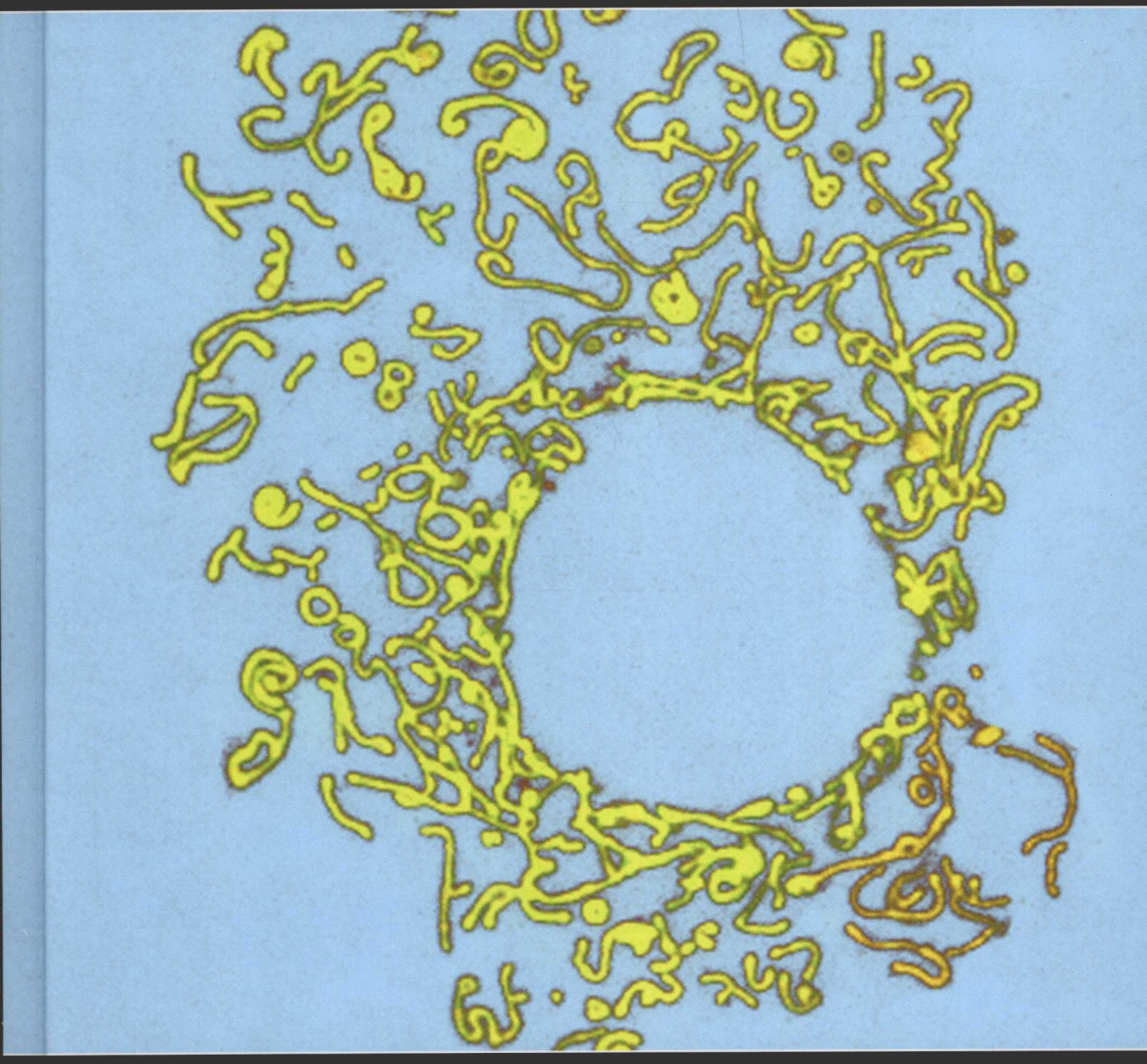
COLD SPRING HARBOR PERSPECTIVES IN BIOLOGY

Mitochondria



EDITED BY Douglas C. Wallace
Richard J. Youle

Mitochondria

A subject collection from *Cold Spring Harbor Perspectives in Biology*

EDITED BY

Douglas C. Wallace

Children's Hospital of Philadelphia
and
University of Pennsylvania

Richard J. Youle

National Institutes of Health

COLD SPRING HARBOR LABORATORY PRESS
Cold Spring Harbor, New York • www.cshlpress.org

Mitochondria

A Subject Collection from *Cold Spring Harbor Perspectives in Biology*
Articles online at www.cshperspectives.org

Printed in the United States of America

Executive Editor	Richard Sever
Managing Editor	Maria Smit
Project Manager	Barbara Acosta
Permissions Administrator	Carol Brown
Production Editor	Diane Schubach
Production Manager/Cover Designer	Denise Weiss
Publisher	John Inglis

Front cover artwork: Mitochondria in one cell are distributed throughout the cytosol and excluded from the central nucleus. When microtubule polymerization and mitochondrial divisions are inhibited, mitochondria form a variety of shapes revealing their plasticity. Image from Stephan Frank (Basel University Hospital, Switzerland).

Library of Congress Cataloging-in-Publication Data

Mitochondria / edited by Douglas C. Wallace, Children's Hospital of Philadelphia and University of Pennsylvania and Richard J. Youle, National Institutes of Health.
pages cm
"A subject collection from Cold Spring Harbor perspectives in biology."
Includes bibliographical references and index.
ISBN 978-1-936113-35-4 (hardcover : alk. paper)
1. Mitochondria. 2. Mitochondrial pathology. 3. Neuromuscular diseases. 4. Cell metabolism.
I. Wallace, Douglas C., editor of compilation. II. Youle, Richard, editor of compilation.

QH603.M5M52 2013
571.6'57--dc23

2013011220

10 9 8 7 6 5 4 3 2 1

Mitochondria

A subject collection from *Cold Spring Harbor Perspectives in Biology*

OTHER SUBJECT COLLECTIONS FROM *COLD SPRING HARBOR PERSPECTIVES IN BIOLOGY*

Signaling by Receptor Tyrosine Kinases
DNA Repair, Mutagenesis, and Other Responses to DNA Damage
Cell Survival and Cell Death
Immune Tolerance
DNA Replication
The Endoplasmic Reticulum
Wnt Signaling
Protein Synthesis and Translational Control
The Synapse
Extracellular Matrix Biology
Protein Homeostasis
Calcium Signaling
The Golgi
Germ Cells
The Mammary Gland as an Experimental Model
The Biology of Lipids: Trafficking, Regulation, and Function
Auxin Signaling: From Synthesis to Systems Biology
The Nucleus
Neuronal Guidance: The Biology of Brain Wiring
Cell Biology of Bacteria
Cell–Cell Junctions
Generation and Interpretation of Morphogen Gradients
Immunoreceptor Signaling
NF-κB: A Network Hub Controlling Immunity, Inflammation, and Cancer
Symmetry Breaking in Biology
The Origins of Life
The p53 Family

SUBJECT COLLECTIONS FROM *COLD SPRING HARBOR PERSPECTIVES IN MEDICINE*

Cystic Fibrosis: A Trilogy of Biochemistry, Physiology, and Therapy
Hemoglobin and its Diseases
Addiction
Parkinson's Disease
Type 1 Diabetes
Angiogenesis: Biology and Pathology
HIV: From Biology to Prevention and Treatment
The Biology of Alzheimer Disease

Preface

LIFE IS THE INTERPLAY BETWEEN STRUCTURE (anatomy), energy (vital force), and the information needed to construct anatomical and energetic systems. To understand biology and the human condition we must understand all three components of this life formula. However, for the past half millennium, biomedical science has been preoccupied with the anatomical differences between organisms and the chromosomal genetics that defines those anatomical differences. Although anatomical differences are essential for understanding differences between species, the important differences between individuals within a species are not anatomical. Hence, the most important factors for human variability are likely to be energetic. Yet only a tiny fraction of biomedical resources have been invested in understanding the energetic differences between people and the importance of energetic variation in health and disease. Perhaps, this is why the common diseases have remained so enigmatic and are routinely referred to as "complex."

This book is about the biology of energy. The prime protagonist is the mitochondrion. Mitochondria are the cellular organelles that produce most of our energy by oxidative phosphoryation (OXPHOS). The eukaryotic cell can be divided into two major functional domains: the nucleus–cytosol, which specializes in anatomy, and the mitochondrion, which specializes in energy. This division of labor is the result of a defining event in the biological history of higher organisms, the symbiosis that brought an oxidative bacterium, the α-proteobacterium, together with an anaerobic archaebacterium about two to three billion years ago, which ultimately yielded the eukaryotic cell. The increased energy provided by the oxidative bacterium permitted the archaebacterium to develop an increasingly large genome, which ultimately coalesced into the nucleus. The presence of multiple cytosolic bacteria provides large amounts of energy for the nucleus–cytosol such that the modern human cell cytoplasm harbors 100s to 1000s of oxidative bacteria and the nucleus encompasses more than 20,000 genes (see the chapter by Gray).

Over time the α-proteobacterium surrendered most of its structural genes to the nucleus. This decreased the energy requirements of the bacterium, permitting more of its energy to be contributed to the nucleus to elaborate an ever-increasing array of nuclear genes. By the time the fungal–animal lineage of multicellular animals evolved, the α-proteobacterium DNA had been reduced to about 14 polypeptide genes plus the rRNAs and tRNAs to sustain the residual bacterial protein-synthesis apparatus, thus creating the modern mitochondrion.

Today's mitochondrial DNA (mtDNA) codes for the core elements of mitochondrial OXPHOS, and thus represents the wiring diagram of the mitochondrial power plant. All of the remaining genes needed for mitochondrial bioenergetics, biogenesis, dynamics, and metabolism now reside in the eukaryotic nuclear DNA (nDNA). The dispersion of the mitochondrial genome between the 100s to 1000s of copies of the mtDNA plus the 1000 to 2000 nDNA genes creates the uniquely complex genetics of mitochondrial energy metabolism.

The cytosolic mitochondria and mtDNAs are confined to the cytosol and thus must be inherited through the cytoplasm. Hence, they are transmitted through the female oocyte and are maternally inherited. Because each cell has 1000s of mitochondria and mtDNAs, a mutant mtDNA can be mixed with normal mtDNAs within a cell at different ratios, a state known as heteroplasmy. Variation in both the nature of mitochondrial gene mutants and the percentage of mutant mtDNAs can affect energy production and human health (see the chapter by Wallace and Chalkia).

The gene products of the more than 1000 nDNA-coded mitochondrial genes must be returned to the mitochondrion. This is accomplished by a sophisticated system that tags proteins translated in the cytosol that are destined for the mitochondrion and selectively imports them into the mitochondrion (see the chapter by Stojanovski et al.).

Within the mitochondrion the mtDNA is packaged into nucleoids by the mitochondrial packaging protein and transcription factor, TFAM (see the chapter by Gilkerson et al.). The mtDNA is replicated within the mitochondrion by the mitochondrion-specific DNA polymerase γ, which has been found to be mutated in a wide spectrum of complex diseases (see the chapter by Stumpf et al.).

The biosynthetic processes of the mitochondrion and the nucleus–cytosol must be coordinated to coincide with the availability of sufficient energy resources to permit mitochondrial and cellular biogenesis and reproduction (see the chapter by Dominy and Puigserver). Hence, a sophisticated system has evolved for communication between the mitochondria and nucleus, which encompasses the epigenome and the cellular signal transduction pathways. The epigenome modulates nDNA gene expression and replication as well as controlling mtDNA replication and mitochondrial proliferation. The epigenome and the signal transduction systems are modulated by high-energy intermediates generated by the mitochondrion in response to environmental energy supplies. Hence, mitochondrial intermediates provide the barometer for the epigenome and the signal transduction systems to determine when there is sufficient energy and biosynthetic intermediates to permit the cell and the mitochondria to grow and multiply.

As mitochondria evolved within the nucleus–cytosol, this intracellular population of bacteria developed a complex system for coordinating their biochemistries. This is achieved by the exchange of gene products through repeated cycles of mitochondrial fusion and fission (see the chapter by van der Bliek et al.).

Because mitochondria continually replicate in cells, even within postmitotic ones, they also must be removed to avoid their becoming too numerous. Damaged mitochondria and mtDNAs must also be preferentially removed before they can accumulate in the cytoplasm and erode cellular function. This intracellular turnover of mitochondria is known as mitochondrial autophagy or mitophagy (see the chapter by Narendra et al.).

To have the maximum energy benefit, the mitochondria must be moved to the sites of greatest energy demand. This challenge is greatest in the neuron, where it is thought that the mitochondria are produced within the cell body but must migrate to the synaptic boutons to energize neurotransmission. Neuronal antigrade and retrograde transport is now understood to be mediated by a complex array of molecular motors and regulatory elements (see the chapter by Schwarz). Thus, mitochondrial fission, fusion, mitophagy, and transport all work together to maintain mitochondrial integrity and, thus, cellular and human health.

There are many instances in which cells must be removed from a tissue, either as a part of development, viral attack, cancer surveillance, or elimination of dysfunctional cells. This process is frequently mediated by an intracellular autodigestion process known as apoptosis (see the chapter by Bender and Martinou). The intrinsic pathway of apoptosis is initiated by mitochondria through their interaction with the Bcl-2 family of proteins. Apoptosis is an energy-requiring process, thus requiring functional mitochondria, and has the benefit of degrading cellular proteins before they can be released into the extracellular space.

The autodigestion of intracellular proteins and DNA is particularly pertinent for the mitochondrion, because mitochondria are the most prevalent bacteria within our bodies. If the cell simply ruptured it would release the mitochondrial antigens, known as mitochondrial damage-associated molecular patterns (DAMPs), into the circulation. This would activate the innate and adaptive immune systems, resulting in inflammation and disease. In cases of trauma where cells are damaged, mitochondrial energy production capacity is impaired, apoptosis fails, and the cell

undergoes necrosis, releasing the mitochondrial DAMPs, which leads to inflammation (see the chapter by Stehling and Lill).

Mitochondria lie at the nexus of a broad spectrum of critical cellular functions. Besides generating much of a cell's energy, mitochondria modulate Ca^{2+} levels, pH levels, multiple catabolic and anabolic pathways, redox homeostasis, reactive oxygen species (ROS) production and signaling, and initiation of the intrinsic pathway of apoptosis through activation of the mitochondrial permeability transition pore (mtPTP). Hence, mitochondrial dysfunction and mutations in mitochondrial genes can perturb a plethora of systems, resulting in a wide range of clinical symptoms (see the chapter by Stehling and Lill). For example, one function of the mitochondrion is to generate iron–sulfur centers for redox enzyme biochemistry. Perturbation of this pathway has been associated with a variety of metabolic diseases (see the chapter by Stehling and Lill). Similarly, mutations in a mitochondrial matrix Fe-containing sulfide dioxygenase block the pathway for detoxifying H_2S. The excess H_2S then inhibits cytochrome *c* oxidase resulting in mitochondrial dysfunction and disease (see the chapter by Tiranti and Zeviani).

Mitochondria also monitor the extracellular physical environment. This was demonstrated by the discovery that mutations in the extracellular matrix protein collagen VI sensitize the mtPTP. The resulting predisposition to activation of the intrinsic mitochondrial apoptosis pathway causes progressive myopathy (see the chapter by Bernardi and Bonaldo).

The central role of the mitochondrion in energy metabolism, intermediary metabolism, cellular homeostasis, and cell death and the unique features of mitochondrial dynamics and genetics have profound implications for both cellular and organismal biology. It is not surprising, then, that our rapidly expanding understanding of mitochondrial biology and genetics is providing new insights into cancer (see the chapter by Gasparre et al.) and aging (see the chapter by Sack and Finkel).

In conclusion, the study of mitochondrial biology and genetics has produced a new bioenergetic systems biology. This is providing powerful new insights into a wide variety of metabolic and degenerative diseases, cancer, and aging. Thus, the biology and genetics of bioenergetics may provide the missing elements necessary for us to elucidate the causes and generate cures for the common "complex" diseases.

Douglas C. Wallace
Richard J. Youle

Contents

Preface, vii

Mitochondrial Evolution, 1
Michael W. Gray

Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease, 17
Douglas C. Wallace and Dimitra Chalkia

Mechanisms of Protein Sorting in Mitochondria, 65
Diana Stojanovski, Maria Bohnert, Nikolaus Pfanner, and Martin van der Laan

The Mitochondrial Nucleoid: Integrating Mitochondrial DNA into Cellular Homeostasis, 83
Robert Gilkerson, Liliana Bravo, Iraselia Garcia, Norma Gaytan, Alan Herrera, Alicia Maldonado, and Brandi Quintanilla

Clinical and Molecular Features of *POLG*-Related Mitochondrial Disease, 93
Jeffrey D. Stumpf, Russell P. Saneto, and William C. Copeland

Mitochondrial Biogenesis through Activation of Nuclear Signaling Proteins, 111
John E. Dominy and Pere Puigserver

Mechanisms of Mitochondrial Fission and Fusion, 127
Alexander M. van der Bliek, Qinfang Shen, and Sumihiro Kawajiri

Mitochondrial Quality Control Mediated by PINK1 and Parkin: Links to Parkinsonism, 143
Derek Narendra, John E. Walker, and Richard Youle

Mitochondrial Trafficking in Neurons, 163
Thomas L. Schwarz

Where Killers Meet—Permeabilization of the Outer Mitochondrial Membrane during Apoptosis, 179
Tom Bender and Jean-Claude Martinou

The Role of Mitochondria in Cellular Iron–Sulfur Protein Biogenesis: Mechanisms, Connected Processes, and Diseases, 193
Oliver Stehling and Roland Lill

Altered Sulfide (H_2S) Metabolism in Ethylmalonic Encephalopathy, 211
Valeria Tiranti and Massimo Zeviani

Mitochondrial Dysfunction and Defective Autophagy in the Pathogenesis of Collagen VI Muscular Dystrophies, 223
Paolo Bernardi and Paolo Bonaldo

Relevance of Mitochondrial Genetics and Metabolism in Cancer Development, 235
Giuseppe Gasparre, Anna Maria Porcelli, Giorgio Lenaz, and Giovanni Romeo

Mitochondrial Metabolism, Sirtuins, and Aging, 253
Michael N. Sack and Toren Finkel

Index, 263

Mitochondrial Evolution

Michael W. Gray

Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3M 4R2, Canada

Correspondence: m.w.gray@dal.ca

Viewed through the lens of the genome it contains, the mitochondrion is of unquestioned bacterial ancestry, originating from within the bacterial phylum α-Proteobacteria (*Alphaproteobacteria*). Accordingly, the endosymbiont hypothesis—the idea that the mitochondrion evolved from a bacterial progenitor via symbiosis within an essentially eukaryotic host cell—has assumed the status of a theory. Yet mitochondrial genome evolution has taken radically different pathways in diverse eukaryotic lineages, and the organelle itself is increasingly viewed as a genetic and functional mosaic, with the bulk of the mitochondrial proteome having an evolutionary origin outside *Alphaproteobacteria*. New data continue to reshape our views regarding mitochondrial evolution, particularly raising the question of whether the mitochondrion originated after the eukaryotic cell arose, as assumed in the classical endosymbiont hypothesis, or whether this organelle had its beginning at the same time as the cell containing it.

In 1970, Lynn Margulis published *Origin of Eukaryotic Cells*, an influential book that effectively revived the long-standing but mostly moribund idea that mitochondria and plastids (chloroplasts) evolved from free-living bacteria via symbiosis within a eukaryotic host cell (Margulis 1970). The discovery in the 1960s of DNA within these organelles together with the recognition that they contain a translation system distinct from that of the cytosol were two of the observations that Margulis marshaled in support of the endosymbiont hypothesis of organelle origins. Indeed, throughout her career, Margulis forcefully argued that symbiosis is a potent but largely unrecognized and unappreciated force in evolution (Margulis 1981). Technological developments in DNA cloning and sequencing in the 1970s and 1980s opened the way to the detailed characterization of mitochondrial genomes and genes, and the generation of key molecular data that were instrumental in affirming a bacterial origin of the mitochondrial and plastid genomes, allowing researchers to pinpoint the extant bacterial phyla to which these two organelles are most closely related. Over the past several decades, numerous reviews have documented in detail the biochemical and molecular and cell biological data bearing on the endosymbiont hypothesis of organelle origins (Gray 1982, 1983, 1989a,b, 1992, 1993, 1999; Gray and Doolittle 1982; Wallace 1982; Cavalier-Smith 1987b, 1992; Gray and Spencer 1996; Andersson and Kurland 1999; Gray et al. 1999, 2001, 2004; Lang et al. 1999; Andersson et al. 2003; Burger et al. 2003a; Bullerwell and Gray 2004). Various endosymbiotic

Cite this article as *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a011403

models proposed over the years have been comprehensively critiqued (Martin et al. 2001), while the debates surrounding the endosymbiont hypothesis have been recounted in an engaging perspective that traces the development of ideas regarding organelle origins (Sapp 1994). Within a historical context, the present article emphasizes more recent data and insights that are relevant to continuing questions regarding how mitochondria originated and have since evolved.

WHAT DO GENETIC, GENOMIC, AND PHYLOGENOMIC DATA TELL US ABOUT THE ORIGIN OF MITOCHONDRIA?

The genetic function of mitochondrial DNA (mtDNA) was first fully revealed by complete sequencing of the ~16-kb mitochondrial genome from several mammalian species (Anderson et al. 1981, 1982; Bibb et al. 1981). This work established the paradigm that mtDNA encodes a small number (13 in mammals) of protein subunits of the mitochondrial electron transport chain and ATP synthase, as well as the ribosomal RNA (rRNA) and transfer RNA (tRNA) components of a mitochondrial translation system. The paradigm was revised when investigations of mtDNA from non-animal species showed quite extraordinary variation in size, physical form, coding capacity, organizational patterns, and modes of expression across the eukaryotic domain (Gray et al. 1998). In some non-animal taxa, additional proteins are encoded in mtDNA, notably extra respiratory proteins and ribosomal proteins (Table 1). On the other hand, some mitochondrial genomes have been reduced drastically in size, losing many of the protein genes encoded in animal mtDNA as well as some or all mtDNA-encoded tRNA genes. At ~6 kb in size, the mitochondrial genome of *Plasmodium falciparum* (human malaria parasite) and related apicomplexans is the smallest known, harboring only three protein genes, highly fragmented and rearranged small subunit (SSU) and large subunit (LSU) rRNA genes, and no tRNA genes (Feagin 2000). In marked contrast, within land plants, mtDNA has expanded substantially in size (>200 kb) if not in coding capacity, with the largest known mitochondrial genome in this lineage (~11,000 kb) exceeding the size of some bacterial and even some nuclear genomes (Sloan et al. 2012). What was evident even early on is that none of the initial mtDNAs investigated by detailed sequencing, including animal mtDNAs, looks anything like a typical bacterial genome in the way in which genes are organized and expressed.

Ribosomal RNA genes are among the few genes universally encoded by mtDNA across eukaryotes, and the corresponding rRNA sequences were exploited early on in phylogenetic reconstructions aimed at elucidating their evolutionary origin. Although mitochondrial large subunit (LSU) and small subunit (SSU) rRNAs are in general very different in size and secondary structure compared with their cytosolic and prokaryotic counterparts, they retain a sufficient degree of primary sequence and secondary structure correspondence that they can be incorporated into the aligned sequence databases on which these phylogenetic reconstructions are based (Gray et al. 1984, 1989; Cedergren et al. 1988). These initial phylogenetic trees showed that mitochondrial rRNA sequences, and presumably the genomes encoding them, emanate from within the domain Bacteria (eubacteria) and not from within Archaea (archaebacteria) or Eucarya (eukaryotes). Plant mitochondrial rRNAs are especially slowly evolving and bacteria-like (Schnare and Gray 1982; Spencer et al. 1984) and were instrumental in pinpointing the α-class of Proteobacteria (*Alphaproteobacteria*) as the specific bacterial lineage from within which they originated (Yang et al. 1985). Subsequent phylogenetic reconstructions using a much larger number of both mitochondrial and bacterial rRNA sequences have consistently confirmed this affiliation and, moreover, pointed to the Rickettsiales, one of six or more orders within *Alphaproteobacteria* (Williams et al. 2007)—and comprising obligate parasites such as *Rickettsia*, *Anaplasma*, and *Ehrlichia*—as especially close relatives of mitochondria (Gray and Spencer 1996; Gray 1998).

These phylogenetic conclusions were considerably bolstered by the sequencing in 1997 of a minimally derived (ancestral) mitochon-

Table 1. Mitochondrial DNA-encoded genes and their functions

(1) Coupled electron transport–oxidative phosphorylation (ATP synthesis)	
Complex I (NADH:ubiquinone oxidoreductase)	***nad1***, ***2***, ***3***, ***4***, ***4L***, ***5***, ***6***, *7*, *8*, *9*, *10*, *11*
Complex II (succinate:ubiquinone oxidoreductase)	*sdh 2*, *3*, *4*
Complex III (ubiquinol:cytochrome *c* oxidoreductase)	***cob***
Complex IV (cytochrome *c*:O_2 oxidoreductase)	***cox1***, ***2***, ***3***
Complex V (F_1F_0 ATP synthase)	*atp1*, *3*, ***4***,[a] ***6***, ***8***,[b] ***9***
(2) Translation	
Ribosomal RNAs	***rnl*** **(LSU)**, ***rns*** **(SSU)**, *rrr5* (5S)
Ribosomal proteins	
Small subunit (SSU)	*rps1*, *2*, *3*, *4*, *7*, *8*, *10*, *11*, *12*, *13*, *14*, *19*
Large subunit (LSU)	*rpl1*, *2*, *5*, *6*, *10*, *11*, *14*, *16*, *18*, *19*, *20*, *27*, *31*, *32*, *34*
Transfer RNAs	***trnA***, ***C***, . . . ***W***, ***Y***
Elongation factor	*tufA*
tm RNA (unstalling of translation)	*ssrA*
(3) Transcription	
Core RNA polymerase	*rpoA*, *B*, *C*
Sigma factor	*rpoD*
(4) RNA processing	
RNase P RNA (5′ tRNA processing)	*rnpB*
(5) Protein import	
ABC transporter	*ccmA* (*yejV*), *ccmB* (*yejW*)
Heme delivery	*ccmC* (*yejU*)
SecY-type transporter	*secY*
Sec-independent transporter	*tatA* (*mttA*)[c], *tatC* (*mttB*)
(6) Protein maturation	
Cytochrome oxidase assembly	*cox11*
Heme *c* maturation	*ccmF* (*yejR*)

A subset of the genes is encoded in the mtDNA of various eukaryotes; for example, those in bold are found in human and other mammalian mitochondrial genomes. Only the most ancestral (gene-rich) mtDNAs, for example, those of jakobid flagellates such as *Reclinomonas americana*, encode all or almost all of this gene set. Rare genes of uncertain origin and function include *dpo* (plasmid-derived DNA polymerase), *rpo* (plasmid-derived RNA polymerase), *rtl* (reverse transcriptase), *mutS* (DNA repair), and *dam* (methyltransferase). See Gray et al. (2004) for details.

[a]Also referred to as *orf25* and *ymf39* (Burger et al. 2003b).

[b]Also referred to as *orfB* and *ymf19* (Gray et al. 1998).

[c]Jacob et al. (2004).

drial genome from *Reclinomonas americana*, a member of an obscure group of protists (eukaryotic microbes) termed jakobid flagellates. The ~69-kb, circular-mapping mtDNA of *R. americana* not only contains more genes than any other characterized mitochondrial genome (including genes specifying 5S rRNA and RNase P RNA) (Table 1), but it also displays evident bacterial characteristics not seen in combination in any other mtDNA, such as operon-like gene clusters, highly bacteria-like rRNA and tRNA secondary structures, and putative Shine–Dalgarno motifs upstream of protein-coding genes. Indeed, *R. americana* mtDNA—so different from that of animal, plant, fungal, and many other protist mtDNAs—has been dubbed "a eubacterial genome in miniature," so striking is its resemblance to a typical bacterial genome (Lang et al. 1997).

At the same time, sequencing of the first Rickettsiales genome, that of *Rickettsia prowazekii*, showed it to be a markedly reduced bacterial genome, superficially resembling mitochondria in its dependence on a host cell (Andersson et al. 1998). However, the genomes of the mitochondrion and members of Rickettsiales are

clearly the products of independent evolutionary reduction, implying that mitochondria were not derived directly from a Rickettsiales taxon (Andersson et al. 1998; Gray 1998); rather, these two groups share a more remote common ancestor. Although the mitochondria–Rickettsiales connection has been a consistent phylogenetic finding (Viale and Arakaki 1994; Gupta 1995; Sicheritz-Pontén et al. 1998; Lang et al. 1999), it is still not certain whether the two are sister groups, or whether mitochondria actually branch within Rickettsiales, which is composed of two distinct families, Rickettsiaceae and Anaplasmataceae (Williams et al. 2007). In several studies, mitochondria appear to be more closely related to the former family, containing various *Rickettsia* species, than to the latter, comprising the genera *Anaplasma*, *Ehrlichia*, and *Wolbachia* (Karlin and Brocchieri 2000; Emelyanov 2001a,b, 2003a,b).

The specific affiliation of mitochondria and Rickettsiales in phylogenetic trees, although robust, has been questioned (e.g., Esser et al. 2004) on the grounds that this inferred relationship might be a phylogenetic artifact due to the high rate of sequence divergence and elevated A+T content of the genomes of Rickettsiales taxa and mitochondria—in other words, a long-branch-attraction (LBA) artifact. For that reason, there has been considerable interest in expanding the database of both mitochondrial and bacterial sequences and in applying a more comprehensive phylogenomics approach (based on many genes) to the reconstruction of phylogenetic trees, in combination with multiple and more sophisticated algorithms, rigorous statistical evaluation methods, and more realistic evolutionary models.

Of particular interest in broadening taxon sampling has been the identification of free-living members of *Alphaproteobacteria* that may be specifically affiliated with the parasitic group Rickettsiales (Williams et al. 2007). These previously unknown α-proteobacterial lineages came to light through the Global Oceanic Survey, metagenomic data from which revealed that oceanic α-Proteobacteria are abundant, with one particular clade (termed SAR11) comprising 30%–40% of total oceanic cell counts. The 1.3-Mb genome of one SAR11 member, *Candidatus Pelagibacter ubique*, is the smallest known genome, encoding the fewest genes, of any free-living bacterium (Giovannoni et al. 2005). Several recent reports have suggested that the SAR11 clade including *Ca. P. ubique* shares a common ancestor with mitochondria, the two together forming a sister group to Rickettsiales (Georgiades et al. 2011; Thrash et al. 2011). Other investigators reject this affiliation, concluding instead that *Ca. P. ubique* is most closely related to soil and aquatic α-Proteobacteria with large genomes (Viklund et al. 2012) and arguing that a rare oceanic group of α-Proteobacteria, termed Oceanic Mitochondria Affiliated Clade (OMAC), represents the closest free-living relatives to mitochondria (Brindefalk et al. 2011). Still other groups have suggested that free-living members of α-proteobacterial orders other than Rickettsiales should be considered as possible sources of the mitochondrial progenitor (Esser et al. 2004; Atteia et al. 2009).

These differing conclusions emphasize the challenges inherent in these sorts of analyses, which have to contend with various types of systematic error (for review, see Thrash et al. 2011), biased taxon sampling, and the highly restricted gene content of mitochondrial genomes. At present, although we know a great deal regarding the mitochondrial family tree, we have to admit that the identity of the immediate next of kin remains elusive.

Phylogenetic and phylogenomic reconstructions strongly and consistently support a monophyletic mitochondrial assemblage, and therefore a single origin of mitochondria. Two other pieces of genomic evidence support the view that extant mitochondrial genomes share a single common ancestor. First, the genes encoded by mitochondrial genomes are, with few exceptions, a subset of the genes encoded by the most gene-rich mtDNA, that of *R. americana*. Because the mitochondrial genome is considered to be a highly reduced version of a much larger α-proteobacterial progenitor genome, it is extremely unlikely that genome reduction in independently acquired bacterial symbionts would separately converge on the same small repertoire of respiratory-chain and ribosomal-

protein genes. Second, in plant and many ancestral protist mtDNAs, ribosomal protein genes are clustered in the same transcriptional order in which they appear in the corresponding bacterial operons. However, some genes present in the bacterial operons are missing from the corresponding mitochondrial gene clusters, either having been moved elsewhere in the mitochondrial genome or to the nuclear genome, as a result of endosymbiotic gene transfer (EGT). The same gene deletions are seen in the homologous mitochondrial ribosomal gene clusters where these occur throughout eukaryotes, arguing that these deletions were already present in the common ancestor of extant mitochondrial genomes (Gray et al. 1999). This conclusion is reinforced by consideration of various nucleus-encoded mitochondrial proteins that are components of the mitochondrial proteome (see below).

HOW DID THE MITOCHONDRIAL SYMBIOSIS HAPPEN?

Endosymbiotic models for the origin of mitochondria (for review, see Martin et al. 2001) are basically variations on two fundamentally different themes that have been referred to, respectively, as the "archezoan scenario" and the "symbiogenesis scenario" (Koonin 2010). In the archezoan scenario, "The host of the proto-mitochondrial endosymbiont was a hypothetical primitive amitochondrial eukaryote, termed archezoan" (Koonin 2010). In contrast, in the symbiogenesis scenario, "A single endosymbiotic event involving the uptake of an α-Proteobacterium by an archaeal cell led to the generation of the mitochondria," followed subsequently "by the evolution of the nucleus and compartmentalization of the eukaryotic cell" (Koonin 2010). The archezoan scenario most closely approximates the classical endosymbiont hypothesis of mitochondrial origin (Margulis 1970; Doolittle 1980), whereas the hydrogen hypothesis (Martin and Müller 1998) is an example of the symbiogenesis scenario. A fundamental difference between these two scenarios is whether the α-Proteobacterial endosymbiosis that gave rise to the proto-mitochondrion happened at the same time as and was integral to the formation of the eukaryotic cell (symbiogenesis scenario) or occurred subsequent to the formation of a primitive, amitochondriate cell that served as host and that was already essentially eukaryotic (archezoan scenario).

Archezoan Scenario

A major boost to the classical endosymbiont hypothesis came from early phylogenetic reconstructions, based initially on SSU rRNA sequences, that showed several eukaryotic lineages branching deeply within the eukaryotic domain of the resulting trees. These early-branching lineages, collectively termed "archezoa," consisted of protists such as microsporidians, diplomonads, and parabasalids, living as parasites in anaerobic environments and characterized by the absence of recognizable mitochondria (Cavalier-Smith 1987a, 1989). Members of archezoa, assumed to be primitively amitochondriate (i.e., never having had mitochondria in their evolutionary history: the so-called "archezoa hypothesis"), could be seen as modern representatives of the sort of ancestral eukaryote that might have played host to an α-Proteobacterial symbiont in a classical endosymbiosis scheme.

Two findings led to the ultimate abandonment of the archezoa hypothesis. First, more rigorous phylogenetic reconstructions combined with other sorts of data convincingly showed that the early-branching position of archezoan taxa in the eukaryotic tree is a methodological artifact, due to a relatively high rate of sequence divergence of the archezoan sequences used in the analysis. This property gives rise to an LBA effect that incorrectly positions these taxa at the base of the eukaryotic clade, closest to the outgroup (prokaryotic) sequences used to root the tree. Microsporidia, for example, were eventually recognized as secondarily degenerate fungi rather than as primitive, early-branching eukaryotes (Hirt et al. 1999; Keeling et al. 2000). Indeed, the current eukaryotic tree is more accurately characterized as a bush, with no one lineage clearly the earliest diverging (Keeling et al. 2005; Koonin 2010). Second, in every apparently amitochondrial lineage that

has been carefully investigated, evolutionary remnants of mitochondria ("mitochondrion-related organelles," or MROs) have been identified (see below). Thus, we currently know of no extant eukaryotic lineages that are convincingly amitochondrial and that therefore might have been primitively amitochondriate (Embley and Hirt 1998). This conclusion does not mean that such lineages do not exist. We may simply not have discovered them yet, or they may have existed at some point in evolutionary history but have now all become extinct. Nevertheless, rejection of the archezoa hypothesis on phylogenetic grounds coupled with the apparent absence of any known amitochondriate eukaryotic lineages considerably weakens the case for an acquisition of mitochondria via the classical endosymbiont route.

Symbiogenesis Scenario

An alternative view, that the host cell for the mitochondrial endosymbiosis was a prokaryote—and specifically an archaeon, not a eukaryote—has recently gained in prominence (Koonin 2010). The "hydrogen hypothesis" (Martin and Müller 1998) is perhaps the best-known example of a symbiogenesis scenario. This scheme proposes that eukaryotes arose

> . . .through symbiotic association of an anaerobic, strictly hydrogen-dependent, strictly autotrophic archaebacterium (the host) with a eubacterium (the symbiont) that was able to respire, but generated molecular hydrogen as a waste product of anaerobic heterotrophic metabolism. The host's dependence upon molecular hydrogen produced by the symbiont is put forward as the selective principle that forged the common ancestor of eukaryotic cells. (Martin and Müller 1998)

Thus, the hydrogen hypothesis "posits that the origins of the heterotrophic organelle (the symbiont) and the origins of the eukaryotic lineage are identical" (Martin and Müller 1998). A corollary of the hydrogen hypothesis and other symbiogenesis scenarios is that the complexity of the eukaryotic cell and its defining features emerged *after* the mitochondrial symbiosis, rather than *before*.

Several arguments can be advanced against a symbiogenesis scenario for the origin of mitochondria (Koonin 2010). Endocytosis (a eukaryotic hallmark) has long been considered an essential function for incorporating a bacterial symbiont, although it is the case that bacterial endosymbioses (e.g., γ-Proteobacteria inside β-Proteobacteria) have been documented (von Dohlen et al. 2001; Thao et al. 2002). In addition, assuming an α-Proteobacterial symbiont as the mitochondrial progenitor in a partnership that simultaneously gave rise to this organelle and the rest of the eukaryotic cell, one might expect that any eubacterial genetic signal in the nuclear genome would be overwhelmingly α-Proteobacterial. However, although an α-Proteobacterial signal does, in fact, predominate (Pisani et al. 2007), in any given eukaryotic lineage collectively more bacterial-type genes appear to derive from diverse non-α-Proteobacterial lineages or fail to affiliate robustly with any specific bacterial phylum. Nevertheless, it is possible that ancestral lineages contributing to a bacterial–archaeal symbiogenesis might have possessed genomes already "scrambled" to a certain extent by horizontal gene transfer (HGT).

A prominent feature of the hydrogen hypothesis is its claim to account simultaneously for the origins of both aerobic and anaerobic energy metabolism in eukaryotes, the assumption being that both pathways were contained in and contributed to the hybrid cell by the α-Proteobacterial partner. It is supposed that the two pathways would have been differentially expressed in the free-living bacterial symbiont when it encountered the appropriate environmental conditions. The hypothesis further posits that genes for aerobic respiration were lost in those eukaryotic lineages in which the mitochondrion was converted to an MRO, some types of which (e.g., hydrogenosome) function in anaerobic energy metabolism (see below). The hydrogen hypothesis predicts that genes of anaerobic energy metabolism in MROs should have been inherited vertically throughout eukaryotes from a common ancestor, clustering as a monophyletic lineage together with α-Proteobacteria in phylogenetic reconstructions. However, a rigorous study of the

phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes has not provided support for this prediction (Hug et al. 2010). Rather, MROs and the enzymatic machinery they contain for anaerobic energy metabolism appear to reflect a high degree of independent and convergent evolution (see below).

Very recently, Lane and Martin (2010) have argued, from a consideration of the energetics of genome complexity, that because eukaryotes encode and express a substantially larger number of proteins than do prokaryotes, this increased expression demands a level of cellular energy that only the mitochondrion is able to satisfy. Accordingly, these investigators view the mitochondrion as the *sine qua non* to eukaryotic genomic and cellular complexity, concluding, rather definitively, that "the host for mitochondria was a prokaryote."

On balance, it would seem that a symbiogenesis scenario (bacterial endosymbiont in an archaeal host) better accommodates the accumulated data relevant to the question of mitochondrion origin than does an archezoan scenario (bacterial endosymbiont in an amitochondriate but essentially eukaryotic host). However, as emphasized above, the latter scenario cannot be entirely discounted at this point. Each scenario raises its own set of issues that are difficult to rationalize without resorting to ad hoc explanations. In the end, each faces the conundrum that there is no straightforward and compelling way to discern how similar the genomes of the proposed prokaryotic ancestors of the eukaryotic cell were to their modern-day descendants.

MITOCHONDRION-RELATED ORGANELLES (MROs)

An extreme form of mitochondrial genome reduction is found in two types of MROs, hydrogenosomes and mitosomes, which entirely lack mtDNA. These two MRO types are distinguished by the fact that hydrogenosomes retain ATP-generating capacity, whereas mitosomes do not. The hydrogenosome, a double-membrane-bound organelle originally discovered in the parabasalid *Trichomonas vaginalis* (Lindmark and Müller 1973), was subsequently identified in various other anaerobic eukaryotes. The *T. vaginalis* hydrogenosome not only lacks a genome but has an incomplete tricarboxylic acid cycle and electron transport chain and no cytochromes. ATP is generated from pyruvate via a substrate-level pathway comprising a characteristic suite of enzymes, including pyruvate: ferredoxin oxidoreductase and an iron–iron hydrogenase. The organelle takes its name from the fact that molecular hydrogen (H_2) is one of the end products of this pathway. The anaerobic metabolism performed by the hydrogenosome initially suggested that the organelle might have originated through an endosymbiosis with an anaerobic bacterium such as *Clostridium* (Whatley et al. 1979). However, subsequent studies have revealed that the *T. vaginalis* hydrogenosome contains several proteins typical of mitochondria, including chaperonins (Bui et al. 1996), the NADH dehydrogenase module of electron transport complex I (Hrdy et al. 2004), and components of the ISC biosynthesis pathway, the mitochondrial machinery for synthesis of iron–sulfur (Fe–S) clusters (Sutak et al. 2004). These results strongly support the view that the *T. vaginalis* hydrogenosome is a relict mitochondrion.

A second group of double-membrane-bound MROs, in this case unable to generate ATP, has been discovered in a number of anaerobic, parasitic protists that were initially considered to be amitochondriate, including the amoebozoons *Entamoeba histolytica* (Clark and Roger 1995; Mai et al. 1999; Tovar et al. 1999) and *Mastigamoeba balamuthi* (Gill et al. 2007), the microsporidians *Trachipleistophora hominis* (Williams et al. 2002) and *Encephalitozoon cuniculi* (Goldberg et al. 2008; Tsaousis et al. 2008), and the diplomonad *Giardia lamblia* (Tovar et al. 2003). The MROs in these protists are collectively termed "mitosomes" (Embley et al. 2003; Embley 2006; Hjort et al. 2010). Here again, identification of typical mitochondrial proteins in this MRO argues that mitosomes, like hydrogenosomes, are evolutionary derivatives of conventional mitochondria, but are even more highly reduced than hydrogenosomes (see Hjort