

Molecular Data Are Transforming Hypotheses on the Origin and Diversification of Eukaryotes

YONAS I. TEKLE, LAURA WEGENER PARFREY, AND LAURA A. KATZ

The explosion of molecular data has transformed hypotheses on both the origin of eukaryotes and the structure of the eukaryotic tree of life. Early ideas about the evolution of eukaryotes arose through analyses of morphology by light microscopy and, later, electron microscopy. Though such studies have proven powerful at resolving more recent events, theories on origins and diversification of eukaryotic life have been substantially revised in light of analyses of molecular data including gene and, increasingly, whole-genome sequences. By combining these approaches, progress has been made in elucidating the origin and diversification of eukaryotes. Yet many aspects of the evolution of eukaryotic life remain to be illuminated.

Keywords: chimeric genome, endosymbiosis, lateral gene transfer, eukaryotic supergroups, molecular systematics

Eukaryotes—organisms composed of one or more cells with nuclei—have inhabited Earth for approximately 1.2 billion to 1.8 billion years (Knoll et al. 2006), are major players in biogeochemical cycling, and cause numerous global diseases (e.g., malaria, African sleeping sickness, amoebic dysentery). The most familiar eukaryotes—plants, animals, and fungi—dominate the visible landscapes of terrestrial systems. Yet these three lineages represent only a small fraction of the estimated 35 to 55 eukaryotic lineages that may be of comparable age (figure 1; Patterson 1999). The small size of the organisms in many of the remaining lineages has traditionally made these groups recalcitrant to study. Moreover, the vast timescale of eukaryotic evolution obscures evolutionary events during their origin and early diversification. Despite these challenges, advances in molecular techniques are making these diverse lineages more accessible and hence are transforming our views on eukaryotic evolution.

In this article we focus on the recent transformation of hypotheses on the origin and diversification of eukaryotes that has occurred with the rise in molecular data (i.e., multigene sequencing, genomics). This increase in molecular data is the result of a confluence of factors, including the development of the polymerase chain reaction and subsequent improvements in high-throughput sequencing of expressed sequences and whole genomes. We first focus on the origin

of eukaryotic features, highlighting the compelling evidence for the chimeric nature of eukaryotic genomes. We then describe the state of knowledge on eukaryotic phylogeny (see box 1), where there has been considerable progress in establishing robust clades, though deep nodes remain elusive. Together, this synthesis highlights both the progress and challenges to our understanding of the evolution of eukaryotic life.

Origins

Hypotheses on the origin of eukaryotic cells must address the evolution of key features that distinguish eukaryotes from bacteria and archaea (see box 1). One such innovation, the presence of a nucleus, is the defining feature of eukaryotes and gives this clade its name: eu = “true” and karyon = “kernel” (referring to the nucleus). Hypotheses on the origin of eukaryotes must further explain the presence of a chimeric genome—a genome in which different genes share most recent common ancestry with varying archaeal and bacterial lineages. A further and perhaps more significant innovation of eukaryotic cells is the cytoskeleton, a network of protein-rich structures underlying eukaryotic morphology and mobility. Bacteria contain homologues of some eukaryotic cytoskeletal proteins, as explained below, but the mechanism underlying the large divergences between these homologues is unclear. Two additional features found in many, but not all,

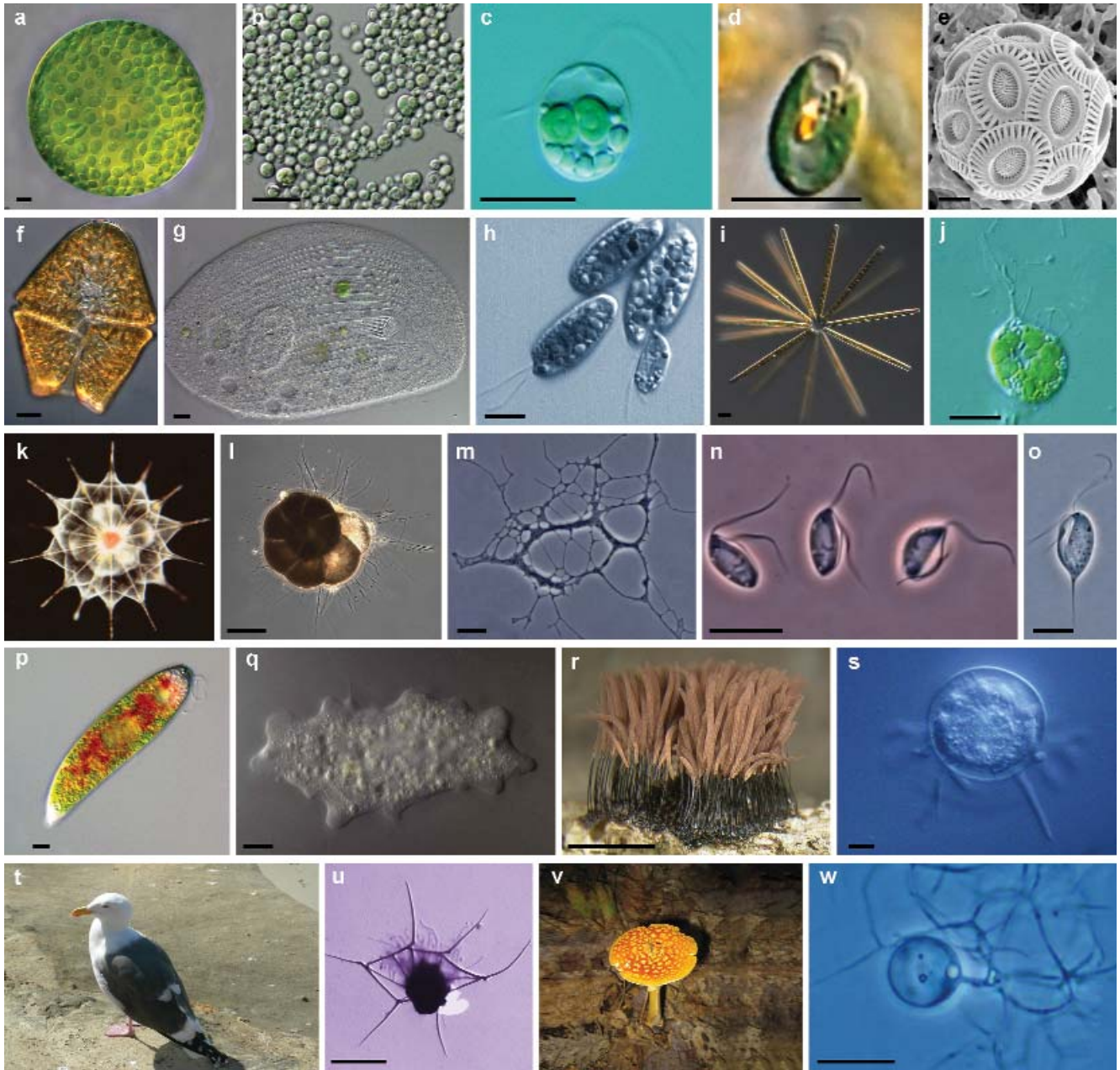


Figure 1. Representative eukaryotic lineages from the six putative supergroups: (a–c) ‘Plantae’: (a) *Eremosphaera viridis*, a green alga; (b) *Cyanidium* species, red algae; (c) *Cyanophora* species, a glaucophyte. (d–i) ‘Chromalveolata’: (d) *Chroomonas* species, a cryptomonad; (e) *Emiliana huxleyi*, a haptophyte; (f) *Akashiwo sanguinea*, a dinoflagellate; (g) *Trithigmostoma cucullulus*, a ciliate; (h) *Colpodella perforans*, an apicomplexan; (i) *Thalassionema* species, colonial diatom (Stramenopile). (j–m) ‘Rhizaria’: (j) *Chlorarachnion reptans*, an autotrophic amoeba (Cercozoa); (k) *Acantharea* species, a radiolarian; (l) *Ammonia beccarii*, a calcareous foraminifera; (m) *Corallomyxa tenera*, a reticulate amoeba. (n–p) ‘Excavata’: (n) *Jakoba* species, a jakobid with two flagella; (o) *Chilomastix cuspidata*, a flagellate retortamonad; (p) *Euglena sanguinea*, an autotrophic Euglenozoa. (q–s) ‘Amoebozoa’: (q) *Trichosphaerium* species, a naked stage (lacking surface spicules) of an unusual amoeba with alternation of generations, one naked and one with spicules; (r) *Stemonitis axifera*, an acellular slime mold; (s) *Arcella hemisphaerica*, a testate amoeba. (t–w) *Opisthokonta*: (t) *Larus occidentalis*, a bird; (u) *Campyloacantha* species, a choanoflagellate; (v) *Amanita flavoconia*, a basidiomycete fungus; (w) *Chytriomycetes* species, a chytrid. All scale bars = 10 micrometers, except bars for (b) and (l) = 100 micrometers and the bar for (r) = 5 millimeters. All images are provided by *micro*scope* (<http://starcentral.mbl.edu/microscope/portal.php>) except (v), which was provided by James Parfrey.

eukaryotes are mitochondria and chloroplasts (see box 1). Some genes from these endosymbiotic organelles have been incorporated into the host cell nucleus through endosymbiotic gene transfer (Martin and Schnarrenberger 1997), further increasing the complexity of eukaryotic genomes.

Certainly, numerous insightful hypotheses on the origin of eukaryotes predate the recent flood of molecular data. These hypotheses can be divided into two major categories: (1) hypotheses involving endosymbiosis, which argue that components of the eukaryotic cell arose by engulfment of prokaryotic organisms (e.g., Sagan 1967, Taylor 1987), and (2) hypotheses for autogenous (“self-birth”) pathways for eukaryotic cell components (e.g., Cavalier-Smith 1978); these earlier hypotheses have been extensively reviewed in the literature (e.g., Roger 1999). Hence, we focus here on the impact of molecular data on hypotheses on eukaryotic origins and evolution.

The chimeric nature of eukaryotic genomes

Eukaryotic genomes are chimeric, as evidenced by varying eukaryotic genes that share common ancestry with diverse archaeal and bacterial lineages. Evidence of the chimeric nature of eukaryotic genomes represents one of the most significant transformations in theories on the origins and diversification of eukaryotes. Rather than evolving in a strictly vertical (treelike) manner, eukaryotic genes have complex histories that very likely involve transfers from endosymbionts, from ingested microbes, or from the original partners in the evolution of the eukaryotic cell (figure 2).

If eukaryotic genomes arose strictly through vertical descent from a common ancestor, we might expect the bulk of eukaryotic genes to trace back to the common ancestor in a single noneukaryotic lineage, depending on patterns of gene gain and loss. In contrast, if endosymbiotic hypotheses are correct and the chimerism of the eukaryotic genome resulted from fusion between two distinct genomes (i.e., one archaeal and one bacterial lineage), we would expect eukaryotic genes to trace back to two, or another relatively small number, of host and donor lineages. However, analyses of individual genes (Brown and Doolittle 1997) and, more recently, whole-genome data (figure 2; Dagan and Martin 2006) reveal a much more complex history of eukaryotic genes, with many potential donor lineages. The chimerism of eukaryotic genomes is argued to be due to a fusion event at the time of the origin of eukaryotes, coupled with continuous lateral transfer of genes from prey organisms into the eukaryotic genome (figure 2; “You are what you eat,” Doolittle 1998). As a result of these processes, different genes within eukaryotic genomes are more similar to varying lineages of bacteria and archaea, as indicated by BLAST (basic local alignment search tool) analysis of 5833 human proteins (figure 2j; Dagan and Martin 2006).

One recent study using a supertree (see box 1) approach did find evidence of substantial input from specific prokaryotic lineages, listed in order of relative contribution: cyanobacteria, proteobacteria, and thermoplasmatales (an order

Box 1. Glossary of key terms.

Archaea: Microbial clade originally defined using culture-independent studies of extreme environments. Now known to be widespread, archaea constitute one of the three domains of life and are defined by ether linkages in their membranes.

Bacteria: Metabolically diverse microbial domain that includes many familiar lineages such as *Escherichia coli* and *Staphylococcus aureus*. The ancestors of both mitochondria and chloroplasts are believed to belong to this domain.

Chloroplast (see *plastid*, below): A membrane-bound organelle in green algae (plant) cells that contains chlorophyll.

Heterogeneous rates of evolution (*sequence rate heterogeneity*): The variation in rates of substitutions among species.

Homology (*homologous*): Similarity in structure or sequence because of common ancestry.

Lateral inheritance (*lateral or horizontal gene transfer*): Genetic inheritance between nonsister lineages; transfer of genes between species.

Mitochondrion: A double membrane-bound organelle primarily involved in oxidative phosphorylation (energy [adenosine triphosphate, or ATP] production) in eukaryotes.

Molecular systematics: A branch of systematics (evolutionary relationships of organisms) that uses molecular data.

Monophyly (*monophyletic group, also known as a clade*): A group of organisms comprising a common ancestor and all of its descendants.

Phylogenomics: In a phylogenetic context, large-scale reconstruction of a species tree using genomic data (usually 100+ genes).

Phylogeny: Evolutionary relationship among organisms or their components (e.g., genes).

Plastid: A double membrane-bound organelle, generally involved in photosynthesis, storage (starch), and synthesis of several macromolecules (e.g., fatty acids). Plastids include chloroplasts and chromoplasts.

SSU-rRNA (*small subunit ribosomal RNA*): An RNA molecule that is one component of ribosomes, which are involved in protein synthesis. SSU-rRNA is extensively used in phylogenetic studies because of its high sequence conservation across all domains of life.

Supertree: A tree assembled from several smaller phylogenies that share some, but not necessarily all, terminal taxa.

Synapomorphy: A derived character shared by two or more groups that originated in their last common ancestor.

Taxonomy: The practice or science of classification.

Ultrastructural identities: Subcellular features (mitochondria, plastids, flagella, root system, nucleus, types of hair, etc.), obtained through electron microscopy, that have been used to define lineages of eukaryotes (e.g., alveolar sacs in alveolates and tripartite hairs in stramenopiles).

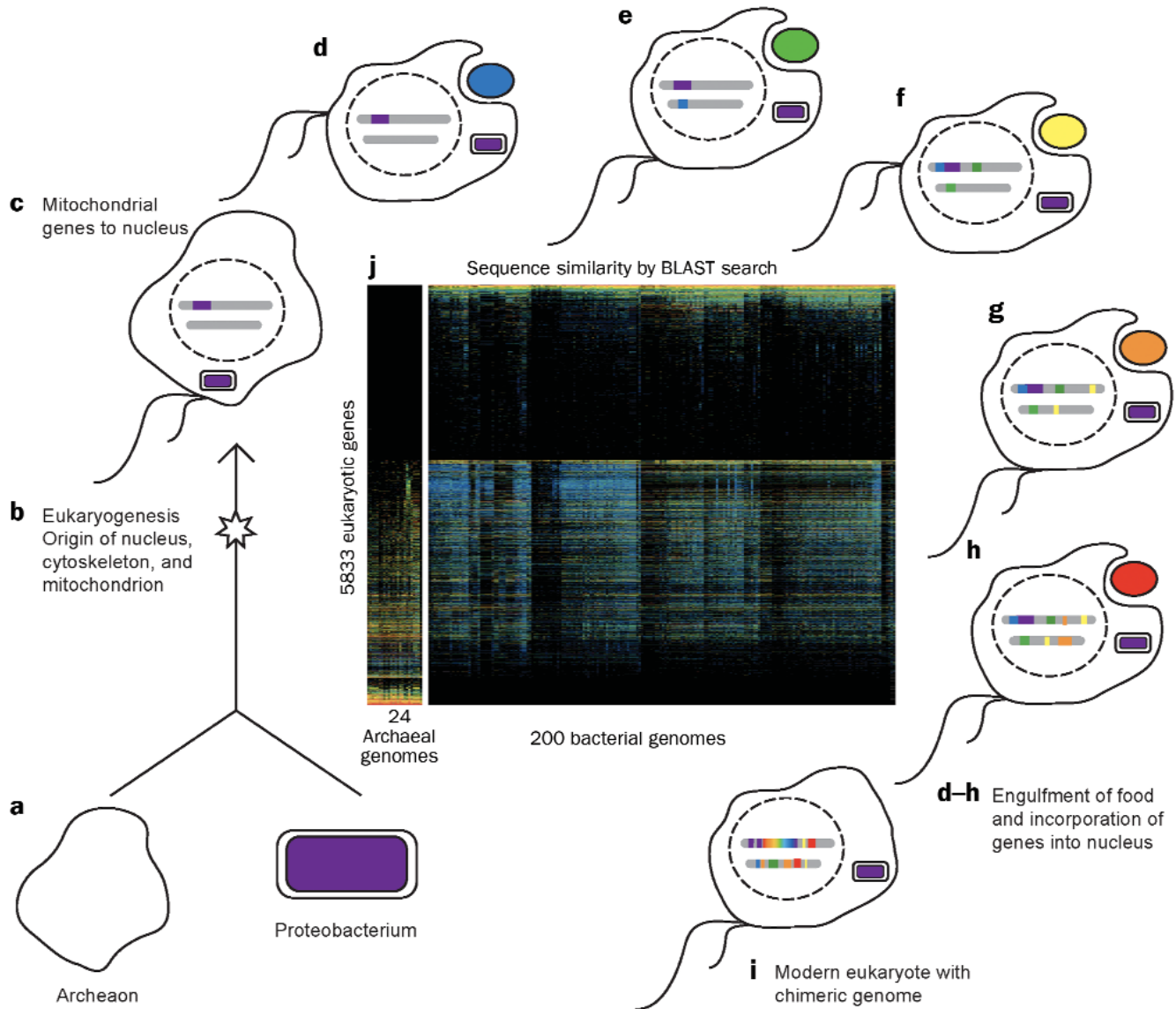


Figure 2. The chimeric eukaryotic genome is consistent with a fusion of an archaeon and a bacterium at the time of the origin of eukaryotes, coupled with subsequent aberrant lateral transfers of genes from food items. (a) An archaeon and proteobacterium that are potential symbiotic partners in origin of eukaryotes. (b) Eukaryogenesis, the origin of the nucleus, cytoskeleton, and mitochondria through unknown mechanisms and events. (c) Mitochondrial genes are transferred to the host nucleus (purple portions of chromosome). (d) Eukaryote engulfs blue food item and incorporates blue genes into host nucleus (in [e]). (e–h) Repeated engulfment of food and incorporation of genes into the host nucleus (hot colors represent more recent events). (i) Modern eukaryote whose chimeric genome is the product of (a)–(h). (j) This figure, modified from Dagan and Martin (2006), depicts the mixed ancestry of eukaryotic genes as elucidated from BLAST (basic local alignment search tool) similarity searches. The 5833 genes from the human genome (rows) were compared with 24 archaeal (columns in the first block) and 200 bacterial genomes (columns in the second block). Colors indicate the level of similarity between eukaryotic and bacterial and archaeal homologues. Warmer colors represented higher BLAST scores, and therefore a higher degree of similarity. Source: Inset modified from Dagan and Martin (2006).

within the Archaea domain; Pisani et al. 2007). Aspects of this pattern are consistent with several of the hypotheses on the origin of eukaryotes that involve a merger between archaeal and proteobacterial genomes at the origin of eukaryotes (e.g., Zillig et al. 1989, Searcy 1992, Martin and Müller 1998, Margulis et al. 2000), coupled with subsequent transfer of genes from cyanobacterial-derived plastids (see box 1 and the

sections on mitochondria and chloroplasts, below). Interpretation of possible donor lineages to the chimeric eukaryotic genome is complicated by the high levels of lateral gene transfer (LGT; see box 1) among bacterial and archaeal lineages over the past billion years or so since eukaryotes evolved; this, combined with deep divergence times, makes phylogenetic inference challenging. Nevertheless, analyses of

completed genomes indicate that the evolutionary history of eukaryotic genes is complex (figure 2) and involves a combination of endosymbiotic gene transfers and aberrant LGTs (defined here as transfers that are not from hosts or symbionts but rather from food, the environment, or another, undetermined source).

Origin of the nucleus

Determining the origin of the nucleus is perhaps one of the most difficult challenges for theories on the origin of eukaryotic cells. The nucleus, which contains the eukaryotic genome, is bounded by the nuclear envelope (with the outer membrane contiguous with the endoplasmic reticulum). Numerous complex components are contained within the nucleus, such as nucleopores and the transcriptional and spliceosomal machinery. As with the case for cytoskeleton described below, hypotheses on the origin of the nucleus must explain the evolution of this complex structure that exists in all extant eukaryotes—there are no clear intermediate taxa for a transition from prokaryote to eukaryote. As with virtually any eukaryotic feature, there are hypotheses that propose endosymbiotic (e.g., Margulis et al. 2000, Hartman and Fedorov 2002) and autogenous (e.g., Cavalier-Smith 2002, Jekely 2007) origins for the nucleus and its associated proteins.

One intriguing recent hypothesis on why eukaryotes evolved a nucleus is that there was selection for a temporal separation between transcription and translation because of the invasion of self-splicing group II introns, which in turn may be precursors to eukaryotic spliceosomal introns (introns removed by the spliceosome) (Martin and Koonin 2006). A comprehensive phylogenomic (see box 1) analysis of proteins in the nuclear envelope, including nucleoporins, lamins, and karyopherins, revealed few clear homologues among bacteria or archaea (Mans et al. 2004). Hence, the challenge remains to develop hypotheses that can explain the intermediate steps that enabled the evolution of the highly complex eukaryotic nucleus.

Origin of the cytoskeleton

Hypotheses on the eukaryotic cytoskeleton have shifted from a focus on the origin of cytoskeletal proteins to questions of how cytoskeletal proteins evolved from homologues in bacteria and archaea into the complex system present in all extant eukaryotes. Detailed comparisons of protein structures have yielded evidence of homologues for numerous cytoskeletal proteins. Most notably, bacterial homologues have been identified for both tubulin and actin, named FtsZ and MreB, respectively (reviewed in Erickson 2007). The bacterial and eukaryotic versions of these proteins are highly divergent, with amino acid similarities ranging from only approximately 10% to 40% (Erickson 2007; see Jenkins et al. 2002 for the special case of *Prostheco bacter*). Contrary to the theory that eukaryotic cytoskeletal proteins are derived from a spirochaete endosymbiont (Margulis 1993, Margulis et al. 2000, 2006), there is no compelling evidence for any specific bacterial donor lineage for eukaryotic cytoskeletal proteins

(Mitchell et al. 2007). Further, neither bacterial nor archaeal homologues have yet been identified for some cytoskeletal proteins, including the diverse myosins (Richards and Cavalier-Smith 2005). Hence, the challenge for theories on the origin of the eukaryotic cytoskeleton is explaining the steps that generated the highly coordinated and complex eukaryotic cytoskeleton found in all extant eukaryotes (Doolittle 1995). Progress in this area will emerge as genome sequences are completed from diverse eukaryotes, coupled with developments in models of protein evolution that allow identification of potential bacterial and archaeal homologues of additional cytoskeletal proteins.

Early acquisition of mitochondrion

There is compelling evidence that the mitochondrion resulted from endosymbiosis of an alpha-proteobacterium, and that during the reduction of the mitochondrial genome, some genes were transferred to the nucleus, contributing to the chimerism of eukaryotic genomes (reviewed in Roger 1999, Embley and Martin 2006). In contrast, the timing of the acquisition of mitochondria has been subject to much debate. Under the Archezoa hypothesis, the ancestral eukaryote lacked mitochondria and the organelle was acquired only after the evolution of numerous amitochondriate lineages (Cavalier-Smith 1983). Early analyses of small subunit ribosomal RNA (SSU-rRNA; see box 1) yielded topologies that were consistent with the Archezoa hypothesis, with amitochondriate lineages such as diplomonads (e.g., *Giardia lamblia*), trichomonads (e.g., *Trichomonas vaginalis*), and microsporidians (e.g., *Encephalitozoon cuniculi*) falling at the base of the eukaryotic tree of life (e.g., Sogin et al. 1989). However, these early trees were based on only a single gene (SSU-rRNA) and were limited by both the number of characters available for analysis (as opposed to more recent multigene studies) and the methodological challenges of dealing with highly heterogeneous rates of evolution in phylogenies.

The taxonomic content of Archezoa changed over time with the emergence of more data and improved analytical tools that suggested that amitochondriate lineages either were not basal or are derived from mitochondrial-containing lineages (see Patterson 1999, Roger 1999, Embley et al. 2003). In addition, the Archezoa hypothesis has been challenged by the presence of genes of mitochondrial origin within the nuclei of putative Archezoa, as well as by the discovery of diminutive organelles such as hydrogenosomes (hydrogen-producing organelles) and mitosomes (degenerate mitochondria of uncertain function) in amitochondriate eukaryotes (Embley et al. 2003). Falsification of the Archezoa hypothesis on the basis of extant amitochondriate eukaryotes leads to the hypothesis that the acquisition of mitochondria occurred at the time of, or soon after, the origin of eukaryotes (e.g., Katz 1998, Gray 1999, Roger 1999). Under such a scenario, this early acquisition of mitochondria contributed to the present-day chimerism of the eukaryotic genome.

Multiple endosymbioses of plastids

Currently, many scientists accept a model of a single primary endosymbiotic origin of plastids into the last common ancestor of red algae, green algae, and glaucocystophytes (a clade of freshwater algae). Under this scenario, all other photosynthetic eukaryotic lineages evolved by subsequent secondary, tertiary, and perhaps even quaternary endosymbiosis, whereby a eukaryote engulfs another eukaryote. Evidence for a single primary acquisition, in which a eukaryote engulfed a cyanobacterium, includes monophyly of plastid genes from diverse eukaryotes (box 1; Bhattacharya et al. 2004, McFadden and van Dooren 2004) and shared transport pathways (McFadden and van Dooren 2004, Steiner et al. 2005). However, other data complicate the hypothesis of a single primary acquisition of chloroplasts in eukaryotes, including the non-monophyly of plastid rubisco genes (Delwiche and Palmer 1996), the nonmonophyly of the putative hosts of this endosymbiosis (Yoon et al. 2008), and the diversity of pigments and light-harvesting complex proteins among photosynthetic eukaryotes (Larkum et al. 2007).

Similarly, there is debate between those who view only a small number of secondary plastid transfers among eukaryotes and those who argue for a much more dynamic history of photosynthesis in eukaryotes. On the one hand is the hypothesis that there have been only three secondary transfers of plastids in eukaryotes into the ancestors of euglenids, chlorarachniophytes, and the supergroup 'Chromalveolata' (single quotation marks denote uncertainty; see "Supergroup hypotheses," below) (Cavalier-Smith 1999, Keeling 2004, Archibald 2006). However, these parsimonious hypotheses are not consistently recovered in analyses of host genomes (Sanchez-Puerta et al. 2007, Yoon et al. 2008). One alternative is articulated in models wherein plastids are particularly likely to be transferred because of their immediate selective benefit to hosts (Grzebyk et al. 2003). As genome-scale data from diverse photosynthetic eukaryotes emerge, hypotheses on the origins of photosynthesis in eukaryotes will be rigorously tested, with emphasis on resolving the chimeric history of the genomes of photosynthetic lineage.

Diversification of eukaryotic lineages

Beyond theories for the origin of eukaryotic cells, there remains the challenge of reconstructing the diversification of eukaryotic life through phylogenetic analyses of extant eukaryotes. Many theories have been proposed in the form of taxonomies (box 1) that seek to describe the organization of eukaryotic diversity (see Parfrey et al. 2006). In the pre-molecular era, numerous proposed taxonomies of eukaryotes were based on morphological features observable through the light microscope and, later, on ultrastructural characters (subcellular features) revealed by the electron microscope (e.g., Patterson 1999, Taylor 1999). Ultrastructural studies led to the recognition of about 70 different lineages of protists with robust patterns of ultrastructural identities or organization (see box 1; Patterson 1999, Taylor 1999). However, evaluation of relationships among these groups has proven difficult,

largely because of the lack of clearly homologous (see box 1) ultrastructural characters.

During the past three decades a wealth of molecular data have been employed in molecular systematics (box 1), greatly advancing our understanding of the diversity of and relationships among eukaryotic lineages (Cavalier-Smith 1983, Sogin et al. 1989, Baldauf et al. 2000, Philippe et al. 2000, Yoon et al. 2008). Systematic studies primarily use sequence data (amino acid and DNA/RNA nucleotides), as they contain numerous characters that can be analyzed once homologous positions are aligned. The amount of sequence data available has grown exponentially as the technology for producing and analyzing sequences has become cheaper. The majority of the microbial lineages are now represented by some molecular data in the genetic databases. Analyses of these data are producing new hypotheses about the evolutionary histories among eukaryotes, and enabling the testing of existing hypotheses.

Supergroup hypotheses

Analyses of emerging molecular phylogenies coupled with ultrastructural studies led to the hypothesis that eukaryotes can be classified into six major supergroups, although subsequent analyses indicate that some of these clades are not robustly supported (e.g., Parfrey et al. 2006, Yoon et al. 2008). The six supergroups, Opisthokonta, 'Amoebozoa', 'Chromalveolata', 'Plantae' (also referred to as 'Archaeplastida'), 'Rhizaria,' and 'Excavata' (Adl et al. 2005), constitute the current classification of eukaryotic diversity that is now emerging in biology textbooks (figure 3). Here, we put the names of many groups in single quotation marks to indicate the uncertainties around these hypothesized eukaryotic relationships, as the support for most of the supergroups is questionable (Parfrey et al. 2006).

We discuss the levels of support the supergroups receive in multigene and phylogenomic analyses, as well as other forms of supporting evidence such as ultrastructural or molecular synapomorphies (see box 1), common organellar origin, and unique molecular features. Groups with ultrastructural identities are by and large supported in molecular studies and provide a standard for robust groups (Patterson 1999, Yoon et al. 2008). Explanations for the instability of the supergroups include the chimeric nature of the genome, LGT, heterogeneous rates of evolution (see box 1), methodological challenges, and limited and biased taxonomic sampling in many studies (see Philippe et al. 2005, Roger and Hug 2006). Others have suggested that the deepest eukaryotic relationships are unknowable because the radiation of eukaryotic lineages proceeded so rapidly that it cannot be resolved after a billion or two years of evolution (Koonin 2007).

Opisthokonta: Supported by multiple sources of data

Opisthokonta, the most robust supergroup, includes animals, fungi, and their microbial relatives (e.g., choanoflagellates, nucleariids) (figure 1t–1w). The animal-fungal clade first emerged in SSU-rRNA gene trees in 1993, and subsequent

studies have provided additional support for the group and added new microbial members (see Parfrey et al. 2006, Steenkamp et al. 2006). Multiple genomic data, including phylogenomic studies (figure 4; Burki et al. 2007, Rodriguez-Ezpeleta et al. 2007), also provide strong support for the Opisthokonta. This diverse clade shares morphological and molecular synapomorphies, such as the presence of single posterior flagellum in those lineages with flagella and a 12-amino acid insertion in the elongation factor 1-alpha gene (Steenkamp et al. 2006) as well as a shared haloarchaeal type tyrosyl-tRNA synthetase that was acquired through LGT (Huang et al. 2005).

Plastid origin supergroup hypotheses: 'Plantae' and 'Chromalveolata'

'Plantae' and 'Chromalveolata', two supergroups with predominantly photosynthetic members, emerged as hypotheses postulating single plastid acquisition through endosymbiosis of a cyanobacterium in 'Plantae' and a red alga in 'Chromalveolata' (see the section above on plastid origins; Adl et al. 2005). Resolving the evolutionary history of these lineages is complicated, as their genomes are a chimera of the original host and plastid genomes. This chimerism is reflected in the incongruence observed between plastid-derived characters (genes of plastid origin, transport pathways into the plastid) that support these groups and host-derived characters (nuclear genes, host ultrastructure), which often do not (Parfrey et al. 2006). Elucidating the evolutionary history of the host is a key test of organelle-based hypotheses, as this will differentiate between a single origin of the organelle (host and plastid phylogenies congruence) and a single plastid source engulfed by multiple hosts (host and plastid phylogenies incongruence) (Delwiche 1999, Grzebyk et al. 2003, Bodyl 2005).

The members of 'Plantae' are the green algae (including land plants), red algae, and glaucocystophytes (figure 1a–1c). Molecular genealogies of plastid-derived genes provide support for the common ancestry of all 'Plantae' plastids, as do other similarities in transport machinery (McFadden and van Dooren 2004). However, as mentioned above, red and green algae have different rubisco protein complexes (Delwiche and Palmer 1996) and light-harvesting compounds (i.e., chlorophyll) (Larkum et al. 2007). Support for the monophyly of 'Plantae' host genomes is generally high in analyses of large data sets (phylogenomics) with limited taxonomic sampling (e.g., Rodriguez-Ezpeleta et al. 2005, 2007, Burki et al. 2007). Conversely, reanalysis of Rodriguez-Ezpeleta and colleagues' (2005) data, including additional taxa of interest and with removal of fast-evolving sites, which may produce spurious relationships in phylogenies, did not support the monophyly

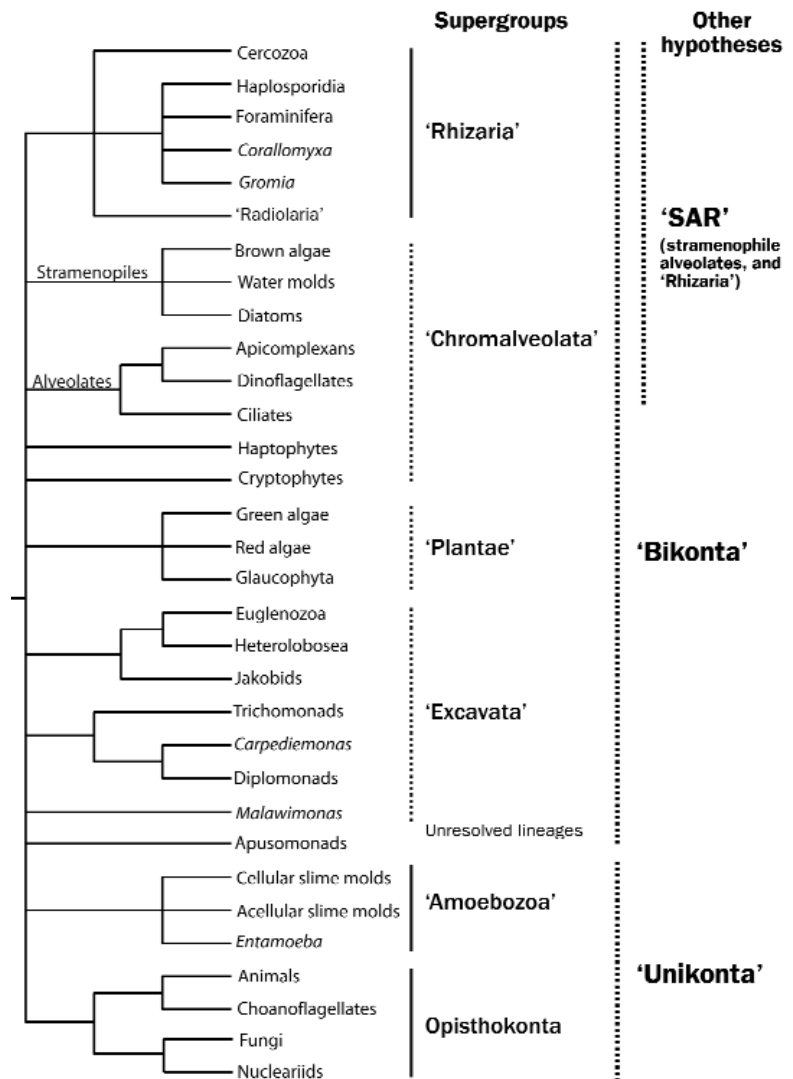


Figure 3. A synthesis of the eukaryotic tree of life that depicts the current classification into supergroups, plus other emerging hypotheses. The unresolved backbone of the tree indicates the uncertainties and lack of support at deeper nodes. Groups that are generally recovered are indicated with solid lines. Dashed lines and single quotation marks are used to denote groups that receive inconsistent support and hypothesized groups whose support is still being evaluated.

of 'Plantae' (Nozaki et al. 2007). Further, our taxon-rich analyses of the four most-sampled markers (SSU, actin, alpha-tubulin, and beta-tubulin) similarly failed to resolve this clade (figure 4; Yoon et al. 2008). These results suggest incongruence due to the conflicting signals of the host and the plastid genomes.

The supergroup 'Chromalveolata' is a highly contentious group that unites the alveolates and the 'Chromista' (Adl et al. 2005). The alveolates are well established by both molecular and ultrastructural evidence and include dinoflagellates, apicomplexans, and ciliates (reviewed in Patterson 1999). The 'Chromista' are a polyphyletic collection of eukaryotes that encompass much of the marine

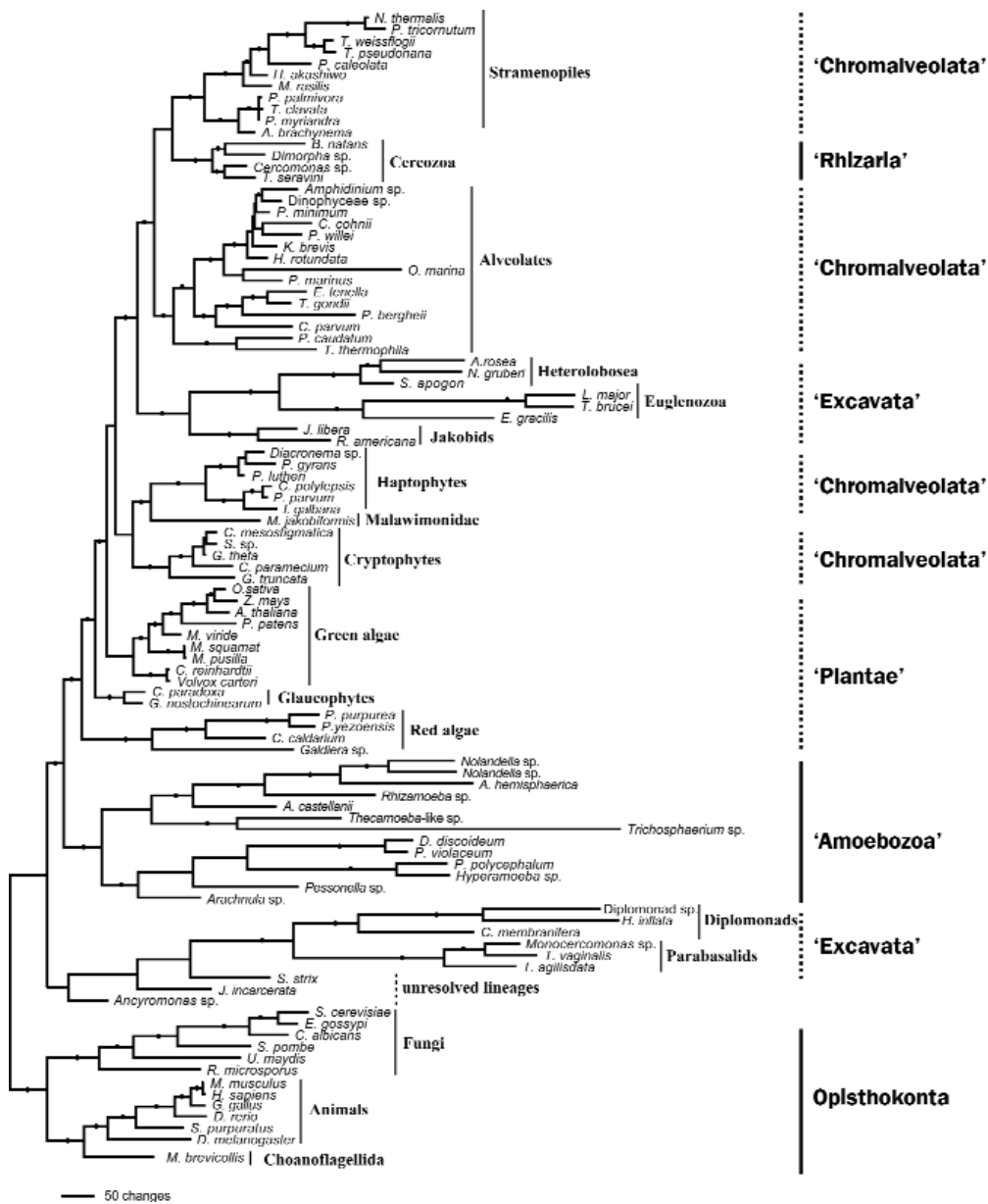


Figure 4. An example of a multigene tree of the major eukaryotic groups inferred from analyses of SSU-rDNA (small subunit ribosomal RNA) and amino acid sequences of actin, alpha-tubulin, and beta-tubulin in MrBayes. Solid black dots indicate Bayesian posterior support values greater than 95%. All branches are drawn to scale. Source: Tree modified from Tekle and colleagues (2008) and used with permission from Elsevier, Inc.

eukaryotic phytoplankton diversity (kelps, diatoms, dinoflagellates, coccolithophorids) (e.g., Cavalier-Smith 2002). ‘Chromalveolata’ remains controversial, as it has consistently failed to form a monophyletic group in nuclear gene analyses (Parfrey et al. 2006), even those with large numbers of genes (Burki et al. 2007, Hackett et al. 2007) and increased sampling of relevant taxa (Yoon et al. 2008). The support for this supergroup is largely limited to those studies that use genetic markers of plastid origin (e.g., McFadden and van

Dooren 2004). However, these plastid genealogies are also consistent with other hypotheses, and a shift is under way from ‘Chromalveolata’ toward alternative hypotheses (Grzebyk et al. 2003, Bodyl 2005, Burki et al. 2007, Sanchez-Puerta et al. 2007).

Amoebae-rich supergroups

One of the impacts of molecular data is the emergence of unsuspected relationships between morphologically dissimilar organisms, as in the case of the heterogeneous supergroups ‘Amoebozoa’ and ‘Rhizaria’ (see Adl et al. 2005). These groups lack defining morphological synapomorphies, but are generally supported by molecular analyses (see Parfrey et al. 2006). The membership of these diverse groups continues to expand as more and more unaffiliated microbial eukaryotes are sampled (e.g., Polet et al. 2004, Tekle et al. 2007, 2008, Yoon et al. 2008). As there is some circularity in continuing to test these groups with the markers that were originally used to define them, robust tests must await additional data, such as phylogenomic studies with broad taxon sampling.

‘Rhizaria’ is a collection of amoebae, parasites, and flagellates, including foraminifera (testate marine amoebae with reticulating pseudopods), chlorarachniophytes (filose amoebae with green algal

plastids), and plasmodiophorids (plant parasites) (figure 1j–1m). ‘Rhizaria’ is frequently supported in molecular analyses of single (or few) genes with decent taxonomic sampling (Nikolaev et al. 2004) and, recently, in phylogenomic analyses with few taxa (Burki et al. 2007, Rodriguez-Ezpeleta et al. 2007). Statistical support for this supergroup is inconsistent in multigene genealogies with larger taxon sampling (Yoon et al. 2008). However, some members of this supergroup (e.g., foraminifera) are characterized by highly divergent

sequence evolution for some markers used, and this might account for its failure to form a group (Habura et al. 2005).

'Amoebozoa' is composed of mostly amoeboid forms, including naked, free-living lobose amoebae (e.g., *Amoeba proteus*), organisms that cause disease in humans (such as *Entamoeba histolytica*, amoebic dysentery), cellular slime molds (*Dictyostelium*), and lobose amoebae that secrete shells (figure 1q–1s). This supergroup is generally supported by molecular analyses with more genetic data (e.g., Burki et al. 2007, Rodriguez-Ezpeleta et al. 2007) and taxonomic sampling (Tekle et al. 2008).

'Excavata': Defined by ultrastructural synapomorphy

The supergroup 'Excavata,' hypothesized on the basis of a homologous ventral feeding groove and associated ultrastructural characters (Simpson and Patterson 1999), has not proven robust in most molecular analyses. This group includes the familiar parasites *Trypanosoma*, *Giardia*, and *Trichomonas*, as well as photosynthetic free-living lineages (e.g., *Euglena gracilis*) (figures 1n–1p). Molecular analyses generally reject the monophyly of 'Excavata' even in studies with good sampling of genes and taxa (Simpson et al. 2006, Rodriguez-Ezpeleta et al. 2007, Yoon et al. 2008). 'Excavata' members generally fall into two separate, well-supported clades (figures 3, 4). The nonmonophyly of 'Excavata' suggests that the excavate groove structure has been lost or modified in intervening lineages, or that it evolved more than once by convergent evolution.

Interrelationships between supergroups

The supergroups are themselves in flux as groups shift with additional sampling of taxa and data. For example, several recent papers have reported a relationship between 'Rhizaria' and members of the 'Chromalveolata,' leading to a hypothesis on the specific relationship between stramenopiles, alveolates, and 'Rhizaria'—the "SAR" hypothesis (figure 3; Burki et al. 2007, Hackett et al. 2007, Rodriguez-Ezpeleta et al. 2007). The emergence of the SAR highlights the complexity of the history of secondary endosymbiosis of green and red algal plastids, which are both found in these lineages. There are no morphological or molecular synapomorphies that support this group, illustrating both the power of molecules to point to unsuspected relationships and the difficulty of resolving the deepest relationships.

At a higher level, the eukaryotic supergroups have been classified as either 'Unikonta' or 'Bikonta' and the root has been drawn between them (Stechmann and Cavalier-Smith 2002). Members of the 'Bikonta' ('Chromalveolata,' 'Excavata,' 'Plantae,' 'Rhizaria') were originally postulated to encompass organisms whose flagella are supported by two flagellar roots, whereas the 'Unikonta' have flagella bearing a single root ('Amoebozoa' and Opisthokonta). These concepts were based on only a few representative lineages, and there are putative 'bikonts' that have one flagellar root, and 'unikonts' with two (Cavalier-Smith 2002). There is some evidence from genomic data that has also been cited in support of 'Unikonta' and 'Bikonta,' including a gene fusion said to be unikont specific

(Stechmann and Cavalier-Smith 2002). However, the gene fusion was later discovered in the genome of a red alga, a putative bikont (Nozaki et al. 2005). The bifurcation of eukaryotes into 'unikonts' and 'bikonts' requires further scrutiny based on analyses of larger taxonomic coverage and unequivocal morphological and molecular synapomorphies.

Molecular data increase resolution of robust eukaryotic clades

The availability of gene and genome sequences has greatly increased our understanding of the phylogenetic relationships among eukaryotes by (a) suggesting relationships between morphologically dissimilar organisms; (b) providing a means to test the robustness of groups, such as those with ultrastructural identities; and (c) providing homologous characters that enable comparison across the diversity of eukaryotes. Below the supergroup level, molecular data have bolstered the support for numerous clades. Molecular data also pointed out some surprising alliances of organisms, such as the placement of microsporidia—a lineage of parasites previously classified among the protozoa—within the fungi (Hirt et al. 1999). Molecular studies have also provided confirmation of relationships suspected on the basis of other forms of evidence such as the stramenopiles, a clade containing diatoms, kelps, and golden algae, as well as many nonphotosynthetic members, including water molds (e.g., *Phytophthora infestans*, the agent of the Irish potato blight). These diverse lineages share the ultrastructural synapomorphy of tripartite tubular hairs on one flagellum (Patterson 1999). Both ultrastructure and molecular studies were instrumental in the emergence of the Alveolata (Taylor 1999). A relationship between ciliates and dinoflagellates was first suggested because of ultrastructural similarities, while molecular analyses added the Apicomplexa to this group. Subsequent work uncovered the ultrastructural synapomorphy of alveolar sacs located directly under the membrane (Patterson 1999).

Conclusions

With the accumulation of gene and genomic data in the past decade, we have seen tremendous progress in our understanding of the origin of eukaryotes. In this era it has become possible to assess different theories of the origin of eukaryotes with empirical data. Examination of genomic data has revealed that the eukaryotic genome is chimeric, as eukaryotic genes share ancestry with different archaeal and bacterial lineages (figure 2). Several lines of molecular evidence indicate an early acquisition of mitochondrion during the evolution of the eukaryotic cell. Genomic data are also playing a role in unraveling the mysteries surrounding the origin of the key innovations of the eukaryotic cell.

Our understanding of the diversification of eukaryotes has also benefited from analyses of molecular data. In particular, molecular data provide a myriad of homologous characters that can be used to reconstruct phylogenies of diverse microbial and macroscopic lineages whose morphological differences preclude comparison by other methods.

Despite varying support, the current classification of eukaryotes into six putative supergroups is a step toward deciphering eukaryotic diversification and evolution. Complex patterns of molecular evolution, methodological problems, and inadequate and biased taxonomic and data sampling have hindered inference of the deep interrelationship of eukaryotes. However, these challenges will diminish through inclusion of more lineages, incorporation of appropriate models of evolution, and refinement of the methodological tools.

Acknowledgments

This work was supported by grants from the National Science Foundation (DEB 043115 and OCE 9648713) and the National Institutes of Health (AREA 1R15GM081865-01) to L. A. K.

References cited

- Adl SM, et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology* 52: 399–451.
- Archibald JM. 2006. Endosymbiosis: Double-take on plastid origins. *Current Biology* 16: R690–R692.
- 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.
- Bhattacharya D, Yoon HS, Hackett JD. 2004. Photosynthetic eukaryotes unite: Endosymbiosis connects the dots. *Bioessays* 26: 50–60.
- Bodl A. 2005. Do plastid-related characters support the chromalveolate hypothesis? *Journal of Phycology* 41: 712–719.
- Brown JR, Doolittle WF. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiology and Molecular Biology Reviews* 61: 456–502.
- Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland Å, Nikolaev SI, Jakobsen KS, Pawlowski J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* 2: e790.
- Cavalier-Smith T. 1978. The evolutionary origin and phylogeny of microtubules, mitotic spindles and eukaryote flagella. *Biosystems* 10: 93–114.
- . 1983. A 6-kingdom classification and a unified phylogeny. Pages 1027–1034 in Schenk HEA, Schwemmler W, eds. *Endocytobiology II: Intracellular Space as Oligogenetic Ecosystem*. Walter de Gruyter.
- . 1999. Zooflagellate phylogeny and the systematics of Protozoa. *Biological Bulletin* 196: 393–395.
- . 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa. *International Journal of Systematic and Evolutionary Microbiology* 52: 297–354.
- Dagan T, Martin W. 2006. The tree of one percent. *Genome Biology* 7: 118.
- Delwiche CF. 1999. Tracing the tread of plastid diversity through the tapestry of life. *American Naturalist* 154: S164–S177.
- Delwiche CF, Palmer JD. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Molecular Biology and Evolution* 13: 873–882.
- Doolittle RF. 1995. The origins and evolution of eukaryotic proteins. *Philosophical Transactions of the Royal Society B* 349: 235–240.
- Doolittle WF. 1998. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in Genetics* 14: 307–311.
- Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature* 440: 623–630.
- Embley TM, van der Giezen M, Horner DS, Dyal PL, Bell S, Foster PG. 2003. Hydrogenosomes, mitochondria and early eukaryotic evolution. *Journal of the International Union of Biochemistry and Molecular Biology* 55: 387–395.
- Erickson HP. 2007. Evolution of the cytoskeleton. *Bioessays* 29: 668–677.
- Gray MW. 1999. Evolution of organellar genomes. *Current Opinion in Genetics and Development* 9: 678–687.
- Grzebyk D, Schofield O, Vetriani C, Falkowski PG. 2003. The mesozoic radiation of eukaryotic algae: The portable plastid hypothesis. *Journal of Phycology* 39: 259–267.
- Habura A, Wegener L, Travis JL, Bowser SS. 2005. Structural and functional implications of an unusual foraminiferal β -tubulin. *Molecular Biology and Evolution* 22: 2000–2009.
- Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D. 2007. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of Rhizaria with chromalveolates. *Molecular Biology and Evolution* 8: 1702–1713.
- Hartman H, Fedorov A. 2002. The origin of the eukaryotic cell: A genomic investigation. *Proceedings of the National Academy of Sciences* 99: 1420–1425.
- Hirt RP, Logsdon JM, 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. *Proceedings of the National Academy of Sciences* 96: 580–585.
- Huang J, Xu Y, Gogarten JP. 2005. The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic. *Molecular Biology and Evolution* 22: 2142–2146.
- Jekely G. 2007. Origin of eukaryotic endomembranes: A critical evaluation of different model scenarios. *Advances in Experimental Medicine and Biology* 607: 38–51.
- Jenkins C, Samudrala R, Anderson I, Hedlund BP, Petroni G, Michailova N, Pinel N, Overbeek R, Rosati G, Staley JT. 2002. Genes for the cytoskeletal protein tubulin in the bacterial genus *Prostheco bacter*. *Proceedings of the National Academy of Sciences* 99: 17049–17054.
- Katz LA. 1998. Changing perspectives on the origin of eukaryotes. *Trends in Ecology and Evolution* 13: 493–497.
- Keeling PJ. 2004. Diversity and evolutionary history of plastids and their hosts. *American Journal of Botany* 91: 1481–1493.
- Knoll AH, Javaux EJ, Hewitt D, Cohen P. 2006. Eukaryotic organisms in Proterozoic oceans. *Philosophical Transactions of the Royal Society B* 361: 1023–1038.
- Koonin EV. 2007. The Biological Big Bang model for the major transitions in evolution. *Biology Direct* 2: 21.
- Larkum AW, Lockhart PJ, Howe CJ. 2007. Shopping for plastids. *Trends in Plant Science* 12: 189–195.
- Mans BJ, Anantharaman V, Aravind L, Koonin EV. 2004. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* 3: 1612–1637.
- Margulis L. 1993. Serial endosymbiosis theory. Pages 1–18 in Margulis L. *Symbiosis in Cell Evolution: Microbial Communities in the Archean and Proterozoic Eons*. W. H. Freeman.
- Margulis L, Dolan ME, Guerrero R. 2000. The chimeric eukaryote: Origin of the nucleus from the karyomastigont in amitochondriate protists. *Proceedings of the National Academy of Sciences* 97: 6954–6959.
- Margulis L, Chapman M, Guerrero R, Hall J. 2006. The last eukaryotic common ancestor (LECA): Acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proceedings of the National Academy of Sciences* 103: 13080–13085.
- Martin W, Koonin EV. 2006. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440: 41–45.
- Martin W, Müller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature* 392: 37–41.
- Martin W, Schnarrenberger C. 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: A case study of functional redundancy in ancient pathways through endosymbiosis. *Current Genetics* 32: 1–18.
- McFadden GI, van Dooren GG. 2004. Evolution: Red algal genome affirms a common origin of all plastids. *Current Biology* 14: R514–R516.
- Mitchell B, Jacobs R, Li J, Chien S, Kintner C. 2007. A positive feedback mechanism governs the polarity and motion of motile cilia. *Nature* 447: 97–101.
- Nikolaev SI, Berney C, Fahrni JE, 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. *Proceedings of the National Academy of Sciences* 101: 8066–8071.
- Nozaki H, Matsuzaki M, Misumi O, Kuroiwa H, Higashiyama T, Kuroiwa T. 2005. Phylogenetic implications of the CAD complex from the primitive red alga *Cyanidioschyzon merolae* (Cyanidiales, Rhodophyta). *Journal of Phycology* 41: 652–657.
- Nozaki H, Iseki M, Hasegawa M, Misawa K, Nakada T, Sasaki N, Watanabe M. 2007. Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Molecular Biology and Evolution* 24: 1592–1595.
- Parfrey LW, Barbero E, Lasser E, Dunthorn M, Bhattacharya D, Patterson DJ, Katz LA. 2006. Evaluating support for the current classification of eukaryotic diversity. *PLoS Genetics* 2: 2062–2073.
- Patterson DJ. 1999. The diversity of eukaryotes. *American Naturalist* 154: S96–S124.
- Philippe H, Germot A, Moreira D. 2000. The new phylogeny of eukaryotes. *Current Opinion in Genetics and Development* 10: 596–601.
- Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. 2005. Heterotachy and long-branch attraction in phylogenetics. *BMC Evolutionary Biology* 5: 50.
- Pisani D, Cotton JA, McInerney JO. 2007. Supertrees disentangle the chimerical origin of eukaryotic genomes. *Molecular Biology and Evolution* 24: 1752–1760.
- Poet S, Berney C, Fahrni J, Pawlowski J. 2004. Small-subunit ribosomal RNA gene sequences of Phaeodarea challenge the monophyly of Haeckel's Radiolaria. *Protist* 155: 53–63.
- Richards TA, Cavalier-Smith T. 2005. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436: 1113–1118.
- Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roue B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Current Biology* 15: 1325–1330.
- Rodriguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, Lang BF. 2007. Toward resolving the eukaryotic tree: The phylogenetic positions of jakobids and cercozoans. *Current Biology* 17: 1420–1425.
- Roger AJ. 1999. Reconstructing early events in eukaryotic evolution. *American Naturalist* 154: S146–S163.
- Roger AJ, Hug LA. 2006. The origin and diversification of eukaryotes: Problems with molecular phylogenetics and molecular clock estimation. *Philosophical Transactions of the Royal Society B* 361: 1039–1054.
- Sagan L. 1967. On the origin of mitosing cells. *Journal of Theoretical Biology* 14: 255–274.
- Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF. 2007. Sorting wheat from chaff in multi-gene analyses of chlorophyll *c*-containing plastids. *Molecular Phylogenetics and Evolution* 44: 885–897.
- Searcy DG. 1992. Origins of mitochondria and chloroplasts from sulfur-based symbioses. Pages 47–78 in Hartman H, Matsuno K, eds. *The Origin and Evolution of the Cell*. World Scientific.
- Simpson AGB, Patterson DJ. 1999. The ultrastructure of *Carpodimonas membranifera* (Eukaryota) with reference to the “Excavate hypothesis.” *European Journal of Protistology* 35: 353–370.
- Simpson AGB, Inagaki Y, Roger AJ. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of “primitive” eukaryotes. *Molecular Biology and Evolution* 23: 615–625.
- Sogin ML, Gunderon JH, Elwood HJ, Alonso RA, Peattie DA. 1989. Phylogenetic meaning of the kingdom concept: An unusual ribosomal RNA from *Giardia lamblia*. *Science* 243: 75–77.
- 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. *Molecular Biology and Evolution* 23: 93–106.
- Steiner JM, Yusa F, Pompe JA, Löffelhardt W. 2005. Homologous protein import machineries in chloroplasts and cyanelles. *Plant Journal* 44: 646–652.
- Taylor FJR. 1987. An overview of the status of evolutionary cell symbiosis theories. *Annals of the New York Academy of Sciences* 503: 1–16.
- . 1999. Ultrastructure as a control for protistan molecular phylogeny. *American Naturalist* 154: S125–S136.
- Tekle YI, Grant J, Cole JC, Nerad TA, Anderson OR, Patterson DJ, Katz LA. 2007. A multigene analysis of *Corallomyxa tenera* sp. nov. suggests its membership in a clade that includes *Gromia*, Haplosporidia, and Foraminifera. *Protist* 158: 457–472.
- Tekle YI, Grant J, Cole JC, Nerad TA, Patterson DJ, Anderson OR, Katz LA. 2008. Phylogenetic placement of diverse amoebae inferred from multigene analysis and assessment of the stability of clades within ‘Amoebozoa’ upon removal of varying fast rate classes of SSU-rDNA. *Molecular Phylogenetics and Evolution* 47: 339–352.
- Yoon HS, Grant J, Tekle YI, Wu M, Chaon BC, Cole JC, Logsdon JM, Patterson DJ, Bhattacharya D, Katz LA. 2008. Broadly sampled multigene trees of eukaryotes. *BMC Evolutionary Biology* 8: 14.
- Zillig W, Klenk HP, Palm P, Leffers H, Puhler G, Gropp F, Garrett RA. 1989. Did eukaryotes originate by a fusion event? *Endocytobiosis and Cell Research* 6: 1–25.

Yonas I. Tekle and Laura A. Katz (lkatz@email.smith.edu) are with the Department of Biological Sciences at Smith College in Northampton, Massachusetts. Laura Wegener Parfrey is with the Organismic and Evolutionary Biology program at the University of Massachusetts in Amherst.