogeny, ecology and evolution of the Ediacara biota. Trends Ecol Evol 24:31–40

Xiao S, Knoll AH, Kaufman AJ, Yin L, Zhang Y (1997) Neoproterozoic fossils in Mesoproterozoic rocks? Chemostratigraphic resolution of a biostratigraphic conundrum from the North China Platform. Precambrian Res 84:197–220

Xiao S, Zhang Y, Knoll AH (1998) Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. Nature 391:553–558

Xiao S, Yuan X, Steiner M, Knoll AH (2002) Macroscopic carbonaceous compressions in a terminal Proterozoic shale: A systematic reassessment of the Miaohe biota, South China. J Paleontol 76:347–376

Xiao S, Knoll AH, Yuan X, Pueschel CM (2004) Phosphatized multicellular algae in the Neoproterozoic Doushantuo Formation, China, and the early evolution of florideophyte red algae. Am J Botany 91:214–227

Yan Y, Liu Z (1993) Significance of eucaryotic organisms in the microfossil flora of the Changcheng System. Acta Micropalaeontol Sin 10:167–180

Yan Y (1995) Discovery and preliminary study of megascopic algae (1700 Ma) from the Tuanshanzi Formation in Jixian, China. Acta Micropalaeontol Sin 12:107–126

Yin L, Yuan X, Meng F, Hu J (2005) Protists of the upper Mesoproterozoic Ruyang Group in Shanxi Province, China. Precambrian Res 141:49–66

Yin L, Zhu M, Knoll AH, Yuan X, Zhang J, Hu J (2007) Doushantuo embryos preserved inside diapause egg cysts. Nature 446:661–663

Yuan X, Xiao S, Yin L, Knoll AH, Zhou C, Mu X (2002) Doushantuo fossils: life on the eve of animal radiation. China University of Science and Technology Press, China

Zhang Y, Yin L, Xiao S, Knoll AH (1998) Permineralized fossils from the terminal Proterozoic Doushantuo Formation, South China. J Paleontol 72(Supp 4):1–52

Zhu M, Gehling JG, Xiao S, Zhao Y-L, Droser M (2008) Eight-armed Ediacara fossil preserved in contrasting taphonomic windows from China and Australia. Geology 36:867–870

Zhu S, Chen H (1995) Megascopic multicellular organisms from the 1700-million-year- old Tuanshanzi Formation in the Jixian area, North China. Science 270:620–622

Zhu S, Sun S, Huang X, He Y, Zhu G, Sun L, Zhang K (2000) Discovery of carbonaceous compressions and their multicellular tissues from the Changzhougou Formation (1800 Ma) in the Yanshan Range, North China. Chin Sci Bull 45:841–846

Chapter 9
Endosymbiosis and the Origin of Eukaryotes

Michael F. Dolan

Search for the term "endosymbiosis" in the Web of Science or the Zoological Record, and you will find a long list of some of the most intriguing and intellectually stimulating articles being published these days. The phenomenon of cells living inside other cells has had an enormous impact on the evolution of life. This has occurred on the macroscale with the endosymbiotic origin of numerous phyla of eukaryotes including the primary photoendosymbiotic green and red algae, and the numerous photosynthetic lineages that evolved by incorporating these algae into their cells, including brown algae, diatoms, and dinoflagellates. This process not only introduced new genes to the eukaryotic genome, but also altered that genome in novel ways. Llyod and Timmis (2011) highlighted its significance in writing, "Endosymbiotic transfer of DNA and functional genes from the cytoplasmic organelles (mitochondria and chloroplasts) to the nucleus has been a major factor driving the origin of new nuclear genes, a process central to eukaryote evolution."

Microevolution, or the origin of species, has also been influenced by endosymbiosis. During the early twentieth-century phase of species description of the termite gut protists, for example, numerous species were named for the ectosymbionts or endosymbionts they harbored (Dolan 2002). Photosynthetic species of animals were similarly named as *Elysia chlorotica*, and *Hydra viridis*.

What role did endosymbiosis play in the origin of eukaryotes? The simple answer is—we still do not know. Although Woese's Three Domains are largely supported, it is clear that the Eukaryotes are a fundamentally chimeric lineage containing both, bacterial and archaeal genes, or put another way—that there have been significant bacterial genetic contributions to eukaryotes. The endosymbiotic origins of mitochondria and chloroplasts are well established, but the extent to which other bacterial genomes have contributed to the eukaryotes still needs research. In this chapter, I will review the latest evidence on the question of ancestrally amitochondriate protists, endosymbiosis from an environmental evolution perspective, Searcy's sulfur syntrophy hypothesis, and termite gut flagellates as an extant model for endosymbiosis.

M. F. Dolan (✉)
Department of Geosciences, University of Massachusetts, Amherst, MA 01003, USA
e-mail: pbi@geo.umass.edu

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_9, © Springer Science+Business Media New York 2013

For a period of 25 years between Cavalier-Smith's formation of the Archaezoa hypothesis (1983) and the paper by Hampl et al. (2008) that concludes that all the archaezoans are derived from mitochondriate cells, evolutionary cell biologists and protistologists actively considered the possibility that some extant eukaryotic lineages never had mitochondria—that the mitochondrial symbiosis occurred after the origin of eukaryotes. The Archaezoa included the diplomonads (eg., *Giardia*), the parabasalids (eg., *Trichomonas*), and the oxymonads, and were mostly anaerobic parasites of animals. Most of these organisms have been found to have hydrogenosomes or mitosomes, membrane-bounded organelles that are thought to be the remnants of mitochondria. This conclusion is due to the fact that these organisms harbor genes in their nuclei that code for proteins targeted to the mitochondria in aerobic organisms, or the hydrogenosomes or mitosomes, which lack their own genome, in these cases.

The hydrogenosome, first described in the parabasalid *Tritrichomonas*, was originally thought to derive from an anaerobic bacterium like *Clostridium*, because it used the [FeFe]-hydrogenase of the obligate anaerobes (Lindmark and Müller 1973). However, the gene for this enzyme has subsequently been found in several microscopic green algae and in other anaerobic protists (*Neocallimastix*, *Nyctotherus*), and appears to have been obtained by these eukaryotes from lateral gene transfer (Hampl et al. 2008). This could be an indication of an additional endosymbiont. With Hampl et al.'s (2008) statement that, "we can now say definitively that all known extant eukaryote lineages diverged after the mitochondrial symbiosis", the term "mitochondrion" has experienced an expansion in meaning to include all organelles derived from the primordial proteobacterial endosymbiont. As a result we can now read of "anaerobic mitochondria" (Mentel and Martin 2008).

These molecular phylogenetic results, combined with new geochemical understanding of the Proterozoic oceans have led to a revision of mitochondrial evolution, and a different candidate, or model for the endosymbiont—*Rhodobacter*. The sedimentary sulfur isotope record indicates that the ocean was sulfidic and anoxic for much of the Proterozoic (Canfield 1998). Since it is likely that eukaryotes, or all extant lineages of eukaryotes, evolved in this sulfidic environment, an endosymbiosis based on sulfur syntrophy is more likely than one favoring aerobic respiration. The endosymbiont, as with many bacteria, was quite versatile metabolically, and contained the genes for the later development of eukaryotes in an oxic environment. As Mentel and Martin (2008) nicely summarize, a generalist proteobacterium can survive by aerobic respiration, anaerobic respiration, fermentation, photolithoautotrophy, and photoheterotrophy. How could this sulfur syntrophy evolve?

Although not given much credit for his pioneering work (eg. Mentel and Martin 2008; van der Giezen 2011), Searcy had exposed the weaknesses of the oxygen-consumption model of mitochondrial endosymbiosis, hypothesized a coherent and parsimonious model for a sulfur syntrophy model, and produced original research to support this hypothesis well before the role of sulfur in eukaryote evolution became appreciated. In textbooks it is commonly reported that mitochondria evolved endosymbiotically so that oxygen could be used as the terminal electron acceptor in the host cell. For example, the mitochondrion evolved from "a stable endosymbiont relation with a bacterium, whose oxidative phosphorylation system they subverted

for their own use" (Alberts et al. 2002). However, there are several reasons why this arrangement would not be favored by natural selection. In an oxic environment, the aerobic bacterium would more likely remain an ectosymbiont while the anaerobic host would be inside. Aerobic respiration releases reactive oxygen species that would not favor endosymbiosis. The aerobic bacterium, and putative mitochondrion, would likely survive just as well on its own, and would not be selected to become an endobiont (Searcy 2003).

In contrast to the oxygen-centered view of mitochondrial origins, Searcy proposed a sulfur syntrophy origin. In this scenario, the bacterial endobiont oxidized hydrogen sulfide to sulfur and the archaeal host cell reduced it back to H_2S (Searcy 2003). Although extant examples of such a sulfur syntrophy involving an archaeal host cell have not been found, bacterial consortia such as *Chlorochromatium aggregatum* are well known and widespread.

Searcy went on to discover remnants of this ancient sulfur syntrophy in a ciliate and in vertebrates. He found that *Tetrahymena* mitochondria consumed sulfide, and that the cell's soluble cytoplasm produced sulfide (Searcy 2006). He had previously found sulfur reduction in the cytoplasm of human red blood cells (Searcy and Lee 1998) and found sulfide oxidation in chicken liver mitochondria (Yong and Searcy 2001). His development of the sulfidostat (2006) that can measure sulfide consumption at levels as low as 10^{-12} M, will enhance research in this area.

The sulfur syntrophy hypothesis has not garnered much attention by those pursuing a genome-driven approach, rather than a physiological or microbial ecological one, although it was only one of two hypotheses for the origin of eukaryotes that passed muster in the 2007 Pisani et al. metaanalysis of molecular phylogenies.

Although it is difficult to recreate ancient biospheres and elucidate which lineages of organisms evolved first, there is one example of endosymbiosis that, while unrelated to marine microbes, could shed light on the processes of endobiosis. This is the three-layered symbioses of the wood-eating termites' guts. These anoxic to microoxic gut environments are perhaps the most prolific centers of endosymbiosis in nature today (Dolan 2011). The eukaryotes involved are the parabasalids, the same group in which Lindmark and Müller described the hydrogenosome. There has been a great proliferation of over 400 species of these protists in the termites' guts, and most of them harbor a wide diversity of endosymbionts. In addition to unique bacterial lineages found only in these protists (eg., the Endomicrobia), these parabasalids host members of the bacteroidiales, verrucomicrobia, actinomycetes, and spirochetes.

For example, the protist *Trichonympha agilis* from the termite *Reticulitermes speratus* harbors an endobiont of *Desulfovibrio* (in addition to a second endobiont of the termite-specific bacterial phylum Termite Group 1, now called the Endomicrobia). It contains the genes for sulfate reduction and hydrogenase, and has been hypothesized to serve as a sink for the hydrogen generated by the protists and other bacteria. The same research group has gone even further, and sequenced the entire genome of a bacteroides endobiont in the protist *Pseudotrichonympha grassii* found in the termite *Coptotermes formosanus* (Hongoh 2008). They found that the endobiont can fix nitrogen and recycle the protists nitrogenous waste into amino acids.

Although this microbial community is based on the digestion of cellulose and not a model for the Proterozoic sulfidic ocean, it presents a unique opportunity to study the establishment of endosymbiosis, the loss or transfer of genes to the host cell's nucleus, and the syntrophic integration of endobiont in host cell. Such extant associations can be used to model the ancient cellular symbioses that led to modern eukaryotes.

References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) Molecular biology of the cell, 4th edn. Garland Science, New York

Canefield DE (1998) A new model for Proteozoic ocean chemistry. Nature 396:450–453

Cavalier-Smith T (1983) Endosymbiotic origin of the mitochondrial envelope. In Schwemmler W, Schenck HEA (ed) Endocytobiology II. de Gruyter, Berlin, pp 265–279

Dolan MF (2002) Speciation of termite gut protists: the role of bacterial symbionts. Inter Microbiol 4:203–208

Dolan MF (2011) The wood-eating termite hindgut: diverse cellular symbioses in a microoxic to anoxic environment. In Seckbach J et al. (ed) Anoxia. Springer, New York

Hampl V, Silberman JD, Stechmann A, Diaz-Triviño S, Johnson PJ, Roger AJ (2008) Genetic evidence for a mitochondriate ancestry in the 'amitochondriate' flagellate *Trimastix pyriformis*. PLoS One 3(1):e1383

Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M (2008) Genome of an endosymbiont coupling N_2 fixation to cellulolysis within in protist cells in termite gut. Science 322:1108–1109

Lindmark DG, Müller M (1973) Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. J Bio Chem 248:7724–7728

Llyod AH, Timmis JN (2011) The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. Mol Biol Evol 7:2019–2028

Mentel M, Martin W (2008) Energy metabolism among eukaryotic anaerobes in light of Proterozoic ocean chemistry. Philos Trans R Soc B 363:2717–2729

Searcy D (2003) Metabolic integration during the evolutionary origin of mitochondria. Cell Res 13:229–238

Searcy DG (2006) Rapid hydrogen sulfide consumption by *Tetrahymena pyriformis* and its implications for the origin of mitochondria. Eur J Protistol 42:221–231

Searcy DG, Lee SH (1998) Sulfur reduction by human erythrocytes. J Exp Zoolog 282:310–322

Traba J, Satrústegui J, del Arco (2011) Adenine nucleotide transporters in organelles: novel genes and functions. Cell Mol Life Sci 68:1183–1206

Van der Giezen M (2011) Mitochondria and the rise of eukaryotes. BioScience 61:594–601

Yong R, Searcy DG (2001) Sulfide oxidation coupled to ATP synthesis in chicken liver mitochondria. Comp Biochem Physiol B Biochem Mol Biol 129:129–137

Part IV
A Planet of Animals and Plants

Chapter 10
Epochal Change: Sweltering Climate at the Paleocene–Eocene Boundary (55 Million Years Ago)

Bruce Scofield

The Paleocene–Eocene Thermal Maximum

Some 55.5 million years (Ma) ago, a brief but very intense global warming occurred simultaneously with a massive perturbation of the carbon cycle. This rapid change in climate began abruptly, possibly within 2,000 years, and lasted another 200,000 years. A return to the previous cooler state took another 150,000–200,000 years. This event, the forerunner and the greatest of several other abrupt warming episodes in the early Eocene, is called the Paleocene–Eocene Thermal Maximum (PETM) (Zachos et al. 1993, 2001). The PETM raises two issues in evolution that are of interest to me. First, it illustrates a correlation between the biota and atmospheric gas composition, specifically how microbiota may precipitate rapid global climate change. Second, it appears to mark a time when evolution, particularly mammalian evolution, was modulated by climate change. The PETM has become an analog for today's anthropogenic greenhouse gas-driven global warming and much research is now focused on understanding it.

The PETM was a short-term worldwide event with rapid onset that coincided with significant changes in marine and terrestrial biota. Oceanic warming at the PETM was most severe at middle bathyl through abyssal depths; it was marked by the most dramatic mass extinction of benthic foraminifera in the past 90 Ma. Concurrent were rapid rates of speciation in planktic foraminifera and dinomastigote population explosions. Accelerated marine molluscan extinction rates, probably due to alterations of sea level and changes in oceanic circulation, are also recorded in the fossil record. This loss of marine taxa was closely followed by an increase of diversity in nonmarine molluscan fauna in the western interior of North America. Closely following the PETM was a dispersion of many northern hemisphere plants and animals, including hoofed mammal order (artiodactyl and perissodactyl), and our own order, primates, from central Asia to southern, eastern, and western regions.

B. Scofield (✉)
Department of Geosciences, University of Massachusetts, Amherst, MA, USA
e-mail: bcscofield@juno.com

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_10, © Springer Science+Business Media New York 2013

These changes mark an event referred to as "fundamental faunal" reorganization at the end of the Paleocene.

The primary evidence for the PETM is a pronounced carbon isotope excursion (CIE), a 3–4 parts per thousand (ppt) negative excursion in stable ^{13}C in the composition of marine and terrestrial sediments, which dated to about 55.5 Ma (Dickens et al. 1995). This CIE required a rapid transfer of reduced carbon to the ocean/atmosphere inorganic reservoir amounting to 75–90 % of the organic carbon reservoir, or about 15 % of the methane hydrate reservoir of today. This is a transfer rate of about 600 gigatons (10^9 tons) in fewer than a thousand years. The consequence was a massive addition of carbon to the Eocene ocean/atmosphere reservoir. During the PETM, deep ocean waters are estimated to have warmed rapidly from 11 to 15 °C. High-latitude sea surface temperatures simultaneously rose by about 6 °C. Dissolved ocean O_2 levels decreased; warm ocean water holds less O_2 gas in solution than does cold water. A global episode of calcium carbonate ($CaCO_3$) dissolution in pelagic sediments, at least partially coeval with the PETM and interpreted as a response to ocean warming, occurred more severely in the Atlantic than in the Pacific. The reduction in latitudinal temperature gradients also resulted in reduced atmospheric circulation.

Possible Causes of the Heat Wave Marking the Paleocene/Eocene Transition

A number of hypotheses have been proposed to explain the negative carbon isotope excursion at the Paleocene–Eocene boundary. One concerns plate tectonics. At 61 Ma ago, North Atlantic Igneous Province (NAIP) volcanism began. This volcanism was most active between 54 and 57 Ma ago. Volcanic degassing during this period is thought to have caused CO_2 gas accumulation in the atmosphere related to a long-term warming trend. The sudden thermal maximum itself may have been triggered by the massive flood basalts that were produced as Greenland and Europe separated. These processes opened up the North Atlantic ocean. The volatilization by massive sill intrusions and contact metamorphism of carbon-rich sediments accumulated in basins near the new plate boundary may have been the source of the ^{12}C-enriched methane that was released into the atmosphere and caused global warming (Storey 2007). Further, volcanism in the Caribbean region at about the same time acted to slow warming at low latitudes, due to the injection of atmospheric dust. The dust decreased the contrast between high- and low-latitude sea surface temperatures and amplified the effects of NAIP volcanism. Warm subtropical latitude surface waters, low in oxygen, became more dense leading to downwelling and reordering of ocean circulation. The warmer waters destabilized methane hydrates on continental shelves, which led to the release and mixing of huge quantities of carbon. (Bralower et al. 1997) With methane in the water column, aerobic methanotrophy in bacterial oxidation produces CO_2 as an end product. Higher CO_2 concentration in seawater lowers ocean pH and leads to dissolution of carbonate rocks.

The dramatic carbon isotope excursion may have been alternatively caused by cometary impact. The hypothetical impact of a comet 10 km in diameter is suggested to have delivered the ^{12}C-enriched carbon that led to global warming (Kent et al. 2003). Evidence for this impact are kaolinite-rich shelf sediments in sediment cores drilled along the Atlantic coastal plain of New Jersey; they contain impact ejecta. The impact and its accompanying dust blanket are hypothesized to have generated geologically instantaneous effects on atmosphere and the oceans that initiated the onset of the carbon isotope excursion at the Earth's surface.

Submarine seismicity, volcanic and gravitational slumping, or a drop in pressure may also have led to slope failure on the continental margins (Bains et al. 1999). The destabilizing of coastal seabed may have resulted in the release of methane hydrates, a large source of ^{12}C-enriched carbon. A study using five genera of foraminifera analyzed isotope data for ODP 690 (Weddell Sea) that shows the CIE was preceded by sea surface warming of 2 °C. The CIE apparently started at the sea surface and moved downward. A decrease in surface ^{18}O indicates that warming preceded the decrease in ^{13}C. (The impact or erosional scenarios would be supported if the decrease of ^{13}C preceded the decrease of ^{18}O.) The onset of peak thermocline warming was correlated with the onset of CIE at the surface (Thomas et al. 2002).

The PETM was not the only dramatic global warming episode that occurred in the early Eocene epoch. Several other hyperthermal events have been recorded, including one at 53.6 Ma ago called the Eocene Thermal Maximum 2 (ETM2) or ELMO. Foram calibrations of marine cores and cyclic sedimentary patterns reveal that these climate changes may have been paced by variations in the Earth's orbit, notably eccentricity. With some adjustments in the dating of these events (called tuning), the lag between the PETM and ELMO is in pace with the 405,000-year-long eccentricity cycle at maxima, which accentuates extreme seasonal contrasts. The result would raise ocean temperatures at intermediate latitudes that would tend to trigger methane hydrate release to the atmosphere from the low temperature deep ocean. Apparently, such maxima follow prolonged minima during which methane hydrate reserves accumulated (Lourens et al. 2005).

Methane and Climate—The Clathrate Gun Hypothesis

The largest reserve of organic carbon on Earth is produced by methanogenic archaebacteria. The archaebacterium *Methanoculleus submarinus*, found deep in marine sediments, is an anaerobic autolithochemotroph that metabolizes the gases CO_2 and hydrogen (H_2) into methane (CH_4) and H_2O. The metabolic product of these bacteria, biogenic methane, is stored in ocean sediments and permafrost in crystalline structures. The clathrate gun hypothesis (Kennet et al. 2003) proposes that increases in the temperature of intermediate water leads to destabilization of methane hydrates on upper continental slopes. Warming intermediate waters induce sediment slump and debris flow that lead to methane release into the ocean/atmosphere system and create

a greenhouse effect. Continuous rise in temperatures and changes in sea level sustain the release of methane from seafloor methane hydrate reservoirs; such positive feedback results in rapid, significant global warming episodes. The residence time of atmospheric methane is fewer than 10 years. After about 10 years, atmospheric methane oxidizes into CO_2, another greenhouse gas. Paleoclimate data record biogenic methane increases at major transitional points in Earth's history, both at the PETM boundary and at the end of the more recent Pleistocene glaciations.

Most of the Earth's methane is biogenic, released as metabolic end products by methane-producing archaebacteria. Methane produced beneath the seafloor becomes trapped in clathrates of ice—microscopic crystalline lattice-like structures. Clathrates are solids that are lighter than water and are kept in place by sediments. They hold up to 160 times the amount of methane molecules in the free gas. They exist in vast quantities at about 0 °C and remain frozen up to 15 °C under pressure between 300 and 2,000 m beneath the seafloor. The total amount of methane stored at present in permafrost and under the seafloor is estimated to be between 5,000 to 20,000 gigatons. Methane trapped in marine sediments in hydrate form constitute such an immense carbon reservoir today that it must be considered a dominant factor in estimating energy resources beyond those of oil, natural gas, or coal. According to the US Geological Survey, global amounts of carbon bound in methane gas hydrates is conservatively estimated to total twice the amount of carbon to be found in all known fossil fuel reserves on Earth. That similar reserves of methane hydrates existed prior to the PETM and were destabilized by temperature changes in ocean currents is a reasonable assumption.

Methanogens, prokaryotic microorganisms, are classified as archae or archaebacteria (Margulis and Chapman 2010). Poisoned by oxygen, they are strict anaerobes. They grow in the total absence of oxygen or in consortia with other bacteria that remove oxygen in the immediate vicinity of the methanogens in oxic environments. Methanogens dwell deep in anoxic ocean floor sediments on continental margins. Their chemoautotrophic metabolism takes place in localities where seashells, animal respiration, limestone, and other forms of calcium carbonate supply carbon in the form of CO_2. Hydrogen from seafloor basalt or bacterial metabolism such as fermentation is provided to methanogens that use the gas as source of energy and electrons. Methane is released by methanogens as their waste product.

$$\rightarrow CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O \text{ (CO_2 reduction)}$$

$$\rightarrow CH_3COOH \rightarrow CH_4 + CO_2 \text{ (acetate fermentation)}$$

A subseafloor methanogen, *Methanoculleus submarinus* (also called Nankai-1), may be the major source of methane in subaqueous methane hydrates. This bacterium was found in deep marine sediments in the Nankai Trough off the eastern coast of Japan. The bacterium is an irregularly shaped coccus. Flagella were indirectly detected but not observed to rotate. The average cell diameter is 0.8–2.0 μm (Mikucki et al. 2003).

The atmosphere concentration of methane, a significant greenhouse gas, apparently oscillates with rapid climate change events. It rapidly increases in a warming climate with a very small lag behind temperature and it both affects and is affected

by climate change. Methane in permafrost underlies 20 % of the Earth's seafloor. Methane that accumulates in clathrates during cold periods is explosively released at warmings; it accelerates the warming process. Methane release to the atmosphere occurs in response to changes in sea level, ocean temperature and pressure, and in response to earthquakes, slumps, and slides. Positive feedback occurs when methane loss from hydrates accelerates further release.

Support for the hypothesis that methane hydrates precipitated the global warming of the PETM is moderately strong. Methane releases have tended to be rapid (decades to centuries), not gradual. Methane hydrates occur close to the ocean floor and are vulnerable to intermediate water temperature changes. An increase in temperature of 1 °C is theoretically equivalent to the effect of a 100-m rise in sea level; a decrease in temperature of 1 °C is theoretically equal to a 100-m drop in sea level. However, methane released from the sea floor has not been shown to pass through the water column and paleoclimate data that show that atmospheric methane directly initiated climate change are not available. Precisely synchronous methane hydrate release with temperature increase has not been satisfactorily demonstrated. For these reasons, the methane hydrate hypothesis for global warming is not universally accepted.

Biotic Changes

The PETM coincided with most severe mass extinction of benthic foraminifera in the past 90 Ma, along with accelerated marine molluscan extinction rates and loss of other taxa. About 50 % of the extinctions of the late Paleocene epoch were probably caused by a combination of highly saline deep water and oxygen gas depletion in warmed waters. The event was most severe at middle bathyal through abyssal depth and was short-term, geologically rapid, and worldwide. This extinction event is dated to about 55.5 Ma at the top of Chron C25, very close to the CIE. Rapid rates of speciation in planktic foraminifera and high population densities of dinomastigotes occurred in the beginning of the Eocene epoch. Short-term eutrophication also probably caused a turnover in foraminifera species. Warmer waters favored larger foraminifera over corals as the main carbonate producers on carbonate platforms at lower latitudes (Scheibner et al. 2005). In the western interior of North America, where much of the evidence for biotic changes at the PETM are found, an increase of diversity in nonmarine molluscan fauna has been detected by paleontological analysis.

Dispersal of northern hemisphere land mammals and plants has been dated to the PETM. Samples from the Lincha Formation in Hunan Province, China, document the first appearance of early hyanenodontids, perissodactyls, artiodactyls, and primates in the fossil record to about this time. A study using isotope and magnetostratigraphy and quantitative biochronology offers modest support for the presence of such taxa in Asia before dispersal to North America and Europe (Bowen et al. 2002). Some 10,000 years after the PETM, animals that are smaller than their immediate descendants are found among the earliest North American fossils of these lineages. Dwarf animal taxa and small-leaved plants dominated, indicating a dry climate. The

crocodile-like reptile *Chamosaurus* and the proprimate mammal *Plesiadapis* went extinct at this time. Evidence of rapid and fundamental faunal reorganization, influx and replacement of archaic by modern species, has been documented in the Clarks Fork–Bighorn Basin of western North America, where the complete sequence of lower Cenozoic sediments known is found. Prior to the PETM in this region, there was a high degree of faunal cosmopolitanism that extended over a wide range with high diversity and low dominance. This ended with the CIE, where postextinction fauna show low diversity and high dominance (Gunnell 1998).

During the time of the PETM, a few important orders of modern mammals (e.g., ungulate and primates) seemed to have evolved very rapidly, in a few hundred or thousand generations. Evidence for such rapid evolution in concert with global climate change raises the question of the relationship between the two. A study of North American mammalian diversity from the late Paleocene through the Pleistocene with a resolution of 1.0 Ma did not show a significant correlation between broad patterns of mammalian evolution and climate change, though there is evidence of higher rates of diversity around 55 Ma (Alroy et al. 2000). Molecular models and the neutral theory of evolution, both emphasizing regular endogenous drivers of change, suggest that the rates of evolution are mostly constant and that life evolves continuously in response to biotic interactions (Gingerich 2006). However, dramatic rapid environmental change described here draws attention to external factors and correlation of evolution with environment. Profound effects on evolution apparently occurred during times of rapid climate change at the Paleocene–Eocene boundary.

References

Alroy J, Koch PL, Zachos JC (2000) Global climate change and north American mammalian evolution. Paleobiology 26(4 Suppl):259–288

Bains S, Corfield R, Norris R (1999) Mechanisms of climate warming at the end of the Paleocene. Science 285:724–727

Bowen GJ, Clyde WC, Koch PL, Ting S, Alroy J, Tsubamoto T, Wang Y, Wang Y (2002) Mammalian dispersal at the Paleocene/Eocene boundary. Science 295:2062–2065

Bralower TJ, Thomas DJ, Zachos JC, Hirschmann MM, Röhl U, Sigurdsson H, Thomas E, Whitney DL (1997) High resolution records of the late Paleocene thermal maximum and circum-Caribbean volcanism: is there a causal link? Geology 25:963–966

Dickens GR, O'Neil JR, Rea DK, Owen RM (1995) Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. Paleoceanography 10:965–971

Gingerich PD (2006) Environment and evolution through the Paleocene-Eocene thermal maximum. Trends Ecol Evol 21(5):246–253

Gunnell GF (1998) Mammalian faunal composition and the Paleocene/Eocene Epoch/series boundary: evidence from the northern bighorn basin. In: Aubry M-P, Lucas S, Berggren WA (eds) Late Paleocene-early Eocene climatic and biotic events in the marine and terrestrial records. Columbia University Press, New York

Kennett James P, Cannariato, Kevin G, Hendy, Ingrid L, Behl, Richard J (2003) Methane hydrates in quaternary climate change: the clathrate gun hypothesis. American Geophysical Union, Washington DC (ISBN 0-87590-296-0)

Kent DV, Cramer BS, Lanci L, Wang D, Wright JD, Van der Voo R (2003) A case for a comet impact trigger for the Paleocene/Eocene thermal maximum and carbon isotope excursion. Earth Planet Sci Lett 211:13–26

Lourens LJ, Sluijs A, Kroon D, Zachos JC, Thomas E, Röhl U, Bowles J, Raffi I (2005) Astronomical pacing of late Palaeocene to early Eocene global warming events. Nature 435:1083–1087

Margulis L, Chapman MJ (2010) Kingdoms and domains: an illustrated guide to the phyla of life on Earth. Elsevier, San Diego

Mikucki JA, Liu Y, Delwiche M, Colwell FS, Boone DR (2003) Isolation of a methanogen from deep marine sediments that contain methane hydrates, and description of *Methanoculleus submarinus* sp. nov. Appl Environ Microbiol 69:3311–3316

Scheibner SRP, Marzouk AM (2005) Turnover of larger foraminifera during the Paleocene-Eocene thermal maximum and paleoclimatic control on the evolution of platform ecosystems. Geology 33(6):493–496

Storey M, Duncan RA, Swisher CC 3rd (2007) Paleocene-Eocene thermal maximum and the opening of the northeast Atlantic. Science 316(5824):587–589

Thomas DJ, Zachos JC, Bralower TJ, Thomas E, Bohaty S (2002) Warming the fuel for the fire: evidence for the thermal dissociation of methane hydrate during the Paleocene-Eocene thermal maximum. Geology 30:1067–1070

Thomas E (1998) Biogreography of the late Paleocene benthic foraminferal extinction. In: Aubry M-P, Lucas S, Berggren WA (eds) Late Paleocene-Early Eocene climatic and biotic events in the marine and terrestrial records. Columbia University Press, New York

Zachos JC, Lohmann KC, Walker JCG, Wise SW (1993) Abrupt climate change and transient climates in the paleogene: a marine perspective. J Geol 100:191–213

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292:686–693

Chapter 11
Insight into Speciation From Long-term Research on Darwin's Finches

Peter R. Grant and B. Rosemary Grant

Darwin's finches on the Galápagos Islands are a classical example of adaptive radiation. A total of 13 species have been formed from a single ancestral species that colonized the Galápagos 2–3 million years ago. Each species occupies a distinct and unique feeding niche. For evolutionary biologists, they pose a challenge. What is the explanation for the evolution of such a variety of species in a relatively short space of time? What governs the rate of speciation and the directions of evolution? Answers to these questions are of general significance in the context of understanding why the world is so rich biologically.

Our research has been directed towards answering these questions. Speciation is the process by which two species form from one, and we would like to know how it happens. To find out, we have concentrated on members of the ground finch genus *Geospiza* because they are generally common and easy to capture for measurement, blood sampling for DNA analysis, and for marking uniquely so that after being released they can be identified and their behavior recorded. We have studied finches throughout the archipelago but concentrated mainly on the island of Daphne Major because it is small (0.34 km^2) and has never been disturbed by human settlement or introductions of alien species. By following the fates of marked birds for more than 30 years, we have discovered several facts that are relevant to questions of species formation. Here, we provide a summary of the major findings. A longer treatment is given in a book (P. R. Grant and B. R. Grant 2008, *How and Why Species Multiply*. Princeton University Press, Princeton, NJ).

The first discovery is that evolution occurs when the environment changes. In 1977, a severe drought affected the whole archipelago. More than 80 % of the medium ground finches (*G. fortis*) on Daphne died, and mortality was size-selective: large birds, especially those with large beaks, survived much better than the rest. They survived by being able to crack open the large and hard woody fruits of *Tribulus*

P. R. Grant (✉) · B. R. Grant
Department of Ecology and Evolutionary Biology, Princeton University, Princeton, USA
e-mail: prgrant@princeton.edu

B. R. Grant
e-mail: rgrant@princeton.edu

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_11, © Springer Science+Business Media New York 2013

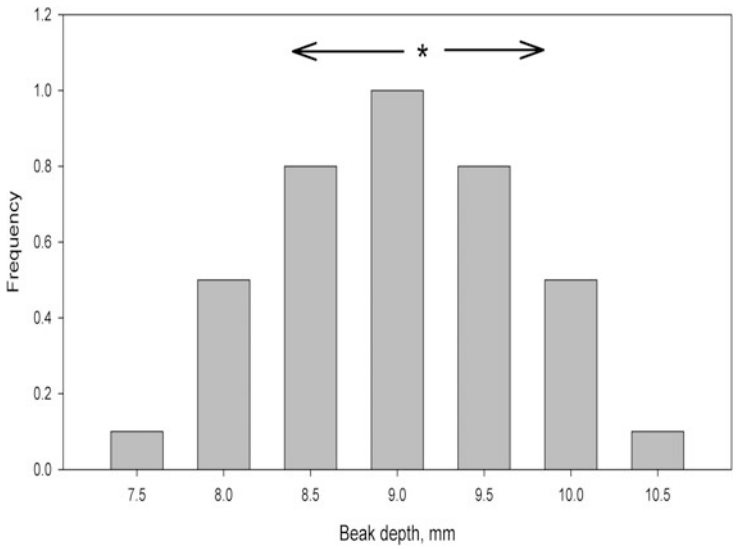


Fig. 11.1 Schematic diagram of the most extreme evolutionary changes in beak depth in response to natural selection; to large size in 1978 and to small size in 2005. The long-term average is indicated by an asterisk, and extreme directional changes are indicated by arrowheads

cistoides that remained in relative abundance after almost all of the small and soft seeds of many other plant species had been consumed. Natural selection gave rise to an evolutionary change in the next generation because beak size is heritable; offspring strongly resemble their parents in beak and body size traits owing to the transmission of genes from parents to offspring.

Second, competition with other species for food influences the direction of evolution. This discovery was made in 2005. In 1982–1983, an exceptionally long El Niño event enabled a population of the large ground finch (*G. magnirostris*) to become established on the island. The numbers gradually built up, reaching a peak just prior to a severe drought in 2003–2004 that was comparable in severity to the drought of 1977. Most finches died, but now, in contrast to 1977, large-beaked members of the medium ground finch population were outcompeted for a limited supply of *Tribulus* fruits by the large ground finch, and they died at a higher rate than birds with smaller beaks. Natural selection occurred, the average beak size fell, and because beak size is heritable, the next generation had small beaks. This is an example of the well-known evolutionary principle of character displacement, in which two species diverge morphologically as a result of competition between them. Thus, over a period of several decades, we have witnessed both selection and evolution oscillating between extremes of large and small beak sizes (Fig. 11.1).

The example of character displacement throws light on what happens during speciation when two populations, having diverged in separate locations such as islands, come together: they compete and diverge further, thereby increasing the chances of long-term coexistence. Finches on Daphne also hybridize, rarely but repeatedly.

This is the third major discovery from our long-term research, and it adds further to an understanding of speciation. Introgressive hybridization has the potential to reveal the genetic factors that usually prevent newly formed species from exchanging genes. However, we discovered that in fact there are no genetic barriers to the exchange of genes between ground finches on Daphne. The medium ground finch hybridizes (rarely) with two other species, the small ground finch (*G. fuliginosa*), and the cactus finch (*G. scandens*), but not the much larger *G. magnirostris*. We found that hybrids survived just as well as the parental species but only when the food supply was suitable for birds of intermediate beak size, i.e., not dominated by large and hard seeds. Moreover, they had no difficulty in obtaining mates, producing fertile eggs, and fledging young.

Our observations on the breeding of finches revealed how mates are chosen and why hybridization occasionally occurs. Both, song and beak morphology are important cues used in mate choice. They differ between species and hence constitute important components of the behavioral barrier to interbreeding. Experiments performed by Robert Bowman on finches in captivity many years ago showed that sons learn their song from their fathers by imprinting on his song at an early age. Thus, the barrier to interbreeding is made up of genetic factors governing the development of species-specific beak sizes and shapes, and culturally transmitted factors governing song production and mate preferences. The normal process of imprinting is occasionally perturbed, for example, when a bird learns the song of another species instead of his father's song. When that happens, hybridization can occur, and it may also lead to backcrossing of the hybrid to one of the parental species in the next generation.

The insight this provides into the process of speciation is that behavioral barriers to interbreeding arise first, at least in these species and probably many other species of birds. Only later do genetic factors that prevent interbreeding species from producing viable and fertile offspring arise by mutation. Hybridization provides an additional insight into when speciation might collapse and when it does not. Speciation may be reversed if there are no disadvantages to hybridization, for example, if the hybridizing species are ecologically similar, whereas it will be sustained if ecological differences between the newly formed species are large.

With regard to the other component of the barrier, beak size, molecular genetic work by Arhat Abzhanov and colleagues has begun to identify the genetic factors functioning in development. Variation in the timing and intensity of expression of two genes (*Bmp4* and *CaM*) early in development creates differences in beak size and shape between species with deep beaks such as the large ground finch, and others with more pointed beaks such as the cactus finch. The genes produce signaling molecules at about day 6 of a 12-day period of embryonic development, and another set of molecules produced by three other genes have similar effects a little later.

In summary, modern molecular genetic research has now been combined with long-term ecological research to illuminate evolutionary processes that Charles Darwin could only guess at, a century and a half ago. The research will be broadly

beneficial to evolutionary biology if it helps other biologists in Galápagos, continental Ecuador, and elsewhere to understand how their study organisms have evolved.

Reference

Grant PR, Grant BR (2008) How and Why Species Multiply. Princeton University Press, Princeton

Chapter 12
Ecological Selection and the Evolution of Body Size and Sexual Size Dimorphism in the Galapagos Flightless Cormorant

Carlos A. Valle

Introduction

Body size is a key life history trait of the ecology of a species (Schmidt-Nielsen 1984; LaBarbera 1989) that frequently is used as both proxy and an indicator of a species' ecological niche within the ecological community (Wilson 1975; Ricklefs and Cox 1977; Ricklefs and Travis 1980; Wiens 1982; Miles and Ricklefs 1984; Peters 1986; White et al. 2007). Within species, however, the sexes frequently differ substantially and consistently in body size, a biological aspect known as sexual size dimorphism (SSD) (Darwin 1871; Anderson 1994; Fairbairn 1997; Blanckenhorn 2000; Fairbairn et al. 2007). The understanding of the evolution of SSD is central to evolutionary ecology; however, debate continues on identifying the most important factors and mechanisms driving its evolution (Anderson 1994; Fairbairn 1997; Abouheif and Fairbairn 1997; Fairbairn et al. 2007).

Several hypotheses involving either sexual or ecological selection or both have been advanced as adaptive explanations for the evolution and maintenance of SSD (Darwin 1871; Selander 1972; and reviews by Shine 1989; Hedrick and Temeles 1989; Anderson 1994; Fairbairn 1997; Fairbairn et al. 2007). Among species where males are larger than females, sexual selection hypotheses include both directional sexual selection for males with larger body size in the competition for mates (the intrasexual competition hypothesis; Darwin 1871; Selander 1972; Clutton-Brock et al. 1977; Alexander et al. 1979; Berry and Shine 1980; Jehl and Murray 1986; Abouheif and Fairbairn 1997; Cox et al. 2003 but see Karubian and Swaddle 2001; Wikelski 2005) or due to female mate choice for larger male size (intersexual selection hypothesis; Lande and Arnold 1985; Szekely et al. 2004; Wikelski and Trillmich 1997; Wikelski 2005). Ecological selection hypotheses on the other

C. A. Valle (✉)
Galapagos Academic Institute for the Arts and Sciences,
Colegio de Ciencias Biológicas y Ambientales,
Universidad San Francisco de Quito, Quito, Ecuador
e-mail: cvalle@usfq.edu.ec

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,*
Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8_12, © Springer Science+Business Media New York 2013

hand suggest that SSD has evolved and is maintained through divergent ecological selection between the sexes and they have taken several forms. The intersexual niche divergence, also known as the "dimorphic niche" or "differential niche utilization" hypothesis (Darwin 1871; Selander 1966, 1972; Lande 1980; Slatkin 1984; Shine 1989; Herrel et al. 1999; Butler et al. 2000) states that SSD evolved through disruptive selection on body size for ecological divergence between the sexes due to intersexual competition for ecological resources (e.g., food). Such a pattern of disruptive selection leading to intersexual ecological divergence would eventually produce an intersexual niche partitioning or the broadening of the ecological niche of the species. Another related ecological hypothesis known as the reproductive role division hypothesis (Hedrick and Temeles 1989) proposes that SSD arises from the sexes specializing on different reproductive roles, which lead to differential selection regimes between the sexes (Wiklund and Karlsson 1988; Hedrick and Temeles 1989; Shine 1989). The female fecundity selection hypothesis emphasizes a relationship between reproductive (fecundity) selection and female body size when there is a reproductive advantage either for smaller body size (i.e., smaller females gain an advantage through earlier reproduction, see Downhower 1976; Ralls 1976; Price 1984) or larger body size (i.e., larger females lay larger eggs, a greater number of eggs or offspring, or increased reproductive frequency; Darwin 1871; Williams 1966; Shine 1988, 1989; Fairbairn and Shine 1993; Fairbairn 1997). Lastly, the secondary ecological dimorphism hypothesis (Shine 1989; Anderson 1994; Fairbairn 1997; Stephens and Wiens 2009) proposes that SSD, once evolved by sexual or fecundity selection, is subsequently exaggerated or reinforced by divergent (disruptive) natural selection due to intersexual ecological competition.

Furthermore, biologists have found in a large variety of animals that the extent of SSD in groups of species within taxa including genera and families usually covaries allometrically with each species' mean body size (Rensch 1960; Ralls 1976; Clutton-Brcok et al. 1977; Jehl and Murray 1986; Shine 1989; Parker 1992; Fairbairn and Shine 1993; Fairbairn 1997; Szekely et al. 2000; Serrano-Meneses and Szekely 2006; Fairbairn et al. 2007; Stephens and Wiens 2008). That pervasive evolutionary ecological trend known as Rensch's rule (Rensch 1960; Reiss 1986; Abouheif and Fairbairn 1997; Szekely et al. 2004; Fairbairn 2005) established that SSD, measured as the ratio of male size to female size, increases with body size (hyperallometry) among species when the male is the larger sex and decreases with the increase in body size (hypoallometry) when the female is the larger sex.

Several hypotheses fundamentally relying upon sexual selection acting on a varied set of conditions attempt to explain the evolution of allometry between SSD and body size (e.g., see review Fairbairn 1997). One hypothesis suggests that interspecific variation in SSD within a clade results from a correlated response of female size to sexual selection on male size due to positive genetic correlation between the sizes of the sexes (Lande 1980; Lande and Arnold 1985; Anderson 1994; Fairbairn and Preziosi 1994; Reeve and Fairbairn 1996). The evolution of such allometric relation between SSD and body size has also been explained by stabilizing selection when the sexes experience differential selection optima for body size (Ralls 1976; Clutton-Brock et al. 1977; Slatkin 1984; Anderson 1994) and as the result of an association

between selection for secondary sexual characters and SSD (Rensch 1960; Anderson 1994; Reiss 1986; Webster 1992). Another hypothesis considers that SSD results from functional constraints on body size (Peters 1983; Schmidt-Nielsen 1984), the allometry for SSD can evolve through sex differences in phenotypic variance for body size (Leutenegger and Cheverud 1982; Cheverud et al. 1985), because large size by itself promotes SSD (Clutton-Brock et al. 1977; Leutenegger 1978; Shine 1979; Gaulin and Sailer 1984), and an opportunity to monopolize mates favors simultaneously both larger body size and SSD (Webster 1992). However, the mechanisms producing this pattern remain controversial (Reiss 1986; Anderson 1994; Fairbairn et al. 2007); these hypotheses remain untested and no general explanation for the evolution of SSD has been established (Fairbairn 1997).

The Galapagos Flightless cormorant (*Phalacrocorax harrisi*), a Galapagos endemic species displays a pronounced male-biased SSD (Livezey 1992; Valle 1994). Although the Flightless cormorant SSD has been emphasized by several authors (e.g., Snow 1966; Harris 1979; Tindle 1984; Snow and Nelson 1984), data on body size and morphometrics was almost nonexistent (i.e., the weight of eight individuals measured by occasional visitors to the Islands; Snow 1966). These authors also speculated that the Galápagos cormorant's remarkable sexual dimorphism in body size has resulted from intersexual competition for food resources leading to a feeding niche partitioning between the sexes. However, before the present study, the only in-depth analysis of the Flightless cormorant body size and morphometric traits was that conducted by Livezey (1992) that was mainly based on museum specimens.

Here, I hypothesize that the exaggerated SSD in the Flightless cormorant is simply accounted by its largest body size among phalacrocoracids, and I use the comparative approach, including other species of cormorants, to test this hypothesis. The test supported the hypothesis and further showed that cormorants (Phalacrocoracidae) do not conform to Rensch's rule either. On the other hand, for the Flightless cormorant, there is no evidence supporting the suggestion that sexual selection has caused the evolution of their larger body size or their remarkable SSD. Therefore, I hypothesize instead that their gigantism and extreme SSD are both novel traits that evolved in situ following colonization of the Galapagos and that ecological selection, rather than sexual selection, drove the evolution of these two character states. Further, I argue that their greater size and SSD is driven by ecological factors due to disruptive intersexual selection involving stronger directional selection for larger size in males coupled with genetic correlation for body size between the sexes that leads to an increase in size in females as correlated response to selection for larger size in males.

Methods

Field Methods

Fieldwork was conducted near Punta Espinosa, Fernandina Island, Galapagos Archipelago, during June and December 1990. Flightless cormorants were captured

by hand, measured, weighed, and marked with a permanent aluminum leg band and three temporary colored plastic bands. Tarsus length was measured by firmly pressing the jaws of the caliper between the hypotarsus and the base of a flattened foot. Bill length (the exposed culmen) was measured by spreading a compass between the tip and base of the upper bill and then measuring the compass span with a ruler. Bill depth and bill width at the level of the nostrils were measured with a caliper. Wing length, the distance from the curvature of the folded wing to the tip of the second phalanx (i.e., feathers excluded), was measured with a flexible ruler. Exclusion of flying feathers from wing length measurement was because of the common variation among individuals in feather length caused by feather wear, absence, or regrowth. All measurements were read to the nearest 0.1 mm except for wing length, which was measured to the nearest 1.0 mm.

Data for other species of cormorants, used for the comparative analyses of sexual dimorphism in body mass and other morphometric traits, comes from Livezey (1992) and Farbairn and Shine (1993), who compiled information from several other sources. I restricted analyses to only 17 species (for body mass) and 10 species (for other traits) of cormorants whose information looked consistent. Sexual dimorphism for every trait was defined as the ratio of male to female. Statistical analyses included model II regression analyses as measurements of body mass and all morphometric traits are expected to involve measurement error (e.g., see Sokal and Rohlf 1995). To test for Rensch's rule, I regressed the logarithms ($\log_{10}$) of male traits to that of female traits and tested the null hypothesis of geometric isometry (slope coefficient $\beta = 1.0$) using model II reduced major axis (RMA) regression (Sokal and Rohlf 1995). Slopes, standard error, and 95 % confidence intervals (not reported) of major axis regressions were estimated for the linear model and jackknifing the original data 1,000 times using RMA software for RMA regression (Bohonak 2004; http://www.bio.sdsu.edu/pub/andy/rma.html). Departures from geometric isometry were inferred from the lack of overlap of the 95 % confidence interval of the estimated slope with unity. These analyses were not corrected for phylogenetic effects because all species were phalacrocoracids. The notion that phylogenetic differences within the family are small is supported by the conservative suggestion by Kennedy et al. (2000) that this group be considered a single genus, *Phalacrocorax*, until current disagreement in their taxonomy is resolved.

Results

Size and Morphometrics

Comparison of body masses (BM) of male (M) and female (F) cormorants and five morphometric traits shows that the Flightless cormorant lies well above the range for all other species, thus demonstrating that it is by far the largest member of the family (Table 12.1). Also, both sexes and particularly male Flightless cormorants are remarkable larger than either one of their two sister species (Kennedy et al. 2009),

Table 12.1 Body mass and morphometrics of cormorants (Phalacrocoracidae) compared with those of the Galapagos Flightless cormorant (GFC). GFC sample sizes were 54 individuals of each sex. Values for other cormorants were pooled from 16 other species for body mass and nine other species for bill length, bill depth, and tarsus length. Data taken from this study (GFC) and from Livezey (1992) and Fairbairn and Shine (1993) for other cormorant species. *Phalacrocorax chelconotos* from Fairbairn and Shine (1993) was excluded from the analyses because of the large discrepancy of measurements reported between the two populations. Standard errors (S) and range (R) are also shown

Trait		Males		Females	
		Cormorants	GFC	Cormorants	GFC
Body mass (g)	Mean	2,037	3,762	1,745	2,754
	S	672		525	
	R	816–3,320		729–2,700	
Bill length (mm)	Mean	61.6	80.9	57.6	72.6
	S	14.5		12.0	
	R	32.1–79.3		32.3–71.4	
Bill depth (mm)	Mean	12.4	23.3	11.0	19.2
	S	3.0		2.3	
	R	9.0–18.2		8.6–15.1	
Tarsus length (mm)	Mean	60.1	82.2	56.3	75.4
	S	10.3		8.1	
	R	39.6–74.0		39.0–64.9	

of which the Neotropic cormorant (*P. brasilianus*) is probably its closest relative (body mass: $M_{BM} = 131\,\%$ cf. $F_{BM} = 82\,\%$, subindices follow those in Table 12.1) and the Double-crested cormorant (*P. Auritus*; $M_{BM} = 69\,\%$ cf. $F_{BM} = 45\,\%$; bill length: $M_{BL} = 37\,\%$ cf. $F_{BL} = 36\,\%$; bill depth: $M_{BD} = 73\,\%$ cf. $F_{BD} = 57\,\%$; tarsus length: $M_{TL} = 31\,\%$ cf. $F_{TL} = 18\,\%$).

There was low variation among individuals within both sexes in all measured traits, with male body mass, the most variable trait, showing a coefficient of variation (CV) of 8 % (Table 12.2). However, the sexes did not differ substantially in their variance for any measured morphometric trait. Correlation between morphometric traits and between morphometric traits and body mass differed both among traits and between the sexes (Table 12.3). Among traits, bill length correlated significantly with all other morphometric traits in females but with only one trait in males (see inset, Table 12.3). The frequency of trait correlation also differed between sexes with females showing the highest frequency (Table 12.3): of the 15 trait pairs for the six measured morphometrics there was correlation between 11 in females but only five in males. Overall, strength of correlation between morphometric traits varied from low in males ($r = 0.10$, not significant (NS) between wing length and bill width) to moderate in females ($r = 0.51$, $p < 0.001$ between wing length and body mass) (Table 12.3). Thus, wing length was the trait showing both the lowest and the highest correlation with another morphometric trait, although with a different trait and on different sexes. The strength of correlation between morphometric traits differed between the sexes, and in general, traits correlated more strongly and more frequently in females than in males. For instance, although bill length did not correlate with body

Table 12.2 Morphometrics, sexual dimorphism indices (SDI = male/female), and coefficients of variation (CV) in Galápagos Flightless cormorants. Data taken from 54 males and 54 females measured between June and December1990. Body mass of males and females in fasting condition was 3,451 g (SD = 153; $N = 10$) and 2,545 g (SD = 71 SD; $N = 11$), respectively. All statistical tests were two-tailed

Trait		Males	Females	SDI
Body mass (g)	Mean	*3,762*	*2,754*	*1.37*
	SD	300	174	
	Range	3,225–4,710	2,497–3,495	
	CV	8.0	6.3	
Bill width (mm)	Mean	*16.1*	*13.2*	*1.22*
	SD	0.7	0.6	
	Range	14.5–17.6	11.8–14.5	
	CV	4.3	4.5	
Bill depth (mm)	Mean	*23.3*	*19.2*	*1.21*
	SD	0.9	0.7	
	Range	21.6–25.3	17.3–20.5	
	CV	3.8	3.6	
Bill length (mm)	Mean	*80.9*	*72.6*	*1.11*
	SD	2.3	2.4	
	Range	75.5–86.5	63.0–78.6	
	CV	2.8	3.3	
Tarsus length (mm)	Mean	*82.2*	*75.2*	*1.09*
	SD	2.4	2.3	
	Range	76.0–87.9	70.5–81.2	
	CV	2.9	3.1	
Wing length (mm)	Mean	*83.1*	*77.0*	*1.08*
	SD	2.4	2.1	
	Range	77.0–89.0	72.0–84.0	
	CV	2.9	2.7	

mass in males ($r = 0.17$, NS), it had a moderate but highly significant correlation in females ($r = 0.43$, $p < 0.001$). A similar pattern occurred between wing length and body mass (males $r = 0.20$, NS; females: $r = 0.51$, $p < 0.001$).

Size-related morphometric traits (mass, tarsus length, and wing length) correlated moderately and at a highly significant level in both sexes except for wing length with body mass in males (Table 12.3). Among feeding structures (bill size and shape-related traits), bill depth and bill width showed the highest correlation in both males and females ($r = 0.39$, $p < 0.01$), whereas there was correlation between bill depth with bill length ($r = 0.38$, $p < 0.01$) and between bill width with bill length (0.27, $p < 0.05$) in females but not males. However, there was no strong correlation between feeding structures and size-related traits; among females, both bill depth ($r = 0.44$, $p < 0.01$) and bill length ($r = 0.43$, $p < 0.01$) correlated with body mass, whereas among males, only bill depth correlated with body mass ($r = 0.29$, $p < 0.05$). Bill length correlated only modestly with tarsus length in females ($r = 0.36$, $p < 0.01$) and did not correlate in males.

Table 12.3 Coefficients of correlation between morphometric traits in male (M) and female (F) Galapagos Flightless cormorants. Traits are body mass (BM), bill depth (BD), bill width (BW), bill length (BL), tarsus length (TL), and wing length (WL)

TRAIT	SEX	BM	BD	BW	BL	TL	WL
BM	M	–					
	F						
BD	M	0.29*	–				
	F	0.44**					
BW	M	0.17	0.39**	–			
	F	0.24	0.39**				
BL	M	0.17	0.17	0.19	–		
	F	0.43**	0.38**	0.27*			
TL	M	0.44***	0.26	– 0.03	0.26	–	
	F	0.42**	0.2	0.07	0.36**		
WL	M	0.20	0.23	0.10	0.35*	0.46***	–
	F	*0.51****	0.42***	0.14	0.40**	0.49***	

	Males	Females
BM	2	4
BD	2	4
BW	1	2
BL	1	5
TL	2	3
WL	2	4

M males, *F* females
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$

Table 12.4 Sexual dimorphism measures (male/female) in Galapagos Flightless cormorants (GFC), Double-crested cormorant (*P. auritus*; DCC), and other phalacrocoracids (cormorant). Means, standard deviations (S), and ranges (R) are shown

		Sexual dimorphism		
Trait		Cormorants	DCC	GFC
Body Mass (g)	Mean	1.16	1.17*	1.37
	S	0.07		
	R	1.03–1.26		
Bill width (mm)		?		1.22
Bill depth (mm)	Mean	1.12	1.11	1.21
	S	0.1		
	R	1.05–1.21		
Bill length (mm)	Mean	1.06	1.10	1.11
	S	0.07		
	R	0.99–1.18		
Tarsus length (mm)	Mean	1.07	1.00	1.09
	S	0.10		
	R	0.99–1.32		
Wing length (mm)				1.08

*$p < 0.05$

Sexual Size Dimorphism

Males and females differed significantly in body mass and all five measured external morphometric traits, with body mass being the most dimorphic trait (males were 37 % larger than females) followed by bill width and bill depth (Table 12.4). Relative to other members of the family, the Flightless cormorant was significantly the most sexually dimorphic in both body mass and bill depth, but their dimorphism in bill length was similar to that of other species (Table 12.4). An across-species comparison shows that sexual dimorphism in both body mass and bill depth were positively correlated to the mean body mass of each sex, respectively (males SDImass: $r = 0.80$, $p = 0.002$, $n = 12$; males SDI_{BD}: $r = 0.86$, $P = 0.006$, $N = 8$; females SDImass: $r = 0.68$, $P = 0.016$, $N = 12$; females SDI_{BD}; $r = 0.82$, $P = 0.01$, $N = 8$). As expected, sexual dimorphism in body mass and bill depth also correlated with the mean body size of both sexes (Fig. 12.1). Thus, variation in sexual dimorphism in body mass and bill depth and other morphometric traits among species within Phalacrocoracidae is fully explained by variation in body size among species. Support for interpretation that the positive relationship between sexual dimorphism and the mean body mass of both sexes is not an artifact resulting from the inclusion of female mass in both variables (e.g., see Clutton-Brock et al. 1977) is demonstrated by the significant regression of male body mass as a function of female mass (Table 12.5). Furthermore, the coefficient of allometry (slope β) of $\log_{10}$ male body mass as a function of $\log_{10}$ female body mass, except for bill depth, was not significantly greater than 1.0 (Table 12.5). Thus, among phalacrocoracids, SSD in most traits has an isometric increase with body size using logarithmic scaling and the Phalacrocoracids do not follow Rensch's rule.

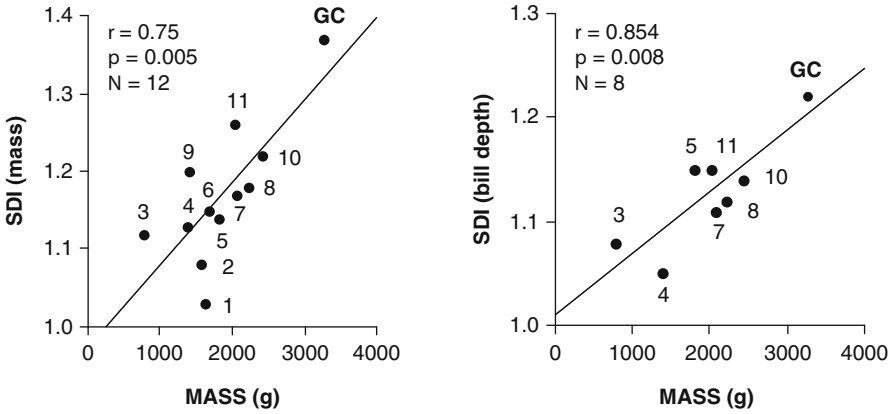


Fig. 12.1 Sexual dimorphism in **a** body mass and **b** bill depth in relation to the average body mass of both sexes in some Phalacrocoracidae. The sexual dimorphism index (SDI) is the mean of males over mean of females for the trait in question. The species included are the Galápagos cormorant *Phalacrocorax harrisi* (*GC*), *P. nigrogularis* (*1*), *P. brasilianus* (*2*), *P. melanoleucus* (*3*), *P. magellanicus* (*4*), *P. penicillatus* (*5*), *P. aristotelis* (*6*), *P. auritus* (*7*), *P. atriceps* (*8*), *P. gaimardi* (*9*), *P. carbo* (*10*), *P. urile* (*11*). Data taken from Livezey (1992)

Table 12.5 Results of model II reduced major axis (RMA) regression while testing for Rensch's rule among species in the family Phalacrocoracidae performed on $\log_{10}$ transformed data testing the null hypothesis of slope (β) = 1.0. Slopes (β), standard error (SE), 95 % confidence intervals (not reported), and coefficients of determination (r^2) of reduced major axis (RMA) regressions are reported for both the linear model and jackknifing the original data 1,000 times using RMA software for RMA regression (see Methods for details)

	RMA slope coefficient (β)		
Morphometric trait		Linear model	Jackknife
Body mass	β estimate	1.117	1.104*
	SE	0.067	0.065
	r^2	0.940	0.950
Tarsus length	β estimate	1.163	1.132*
	SE	0.160	0.085
	r^2	0.848	
Bill depth	β estimate	*1.144**	1.145*
	SE	0.039	0.0382
	r^2	0.991	0.993
Bill length	β estimate	1.146	1.108
	SE	0.072	0.075
	r^2	0.968	1.000

*$p < 0.05$

Discussion

Size and Morphometrics

The ancestor of the Flightless cormorant colonized the Galapagos Islands about 2 million years ago (Kennedy et al. 2009). Consequently, this species is perhaps

one of the most recently diverged species in its family. Assuming a relatively recent divergence, in comparison with the rest of the cormorants and particularly relative to either one of its two sister species, the larger size (gigantism) and remarkable SSD of the Flightless cormorant probably are both novel traits that have evolved in situ in the Galapagos following colonization.

This study, as well as Livezey's (1992), confirmed that the Flightless cormorant is the largest member of its family. The use of body mass as a predictor for structural size is widespread (e.g., see Rising and Somers 1989), but body mass has several confounding effects (Piersma and Davidson 1991). Among cormorants, which usually eat relatively large meals (see Valle 1994; C. Valle, unpublished data), field measurements of body mass may be substantially variable depending on when the individual last fed. Levels of correlation between external morphometric traits indicated that in the Flightless cormorant, the best predictor of body mass for males was tarsus length, but for females it was wing length. However, as the correlations between body mass and tarsus length were also moderate and similar between sexes, tarsus length is perhaps the best predictor of structural size for the species and a useful index for interspecific comparisons. Low individual variation in body size and morphometric traits in both sexes is suggestive of strong directional selection (e.g., Price et al. 1984; Gibbs and Grant 1987; Grant and Grant 2008) because all the traits have clear functionality, except perhaps wing length. Low variation in wing length in a flightless bird is surprising, as wings have no apparent function for swimming, prey capture, or courtship. A feasible explanation could be genetic correlation of wing length with other traits constrained by selection. In the Flightless cormorant, this explanation is supported by a relatively strong correlation of wing length with both body mass and tarsus length.

Several factors may have facilitated and lead to an increase in size in the Flightless cormorant (e.g., see Snow 1966), a species in male-biased SSD. It has been hypothesized that in species in which the male is larger than the female, larger size and sexual dimorphism in body size can be driven by intrasexual selection due to male–male competition for access to mates (Darwin 1871; Clutton-Brock et al. 1977; Alexander et al. 1979). However, the Flightless cormorant is sequentially polyandrous and role-reversed in the sense that females rather than males aggressively compete for access to mates (Valle 1994; Valle 2009). Furthermore, there is no strong evidence that in this species, intersexual selection (female mate choice) leads to a pattern of disassortative mating (i.e., smaller females preferentially choosing larger males as mates; Valle 1994; but see Tindle 1984). With a current lack of evidence for either intrasexual or intersexual selection as a mechanism favoring an increase in size for this species, it appears more likely as an adaptation resulting from ecological selection.

Here, I hypothesize that ecological selection for larger body size in the Flightless cormorant has resulted from the advantage gained by largest males being able to take larger prey. Largest males are predicted to capture larger prey, thus enhancing survival because they can more easily satisfy daily food demands and reduce the competition with females. On the other hand, given that enough small prey is available, it is predicted that larger size will not be selected for in females because of a stronger competition with males. This scenario of direct selection for larger body size only in

males is consistent with an ancestral male-biased condition judging from the pattern of male-biased SSD found among phalacrocoracids (Livezey 1992; Johnsgard 1993). Flightless cormorants' prey vary greatly in size (Valle 1994, C. Valle, unpublished data), ranging from small blenny fish (*Dialommus fuscus*; < 50 mm in length and < 50 g body mass) captured mainly by females to very large prey (e.g., moray eels, *Muraena lentiginosa*, Galapagos reef octopuses, *Octopus oculifer*, and flag cabrilla, *Epinephelus labriformis*), whose capture and consumption—almost exclusively by male cormorants— requires great effort and strength. I suggest that a great increase in body size allowed males to feed upon this unusually large type of prey found in their new habitat upon colonization, thus broadening the species feeding niche. Selection pressures for larger body size and feeding structures, especially beak depth, which correlate with jaw strength in birds (Grant 1968; Grant 1999; Grant and Grant 2008; Wilson 1972; Lederer 1975) and may affect prey size in Flightless cormorants, are expected to remain strong preferentially on males that feed disproportionally on very large prey (Valle 1994). The remarkable increase in body size and feeding structures of the Flightless cormorant relative to their two sister species attest for this possibility. An opportunity for character release may also have facilitated the great increase in size in the Galapagos cormorant, as there are no other large piscivores in the area with a similar diet and foraging niche. Competitive exclusion with the smaller sympatric Galapagos penguin (*Spheniscus mendiculus*) could have been an important factor for body size increase (e.g., Snow 1966; Thorton 1971; Snow and Nelson 1984). I question that, however, because Flightless cormorants feed upon demersal, slow-moving sedentary prey by searching among rocks and crevices (Snow 1966; personal observations by C. Valle). Penguins, on the other hand, normally feed on very active and pelagic prey.

Sexual Size Dimorphism

Here, I propose that the greater SSD in Flightless cormorants relative to other phalacrocoracids is simply explained by the species' larger size. That larger body size and greater SSD have in turn evolved through intersexual disruptive selection (Selander 1966, 1972; Shine 1989; Herrel et al. 1999; Butler et al. 2000), with overall directional selection for larger body size but with males increasing at a greater rate than females. Such a pattern of disruptive selection leading to a general selection for larger body size is expected to result if disruptive selection involves stabilizing selection for female smaller size or if selection for larger size in males is stronger than selection for smaller size in females (see Slatkin 1984), coupled with an imperfect genetic correlation for body size between the sexes (e.g., see Maynard Smith 1977; Lande 1980; Lande and Arnold 1985; Fairbairn and Preziosi 1994).

Current empirical evidence and theoretical research support a number of predictions derived from these hypotheses of the evolution of body size and SSD in the Flightless cormorant and perhaps among other phalacrocoracids, a taxon with male-biased, sexually dimorphic but essentially socially monogamous species

(Johnsgard 1993). First, the hypothesis that the greater SSD dimorphism in the Flightless cormorant relative to other phalacrocoracids is simple explained by their largest size is fully supported by the comparative approach. In the comparison of phalacrocoracids, the correlations of SSD and other morphometric traits showed that body size differences fully account for SSD differences between species and that cormorants (Phalacrocoracidae) do not conform to Rensch's rule.

Second, the main prediction from intersexual disruptive selection, using diet as a proxy for intersexual competition and a divergent foraging niche, is that the sexes differ substantially in prey size or type. The evidence for a sexually dimorphic foraging niche at least at first glance seems compelling for this species. The sexes differ dramatically in the size of prey they captured (Valle 1994).

Third, strong intraspecific competition for food is expected in this population because flightlessness and foraging behavior severely restrict their foraging range (Valle 1994; Valle 1995; Wilson et al. 2008); as predators relaying almost exclusively on benthic prey, their foraging grounds are restricted to the shallower and severely limited littorine areas along the coasts of the two of the Galapagos Islands. Also, as both sexes forage in the same areas and no aggression or direct food competition has been observed while foraging, competition within each sex and between the sexes is expected to be indirect or scramble competition.

Fourth, an increase in size in females relative to size of the ancestral species is predicted despite intersexual disruptive selection favoring female's small size, because female size is expected to increase as a correlated response to selection for larger size in males due to genetic correlation (Lande and Arnold 1985; Fairbairn 1997); thus, indirect selection for larger size in females due to genetic correlation (e.g., see Fairbairn 1997) is expected to counterbalance stabilizing selection for female size or selection for small size in females. Indeed, data showed that female Flightless cormorants are substantially larger than either of their two sister species, which are assumed to more closely represent the ancestral character state for body size in the Flightless cormorant. Another complementary prediction that could be derived from the correlated response selection hypothesis is that both sexes should have evolved a change in size in the same direction but at different rates. This prediction is fulfilled in the Flightless cormorant as attested by the large difference in the proportion of size increase achieved by male and female Flightless cormorants when compared with their two sister species.

The lack of allometry found in the comparative analyses of SSD found among phalacrocoracids also agrees with the theoretical prediction of the correlated response selection hypothesis, although allometry for SSD should occur during early stages of evolutionary divergence (Lande 1980; Zeng 1988; Emerson 1994; Fairbairn 1997). The sex differences in the strength of selection for larger size may not be enough to promote an allometric relationship in a log–log scale, thus accounting for the lack of the pattern known as Rensch's rule among the Phalacrocoracidae. This finding is also in agreement with a recent study showing that SSD dimorphism in seabirds does not conform to Rensch's rule, a result attributable to the socially monogamous mating system of most seabirds (Serrano-Meneses and Szekely 2006). However, among phalacrocoracids, because they are essentially monogamous, there is no confounding

of body size and mating system (polygyny), which may drive the increased male-biased dimorphism in larger, more polygynous species. (e.g., Clutton-Brock et al. 1977) and explain the recurrence of Rensch's rule among a large number of taxa (Fairbairn 1997; Fairbairn et al. 2007).

Comparative studies of the relationship of SSD and body size have diminished the role of ecological mechanisms and thus have overemphasized the role of sexual selection on males as the main driver of interspecific differences in SSD and the evolution of allometry for SSD (Jehl and Murray 1986; Fairbairn 1997; Abouheif and Fairbairn 1997; Fairbairn et al. 2007; but see Shine 1989). Results of this study suggest that disruptive ecological selection could be the driving force causing the evolution of both body size changes and SSD within a species and across species within a clade. However, the setting for the evolution of SSD in the Flightless cormorant seems to be even more complex than that for other species as a result of the evolution of sequential polyandry (Valle 1994, 2009). In the Flightless cormorant, males are in short supply because they provide longer parental care. Therefore, although the adult sex ratio is near unity, females rather than males aggressively compete with each other for access to mates and presumably females may currently be undergoing high levels of intrasexual selection targeting overall body size, thus further complicating our understanding of the evolution and maintenance of SSD.

Acknowledgments I gratefully acknowledge financial support from Princeton University (Department of Ecology and Evolutionary Biology) and National Geographic that allowed me to conduct field research. I thank Universidad San Francisco de Quito and the Galapagos Academic Institute for the Arts and Sciences (GAIAS) through Santiago Gangotena (President), Carlos Montufar (Vice President), and Diego Quiroga (Dean for Scientific Research) for their encouragement and for allowing the necessary time for data analyses and preparing the manuscript for publication. I also thank Carlos Mena and Steve Walsh for their encouragement. I would like to thank John Krenz for his valuable comments and suggestions.

References

Abouheif E, Fairbairn DJ (1997) A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. Am Nat 149:340–362

Alexander RD, Hoogland JL, Howard RD, Noonan KM, Sherman PW (1979) Sexual dimorphism and breeding systems in pinnipeds, ungulates, primates and humans. In: Chagnon NA, Irons W, Scituate MA (eds) Evolutionary biology and human social behavior: an anthropological perspective. Duxbury, pp 402–435

Anderson M (1994) Sexual selection. Princeton University Press, Princeton

Berry JF, Shine R (1980) Sexual dimorphism and sexual selection in turtles (order Testudines). Oecologia 44:185–191

Blanckenhorn WU (2000) The evolution of body size: what keeps organisms small? Q Rev Biol 75:385–407

Bohonak AJ (2004) Software for reduced major axis regression, V.1.17. San Diego State University

Butler MA, Schoener TW, Losos JB (2000) The relationship between sexual size dimorphism and habitat use in Greater Antillean *Anolis* lizards. Evolution 54:259–272

Calder WA (1984) Size, function, and life history. Cambridge, Harvard University Press

Cheverud JM, Dow MM, Leutenegger W (1985) The quantitative assessment of Phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. Evolution 39:1335–1351

Clutton-Brock TH, Harvey PH, Rudder B (1977) Sexual dimorphism, socionomic sex ratio and body weight in primates. Nature 269:797–800

Cox RM, Skelly SL, John-Alder HB (2003) A comparative test of adaptive hypotheses for sexual size dimorphism in lizards. Evolution 57:1653–1669

Darwin C (1871) The descent of man and selection in relation to sex, 1st edn. John Murray, London

Downhower JF (1976) Darwin's finches and the evolution of sexual dimorphism in body size. Nature 263:558–563

Emerson SB (1994) Testing pattern predictions of sexual selection. Am Nat 143:848–869

Fairbairn DJ (1997) Allometry for sexual size dimorphism: pattern and process in the coevolution of body size in males and females. Ann Rev Ecol Syst 28:659–687

Fairbairn DJ (2005) Allometry for sexual size dimorphism: testing two hypotheses for Rensch's rule in the water strider *Aquarius remigis*. Am Nat 166:69–84

Fairbairn DJ, Preziosi RF (1994) Sexual selection and the evolution of allometry for sexual size dimorphism in the water strider, *Aquariusremigis*. Am Nat 144:101–108

Fairbairn DJ, Shine R (1993) Patterns of sexual size dimorphism in seabirds of the southern hemisphere. Oikos 68:139–145

Fairbairn DJ, Blanckenhorn WU, Szekely T (2007) Sex, size, and gender roles: evolutionary studies of sexual size dimorphism. Oxford University Press, Oxford

Gaulin SJC, Sailer LD (1984) Sexual dimorphism in weight among the Primates: the relative impact of allometry and sexual selection. Int J Primatol 5:515–535

Gibbs HL, Grant PR (1987) Oscillating selection on Darwin's finches. Nature 327:511–513

Grant PR (1968) Bill size, body size, and the ecological adaptations of bird species to competitive situations on islands. Syst Zool 17:319–333

Grant PR (1999) Ecology and evolution of Dartwin's finches, 2nd edn. Princeton University Press, Princeton

Grant PR, Grant BR (2008) How and why species multiply: the radiation of Darwin's finches. Princeton University Press, Princeton

Harris MP (1979) Population dynamics of the Flightless cormorant Nannopterum harrisi. Ibis 121:135–146

Hedrick AV, Temeles EJ (1989) The evolution of sexual dimorphism in animals: hypotheses and tests. Trends Ecol Evol 4:136–138

Herrel A, Spithoven L, Van Damme R, De Vree F (1999) Sexual dimorphism of head size in *Gallotia galloti*: testing the niche divergence hypothesis by functional analyses. Funct Ecol 13:289–297

Jehl JR Jr, Murray BG Jr (1986) The evolution of normal and reverse sexual size dimorphism in shorebirds and other birds. Curr Ornithol 3:1–86

Johnsgard PA (1993) Cormorants, darters, and pelicans of the world. Smithsonian Institution Press, Washington

Karubian J, Swaddle JP (2001) Selection on females can create 'larger males'. Proc R Soc Lond B Biol Sci 268:725–728

Kennedy M, Gray RD, Spence HG (2000) The phylogenetic relationships of the shags and cormorants: can sequence data resolve a disagreement between behavior and morphology? Mol Phyl Evol 17:345–359

Kennedy M, Valle CA, Spencer HG (2009) The phylogenetic position of the Galápagos Cormorant. Mole Phylogenet Evol 53:94–98

LaBarbera M (1989) Analyzing body size as a factor in ecology and evolution. Annu Rev Ecol Syst 20:97–117

Lande R (1980) Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution 34:292–305

Lande R, Arnold SJ (1985) Evolution of mating preference and sexual dimorphism. J Theor Biol 117:651–664

Lederer RJ (1975) Bill size, food size, and jaw forces of insectivorous birds. Auk 92:385–387
Leutenegger W, Cheverud J (1982) Correlates of sexual dimorphism in primates: ecological and size variables. Int J Primatol 3:387–402
Leutenegger W (1978) Scaling of sexual dimorphism in body size and breeding system in primates. Nature 272:610–611
Livezey BC (1992) Flightlessness in the Galapagos Cormorant (*Compsohalieus* [*Nannopterum*] *harrisi*): heterochrony, giantisms and specialization. Zool J Linn Soc 105:155–224
Maynard SJ (1977) Parental investment: a prospective analysis. Anim Behav 25:1–9
Miles DB, Ricklefs RE (1984) The correlation between ecology and morphology in deciduous forest passerine birds. Ecology 65(5):1629–1640
Parker GA (1992) The evolution of sexual size dimorphism in fish. J Fish Biol 41(Suppl B):1–20
Peters RH (1983) The ecological implications of body size. Cambridge University Press, New York
Peters RH (1986) The ecological implications of body size. University Press, Cambridge (Cambridge Studies in Ecology, May 22, 1986)
Piersma T, Davidson KC (1991) Confusions of mass and size. Auk 108:441–444
Price TD, Grant PR, Gibbs HL, Boag PT (1984) Recurrent patterns of natural selection in a population of Darwin's finches. Nature 309:787–789
Price TD (1984) The evolution of sexual size dimorphism in Darwin's finches. Am Nat 123:500–518
Ralls K (1976) Mammals in which females are larger than males. Q Rev Biol 51:245–276
Reeve JP, Fairbairn DF (1996) Sexual size dimorphism as a correlated response to selection on body size: an empirical test of the quantitative genetic model. Evolution 50:1927–1938
Reiss MJ (1986) Sexual dimorphism in body size: are larger species more dimorphic? J Theor Biol 121:163–172
Rensch B (1960) Evolution above the species level. Columbia University Press, New York
Ricklefs RE, Cox GW (1977) Morphological similarity and ecological overlap among passerine birds on St. Kitts, British West Indies. Oikos 28:60–66
Ricklefs RE, Travis J (1980) A morphological approach to the study of avian community organization. Auk 97:321–338
Rising JD, Somers KM (1989) The measurement of overall body size in birds. Auk 106:666–674
Schmidt-Nielsen K (1984) Scaling. Why is animal size so important? Cambridge University Press, New York
Selander RK (1966) Sexual dimorphism and differential niche utilization in birds. Condor 68:113–151
Selander RK (1972) Sexual selection and dimorphism in birds. In: Campbell B (ed) Sexual selection and the descent of man1871–1971. Aldine Publ. Co., Chicago
Serrano-Meneses MA, Szekely T (2006) Sexual size dimorphism in seabirds: sexual selection, fecundity selection and differential niche-utilisation. Oikos 113:385–394
Shine R (1979) Sexual selection and sexual dimorphism in the Amphibia. Copeia 2:297–306
Shine R (1988) The evolution of large body size in females: a critique of Darwin's "fecundity advantage" model. Am Nat 131:124–131
Shine R (1989) Ecological causes for the evolution of sexual dimorphism: a review of the evidence. Q Rev Biol 64:419–461
Slatkin M (1984) Ecological causes of sexual dimorphism. Evolution 38:622–630
Snow BK (1966) Observations on the behaviour and ecology of the Flightless Cormorant *Nannopterum harrisi*. Ibis 108:265–280
Snow DW, Nelson JB (1984) Evolution and adaptations of Galapagos sea-birds. Biol J Linn Soc 21:137–155
Sokal RR, Rohlf FJ (1995) Biometry, 3rd edn. W. H. Freeman and Company, New York
Stephens PR, Wiens JJ (2008) Testing for evolutionary tradeoffs in a phylogenetic context: ecological diversification and evolution of locomotor performance in emydid turtles. J Evol Biol 21:77–87
Stephens PR, Wiens JJ (2009) Evolution of sexual size dimorphisms in emydid turtles: Ecological dimorphism, Rensch's rule, and sympatric divergence. Evolution 63: 910–925

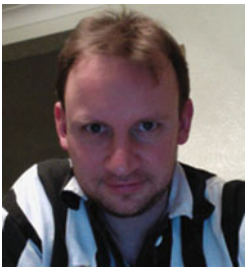
Szekely T, Reynolds JD, Foguerola J (2000) Sexual size dimorphism in shorebirds, gulls, and alcids: the influence of natural selection. Evolution 54:1404–1413
Szekely T, Freckleton RP, Reynolds JD, Southwood R (2004) Sexual selection explains Rensch's rule of size dimorphism in shorebirds. Proc Natl Acad Sci U S A 101:12224–12227
Thorton I (1971) Darwin's Islands: a natural history of the Galapagos. The Natural History Press, New York
Tindle R (1984) The evolution of breeding strategies in the Flightless cormorant (Nannopterum harrisi) of the Galapagos. Biol J Linn Soc 21:157–164
Valle CA (1994) The ecology and evolution of sequential polyandry in Galapagos cormorants, Ph.D. Dissertation. Princeton University, Princeton
Valle CA (1995) Effective population size and demography of the rare flightless Galapagos cormorant. Ecol Appl 5:601–617
Valle CA (2009) The Flightless cormorant: the evolution of female rule In: de Roy T (ed) Galapagos: preserving Darwin's legacy. David Bateman Ltd., New York
Webster MS (1992) Sexual dimorphism, mating system, and body size in New World blackbirds (Icterinae). Evolution 46:1621–1641
White EP, Morgan SK, Kerkhoff AJ, Enquist BJ (2007) Relationships between body size and abundance in ecology. Trends Ecol Evol 22:323–330
Wiens JA (1982) On size ratios and sequences in ecological communities: are there no rules? Ann Zool Fennici 19:297–308
Wikelski M (2005) Evolution of body size in Galapagos marine iguanas. Proc R Soc B 272:1985–1993
Wikelski M, Trillmich F (1997) Body size and sexual size dimorphism in marine iguanas fluctuate as a result of opposing natural and sexual selection: an island comparison. Evolution 51:922–936
Wiklund C, Karlsson B (1988) Sexual size dimorphism in relation to fecundity in some Swedish Satyrid butterflies. Am Nat 131:132–138
Williams GC (1966) Adaptation and natural selection. Princeton University Press, Princeton
Wilson MF (1972) Seed size preference in finches. Wilson Bull 84:449–455
Wilson DS (1975) The adequacy of body size and a niche difference. Am Nat 109:769–784
Wilson RP, Vargas FH, Steinfurth A, Riordan P, Ropert-Coudert Y, MacDonald DW (2008) What grounds some birds for life? Movement and diving in the sexually dimorphic Galapagos cormorant. Ecol Monographs 78:633–652
Zeng Z-B (1988) Long-term correlated response, interpopulation covariation, and interspecific allometry. Evolution 42:363–374

About the Editors and Authors

Gabriel A. Trueba Gabriel is Professor of Microbiology, Bacterial Genetics at Universidad San Francisco de Quito (USFQ), Ecuador. He is also the head of the Microbiology Institute at USFQ. He is Member of Editorial Board of The Open Tropical Medicine Journal and Scientific Board of revista Hechos Microbiológicos. He has received the Bruce Hoeneisen research award from USFQ. He holds a PhD from Iowa State University, USA.

Carlos Montufar Carlos is Professor of Physics and Introductory Courses in sciences at Universidad San Francisco de Quito (USFQ). Carlos is Vice Chancellor, Academic Dean and Co-Founder and Vice President of Corporación de Promoción Universitaria, Executive Director, Co-Founder of Colegio Menor San Francisco de Quito. Carlos holds a PhD from University of Notre Dame, USA.

Sandra L. Baldauf Professor at Department of Systematic Biology, Evolutionary Biology Centre, Uppsala University, Norbyvagen 18D, SE-752 34 Uppsala, Sweden and Professor and Chair, Program in Systematic Biology, Uppsala University. Sandra is also Member of IEGS Departmental Board and Editorial Board, Quarterly Review of Biology and Associate Editor of *Protist*. She is also Advisor to Chinese Academy of Sciences. She holds a PhD from University of Michigan, USA.



Martin Carr Martin is a Research Fellow of the Biotechnology and Biological Sciences Research Council at University of York, Department of Biology. Martin holds a PhD from Institute of Genetics, University of Nottingham, Germany.

Michael F. Dolan Michael is Lecturer in Biology at University of Massachusetts and Co-Director of the NASA Planetary Biology Internship Program at the Marine Biological Laboratory. He is also a Research Associate in the Division of Invertebrate Zoology at the American Museum of Natural History in New York. He holds a PhD from University of Massachusetts, USA.

Edward Feil Professor of Microbial Evolution at the Department of Biology and Biochemistry, University of Bath. Edward is Senior Research Fellow at the Department of Biology and Biochemistry, University of Bath. He is also Welcome Trust Prize Fellow of University of Sussex. Edward holds a PhD from University of Sussex, UK.

Peter R. Grant Peter is Emeritus Professor of Zoology at the Department of Ecology and Evolutionary Biology, Princeton University. He has received numerous awards such as Honorary PhD degrees from Uppsala University and University of Zurich, and DSc degrees from Universidad San Francisco, Quito and McGill University, the Alexander von Humboldt Foundation Senior Scientist Research Prize, the E. O. Wilson Prize, the American Society of Naturalists, the Darwin-Wallace Medal of the Linnean Society, the Kyoto Prize in Evolutionary Biology, etc. He holds a PhD from University of British Columbia, Canada.

Rosemary Grant Rosemary is Emeritus Professor of Zoology at the Department of Ecology and Evolutionary Biology, Princeton University. She has received numerous awards such as Honorary PhD degrees from Uppsala University and University of Zurich, and DSc degrees, Universidad San Francisco, Quito and McGill University, the Darwin Medal, the Royal Society of London, the Kyoto Prize in Evolutionary Biology, the Darwin-Wallace Medal of the Linnean Society, etc. She holds a PhD from Uppsala University, Sweden.

Ricardo Guerrero Ricardo is Professor of Microbiology at the University of Barcelona, Barcelona, Spain, and Adjunct Professor of Microbiology, University of Massachusetts-Amherst, Amherst, USA. He is also director of the Ecogenètica I Diversitat Microbianesat University of Barcelona, Spain, and Editor-in-Chief of International Microbiology. He holds a Doctorate in Sciences from University of Barcelona, Spain.

Amparo Latorre Amparo is Professor of Genetics, Evolution of Genomic Complexity and Comparative Genomics at University of Valencia, Spain, and Director of the Cavanilles Institute for Biodiversity and Evolutionary Biology (ICBiBE). She holds a PhD from University of Valencia, Spain.

Antonio Lazcano Professor, Origins of Life, School of Sciences, Universidad Nacional Autónoma de México (UNAM), Mexico, and Distinguished Professor of the University of Henan, China. Antonio is also President of the International Society for the Study of the Origins of Life (ISSOL), Member of Editorial Board of The Journal of Molecular Evolution, Origins of Life, among other scientific journals. He has also received several awards like The Golden Medal for Biological Research, Puebla, Mexico, and Doctor Honoris Causa from the University of Milan, Italy. Antonio holds a Doctorate in Sciences from Universidad Autonoma de Mexico, Mexico.

Lynn Margulis Lynn is Distinguished University Professor, Department of Geosciences, University of Massachusetts, Amherst, MA, USA, and Co-Administrator of Planetary Biology Internship (PBI). Lynn is member of the National Academy of Sciences USA and the Russian Academy of Sciences. Lynn has also received many awards such as Darwin-Wallace Medal, Linnean Society of London, Alexander von Humboldt Prize, Berlin, among others. Lynn holds a PhD degree from Department of Genetics, University of California, Berkeley, USA.

Andrés Moya Andrés is Professor of Genetics at Universitat de València (Instituto Cavanilles de Biodiversidad y Biología Evolutiva), Valencia, Spain, and Head, Genomics and Health Area, Centre for Public Health Research (CSISP), Conselleria de Sanita. He is also President of Sociedad Española de Biología Evolutiva and has received awards such as „Ciudad de Barcelona de Investigación Científica“ and „Premio extraordinario de doctorado, Universidad de Valencia, Facultad de Filosofía y CC. de la Educación“. He holds a PhD from University of Valencia, Spain.

Maria Romeralo Maria is a Scholar at Department of Systematic Biology, Evolutionary Biology Centre, Uppsala University. Maria is also Member of the Sociedad Española de Biología Evolutiva. Maria holds a PhD in Biology from Complutense University, Madrid, Spain.

Bruce Scofield Bruce is Instructor of Psychology and Astronomy at Kepler College in Seattle, Washington, and Instructor of Environmental Evolution, University of Massachusetts. He is the author of several books on ancient Mesoamerican astronomy and astrology. He holds a PhD degree from Geosciences, University of Massachusetts, USA.

Vaughan Southgate Vaughan is currently President of the Linnean Society and Research Associate of the Department of Zoology at Natural History Museum. He has been Director of the World Health Collaborating Centre of the World Health Organization, Chairman of Meetings Committee and Honorary Secretariat of the Royal Society of Tropical Medicine & Hygiene and Editor of Transactions of the Royal Society of Tropical Medicine & Hygiene. Vaughan has been Member of Expert Advisory Panel on Parasitic Diseases (Schistosomiasis) for the World Health Organization. He holds a PhD from Molteno Institute, Christ's College, University of Cambridge, UK.

Carlos A. Valle Carlos is Professor of Evolution at the Biological and Environmental Sciences College, Universidad San Francisco de Quito, Ecuador, and Co-Director, Galapagos Academic Institute for the Arts and Sciences (GAIAS), Galapagos Islands, Quito, Ecuador. He is also Member Charles Darwin Foundation Board of Directors, for the Galapagos Islands and Member of the General Assembly of Charles Darwin Foundation for the Galapagos Islands. He holds a PhD from Princeton University, USA.

Shuhai Xiao Professor, Geosciences, Virginia Tech, Guest Professor, Peking University and other universities in China. Shuhai is also Associate Editor, *Solid Earth*, Deputy Editor-in-Chief, Journal of Stratigraphy, and editor of other journals in this field. He has received Charles Schuchert Award, Paleontological Society. He holds a PhD from Harvard University, USA.

Ada Yonath Director, The Helen and Milton A. Kimmelman, Center for Biomolecular Structure and Assembly, Structural Biology Department, Weizmann Institute of Science, Rehovot, Israel and Professor, Dept. of Structural Biology, Weizmann Institute of Science, Rehovot, Israel. Ada has received numerous awards among which the 2009 Nobel Prize for Chemistry, Honorary Doctorate, Oxford University, England, and The UNESCO-L'Oréal Award for European Woman in Life science, Paris. Ada holds a PhD from Weizmann Institute, Israel.

Index

G. Trueba, C. Montúfar (eds.), *Evolution from the Galapagos,* Social and Ecological Interactions in the Galapagos Islands 2,
DOI 10.1007/978-1-4614-6732-8, © Springer Science+Business Media New York 2013