The Eukaryotic Tree of Life from a Global Phylogenomic Perspective

Fabien Burki

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Correspondence: burkif@mail.ubc.ca

Molecular phylogenetics has revolutionized our knowledge of the eukaryotic tree of life. With the advent of genomics, a new discipline of phylogenetics has emerged: phylogenomics. This method uses large alignments of tens to hundreds of genes to reconstruct evolutionary histories. This approach has led to the resolution of ancient and contentious relationships, notably between the building blocks of the tree (the supergroups), and allowed to place in the tree enigmatic yet important protist lineages for understanding eukaryote evolution. Here, I discuss the pros and cons of phylogenomics and review the eukaryotic supergroups in light of earlier work that laid the foundation for the current view of the tree, including the position of the root. I conclude by presenting a picture of eukaryote evolution, summarizing the most recent progress in assembling the global tree.

It is redundant to say that eukaryotes are diverse. Plants, animals, and fungi are the charismatic representatives of the eukaryotic domain of life, but this narrow view does not do justice to the eukaryotic diversity. Microscopic eukaryotes, often unicellular and known as the protists, represent the bulk of most major groups, whereas multicellular lineages are confined to small corners on the global tree of eukaryotes. If all eukaryotes possess structures enclosed within intracellular membranes (the organelles), an infinite variation of forms and feeding strategies has evolved since their origin. Eukaryotic cells can wander on their own, sometimes forming hordes of free-living pico-sized organisms that flourish in oceans. They can be parasites or symbionts, or come together by the billions in tightly packed, highly regulated multicellular organisms. Eukaryotes have occupied just about every ecological niche on Earth. Some actively gather food from the environment, others use plastids (chloroplasts) to derive energy from the light; many can adapt to variable conditions by switching between autotrophy and the predatory consumption of prey by phagotrophy. Eukaryotes also show a great deal of genomic variation (Lynch and Conery 2003). Some amoebozoan protists, for instance, have the largest known genomes—more than 200 times larger than that of humans (Keeling and Slamovits 2005). Conversely, microbial parasites can have highly compact, bacterial-size genomes (Corradi et al. 2010). Even smaller are the remnant nuclear genomes (nucleomorphs) of what were once free-living microbial algae. At around 500,000 nucleotides and
hardly encoding a few hundreds genes, nucleo-
morphs are the smallest nuclear genome of all (Douglas et al. 2001; Gilson et al. 2006; Lane et al. 2007).

Recognizing this great diversity and pushed by a desire to establish order, biologists have long attempted to assemble a global eukaryotic tree of life. A fully resolved phylogenetic tree including all organisms is not only the ultimate goal of systematics, it would also provide the foundation to infer the acquisition and evolution of countless characters through the history of long-dead species. But early attempts to resolve the eukaryotic tree, most of which were based on comparisons of morphology and nutrition modes, faced the impossible challenge of describing in an evolutionary sensitive way a world in which most of the diversity occurs among tiny microbes. For decades, biology textbooks assigned the eukaryotes to evolutionary entities called “kingdoms” in which the lords were the animals, plants, and fungi (Copeland 1938; Whittaker 1969; Margulis 1971). This is not to say that biologist ignored protists, and they have been in fact recognized as a kingdom for more that a century (Haeckel 1866), but protists were considered to be “simple” organisms from which more elaborate, multicellular species emerged. Although these early proposals succeeded in recognizing several major assemblages, such as animals and plants, they were less successful in resolving the relationships between the groups and, with the benefit of hindsight, failed to account for the fundamental paraphyletic and complex nature of the protist lines.

A MOLECULAR (R)EVOLUTION

The backbone of the eukaryotic tree has gone through some profound rearrangements in the past 20 years. Comparing nucleotide or amino acid sequences is now the tool of choice for reconstructing evolutionary histories. This is particularly true for protists because the interpretation of their morphological characters alone is problematic. For years, the go-to molecular marker for phylogenetics has been the small subunit ribosomal RNA (SSU rRNA). It is easy to amplify and contains both hypervariable and conserved regions, allowing researchers to investigate different depths of phylogenetic resolution. As a result, the SSU rRNA dominates molecular databases, and the majority of the known eukaryotic diversity, when characterized molecularly, is still defined solely by this marker.

The pioneering molecular phylogenies consistently recovered a handful of deeply diverging protist lineages (e.g., diplomonads, parabasalids, microsporidians, Archamoebae), progressively emerging from the distant prokaryotic root, and followed by a densely branched “crown,” nesting the more familiar eukaryotic diversity (Sogin et al. 1986, 1989; Friedman et al. 1987; Woese et al. 1990; Sogin 1991). This was an appealing picture of evolution because these early diverging species were seemingly morphologically simple single-cell organisms that lacked mitochondria and other typical eukaryotic structures, such as peroxisomes (Keeling 1998). These phylogenies were also consistent with the archezoa hypothesis, which postulated that amitochondriate eukaryote lineages diverged before the endosymbiotic event that gave rise to mitochondria (Cavalier-Smith 1983, 1987, 1989). Even more convincing was that other molecular markers, including various elongations factors and RNA polymerase subunits, corroborated the deep-branched position of archezoaan taxa, altogether supporting the prediction that they should branch earlier than the mitochondrion-containing eukaryotes if they predate the origin of this organelle (Brown and Doolittle 1995; Klenk et al. 1995; Kamaishi et al. 1996; Yamamoto et al. 1997).

At the other end of the tree, the so-called crown, contain the major clades of eukaryotes, appearing tightly bunched together as if they diverged almost simultaneously (Sogin 1991; Knoll 1992). These clades included animals, fungi, and plants as well as diverse protist lineages such as alveolates and stramenopiles (see below). The branching pattern among the SSU rRNA crown taxa, however, could not be resolved even with the help of several protein markers (Baldauf 1999; Hirt et al. 1999; Roger
et al. 1999; Moreira et al. 2000). This lack of molecular resolution was interpreted as evidence that not enough phylogenetic signal could accumulate in the sequences because most phyla emerged in a very short period of time, like in a “big-bang” explosion of species diversification (Philippe et al. 2000a,b).

TRIMMING THE TREE

The early molecular-based interpretation of the eukaryotic tree showing the archeozoan-crown dichotomy did not last very long. First, as more and more lineages were being sequenced, mitochondriate protist groups such as Euglenozoa and Foraminifera squeezed in between the archezoan taxa and the crown (Sogin et al. 1986; Clark and Cross 1988; Pawlowski et al. 1996). Moreover, the archezoans were characterized by elevated rates of molecular evolution, which translates into long branches in phylogenetic trees (Philippe et al. 2000b). This meant that for these “basal” taxa, most standard molecular markers such as the SSU rRNA were mutationally saturated, a feature that can lead to the notorious long branch attraction (LBA) artefact in which distantly related species with fast evolving sequences are erroneously clustered together (Felsenstein 1981). Ultimately, new genes, more taxa, and better phylogenetic methods showed that if the archezoan taxa appeared to diverge early, it was not because they were “primitive” eukaryotes but rather because of their artificial attraction to the base of the tree by distant outgroups (Embley and Hirt 1998; Roger 1999; Baldauf et al. 2000; Philippe et al. 2000a,b). At the same time, mitochondrial-derived genes were progressively discovered in archezoan genomes, the products of which were shown to be targeted to reduced double-membrane-bounded organelles of mitochondrial ancestry (hydrogenosomes and mitosomes), suggesting that all “amitochondriate” eukaryotes once possessed mitochondria (Clark and Roger 1995; Bui et al. 1996; Keeling 1998; Tovar et al. 1999, 2003; Williams et al. 2002; Embley and Martin 2006; Goldberg et al. 2008; Hjort et al. 2010). This marked the end of both the archezoa hypothesis and the concept of a eukaryotic crown, and no longer supported the idea that the prokaryote-to-eukaryote transition was a progressive transformation involving intermediate amitochondriate forms.

More generally, the whole eukaryotic tree was shaken up by important discrepancies between SSU rRNA-based phylogenies and those inferred from a growing number of protein-coding genes, as well as discrete molecular characters such as shared indels (insertions/deletions) or gene fusions, or the systematic analysis of light and electron microscopy data (e.g., Baldauf and Palmer 1993; Keeling and Doolittle 1996; Fast et al. 1999; Baldauf et al. 2000; Moreira et al. 2000; Cavalier-Smith 2000; Nipkow et al. 2004; Harper et al. 2005). The integration of these various kinds of data led to the conception that most, if not all, eukaryotic diversity can be assigned to one of several major assemblages, called “supergroups” (Baldauf 2003; Keeling 2004; Simpson and Roger 2004; Adl et al. 2005; Keeling et al. 2005; Parfrey et al. 2006). In this framework, animals occupy just one branch among hundreds, and are far outnumbered at the level of major lineages by unicellular eukaryotes. In its original form, this new tree of eukaryotes was an unrooted polytomy with six main stems, each representing a supergroup: Opisthokonta, Amoebozoa, Excavata, Archaeplastida, Rhizaria and Chromalveolata. All six of these supergroups emerged from a common point, and their order of divergence was mostly unknown.

PHYLOGENETICS + GENOMICS = PHYLOGENOMICS

Fast forward to present day, we are well into the genomic era and reconstructing the tree of eukaryotes is no longer the job of a few genes. Instead, huge phylogenomic data sets containing hundreds of genes for an always-increasing number of taxa can now be used. The tedious and expensive Sanger sequencing of the early 2000s (the human genome cost in the range of a billion U.S. dollars) has been replaced by fast and cheap next-generation sequencing, which can produce genome-scale data at unprecedented depths for a taxonomically broad sampling.
Once strongly biased toward organisms relevant for human well-being (of economical or medical importance), there has been an explosion of the taxonomic distribution of species for which extensive genomic data are available. As of 2012, at least one species from each major lineage of eukaryotes has had its genome fully sequenced, not to mention the numerous smaller scale genomic surveys and transcriptomic data sets that have been generated. With this wealth of sequence data at hand, phylogenomics started off as a way to predict gene functions by evolutionary analysis in which uncharacterized genes were predicted by their phylogenetic position relative to genes with known functions (Eisen 1998). Although this area of phylogenomics is still very active, it also rapidly emerged as a new domain of phylogenetics and became an essential tool for addressing controversial evolutionary questions, such as the transitions that led to the current diversity of eukaryotes.

In its most popular application to phylogeny, phylogenomics relies on multiple-sequence alignments, much like the single-gene approach, but here the genes are often concatenated into large supermatrices (Delsuc et al. 2005). With this approach, too, it is important to ensure the vertical transmission of the characters (orthology)—that the genes in related species are inherited from a common ancestor. This is not an easy task because the genes available from genomic-scale projects are still poorly sampled, substantially more so than for widely used phylogenetic protein markers such as actin, α- and β-tubulins, or translation elongation factor 2. Consequently, differentiating between orthology and, for example, the undesirable horizontally transferred genes (HGTs), independent gene losses or partially sampled genes can be difficult to achieve with a limited taxon selection. Moreover, many of the general biases that apply to single-gene phylogenetics are exacerbated in a phylogenomic context (Jeffroy et al. 2006; Rodriguez-Ezepeleta et al. 2007b; Philippe et al. 2011). Phylogenetic reconstruction involves two opposing forces: the true phylogenetic signal carrying the evolutionary history, and the nonphylogenetic signals (noise) resulting from a combination of one or more causes such as saturated positions in the data set (stochastic errors) or model misspecifications (systematic errors). Stochastic errors arise when the number of positions in an alignment is small meaning that the random background noise, which inevitably accumulates through time because of homoplasy, will have a neutralizing effect on the positions that contain the genuine phylogenetic signal. Model misspecifications leading to systematic errors include, for example, the heterogeneity of nucleotide or amino acid composition, which tends to incorrectly cluster together species sharing the same composition, or the heterogeneity of the evolutionary rates, which can result in the LBA artefact (Felsenstein 1981; Philippe 2000; Lopez et al. 2002; Foster 2004; Ho and Jermiin 2004; Jermiin et al. 2004).

Logically, the goal of a phylogenetic inference is to minimize the noise and maximize the true phylogenetic signal (Philippe et al. 2005, 2011). Single-gene phylogenies, because of the limited information they contain, are especially susceptible to stochastic errors; to counter this, the obvious solution is to gather more data in the hope that enough phylogenetic signal is recovered (i.e., synapomorphy will dominate homoplasy). Systematic errors, on the other hand, tend not to vanish with the addition of more data as their causes do not average out over longer alignments (Rodriguez-Ezepeleta et al. 2007b). Yet, the key advantage of phylogenomics is precisely that more data is available to start with, making it possible to apply strategies to diminish the known sources of systematic errors while still maintaining most of the phylogenetic signal (see Delsuc et al. 2005; Philippe et al. 2005, 2011 for reviews). Thus, when combined with other options developed and tested on smaller data sets to increase the phylogenetic accuracy, such as the use of more accurate phylogenetic methods or better taxon samplings, phylogenomics becomes a very powerful tool.

Phylogenomics has confirmed the existence of most supergroups, although various degrees of controversy remain for each of them. The increased power in resolution has also led to some important shuffling among the super-
groups and allowed the placement in the tree of several “orphan” lineages. Below, in the next section, I briefly introduce each supergroup, and discuss the eukaryote phylogeny in light of the most recent advances (see Fig. 1).

THE EUKARYOTIC SUPERGROUPS

Opisthokonta
This supergroup contains animals (Metazoa) and fungi, as well as several lines of heterotrophic protists. Animals emerged from within a paraphyletic assemblage of protists, including the choanoflagellates (e.g., Monosiga), Filasteria (e.g., Capsaspora), and Ichthyospora (e.g., Sphaeroforma; Steenkamp et al. 2006; Ruiz-Trillo et al. 2008; del Campo and Ruiz-Trillo 2013). Fungi, including the monophyletic and early assemblage comprising Cryptomycota (e.g., Rozella), aphelids, and microsporidians (Lara et al. 2010; Jones et al. 2011a,b; James et al. 2013; Karpov et al. 2013) are closely related to nucleariid amoeba (Nuclearia simplex) and the aggregative slime mold Fonticula (Brown et al. 2009; Liu et al. 2009, 2012). Opisthokonts are putatively united by the presence of a single posterior flagellum in several representatives (Cavalier-Smith and Chao 1995), and many molecular-based evidences (single-gene phylogenies, phylogenomics, and indels) have consistently supported the existence of this group (e.g., Baldauf and Palmer 1993; Wainright et al. 1993; Baldauf et al. 2000; Brown et al. 2009; Liu et al. 2009). It is one of the most reliable supergroups.

Amoebozoa
This supergroup includes mostly amoeboid, heterotrophic protists such as the classical naked and testate lobose amoebae with broad pseudopods (e.g., Amoeba), but also contains some amitochondriate parasitic lineages of critical medical importance (e.g., Entamoeba), flagellated cells (e.g., Phalansterium, Multicilia), or the mycetozoan slime molds capable of aggregative multicellularity (e.g., Dictyostelium; see Pawlowski and Burki 2009 for a recent review). In addition to morphological similarities of the amoeboid cells, evidence that they form a monophyletic group is based on single-gene phylogenies (e.g., Fahrni et al. 2003; Smirnov et al. 2005), phylogenomics (Bapteste et al. 2002; Minge et al. 2009; Brown et al. 2012), and similarities in mitochondrial genome architecture between Acanthamoeba and Dictyostelium (Lonergan and Gray 1996; Iwamoto et al. 1998). Opisthokonts and Amoebozoa are often united in a larger supergroup called the unikonts (Cavalier-Smith 2002) or, more recently, Amorphea (Adl et al. 2012), which is supported by a gene duplication/fusion (Stechmann and Cavalier-Smith 2002, 2003) as well as single-gene phylogenies (e.g., Baldauf and Palmer 1993) and phylogenomics (Rodriguez-Ezpeleta et al. 2007a; Hampl et al. 2009; Brown et al. 2012; Derelle and Lang 2012). Amorphea also contains protist lineages of unclear placement, such as the apusomonads, ancryomonads, and breviate amoebae. Based on phylogenomic analysis, opisthokonts, apusomonads, and breviates were recently grouped into a larger entity named Obazoa (Brown et al. 2013). The validity of this assemblage, however, relies on the position of the eukaryotic root, which remains hypothetical (see below).

Excavata
This supergroup is composed of diverse and mainly heterotrophic protists, many of which are anaerobes and/or parasites and possess hydrogenosomes or mitosomes instead of mitochondria (e.g., Giardia, Trichomonas). A group including lineages with plastids of green algal origin, the euglenids (e.g., Euglena), also belongs to this assemblage. Excavates were originally proposed based on a distinctive feeding groove supported by a particular set of cytoskeletal features that are found in some, but not all, of these organisms (Simpson 2003). A strong molecular confirmation of this grouping is currently lacking, which is at least, in part, attributable to the high rates of sequence evolution of most of its putative constituent lineages. Single-gene phylogenies (e.g., Kolisko et al. 2008; Takishita et al. 2012) and phylogenomics (Rodriguez-Ezpeleta
et al. 2007a; Hampl et al. 2009) have generally recovered the monophyly of the group when the fast-evolving taxa are excluded, but some lineages, such as *Malawimonas*, have eluded robust placement (Hampl et al. 2009; Zhao et al. 2012).

**Figure 1.** Global tree of eukaryotes from a consensus of phylogenetic evidence (in particular, phylogenomics), rare genomic signatures, and morphological characteristics. Numerous eukaryotic groups are shown (not exhaustively), regardless of their taxonomic rank. Cartoons illustrate the diversity constituting the largest assemblages (colored boxes). The branching pattern does not necessarily represent the inferred relationships between the lineages. Dotted lines denote uncertain relationships, including conflicting positions. Note the solid branch leading to haptophytes and rappemonads: This illustrates the strong support for placing haptophytes as sister to SAR (stramenopiles, alveolates, and Rhizaria) in a recent study (Burki et al. 2012b), but this lineage is not included in a colored assemblage because confirmation is needed. The arrows point to possible positions for the eukaryotic root; the solid arrow corresponds to the most popular hypothesis (Amorphea-bikont rooting), the broken arrows represent the alternative hypotheses discussed in the text. (This figure was inspired by a template provided by Y. Eglit.)

**Archaebacteria (Plantae)**

This supergroup is composed of the three main lineages of primary photosynthetic taxa: organisms that harbor plastids directly derived from the cyanobacterial endosymbiosis. (1) The glau-
cophytes (e.g., *Cyanophora*) are a small group of enigmatic freshwater microscopic algae with uniquely cyanobacterial-like plastids (they have retained the prokaryotic peptidoglycan layer between the two plastid membranes). (2) The red algae (rhodophytes) form a diverse group of unicellular algae and large seaweeds, for example, the genus *Porphyra* commonly used to wrap sushi (nori). (3) The “green” organisms (*Viridiplantae*), including the green algae, with mostly free-living unicellular (e.g., *Chlamydomonas*), and colonial or multicellular taxa (e.g., *Volvox*, *Caulerpa*), but nonphotosynthetic parasitic taxa (e.g., *Prototheca*, *Helicosporidium*), are also known, as well as the land plants (mosses, ferns, angiosperms, etc.). Single-gene phylogenies and phylogenomics of plastid genes have strongly supported a common origin of the primary plastids in the ancestor of Archaeplastida (Chu et al. 2004; Hagopian et al. 2004; Rodriguez-Ezpeleta et al. 2005). Cell biology also supports a unique endosymbiosis with a shared plastid protein import machinery (Palmer 2003; McFadden and van Dooren 2004; Price et al. 2012). Nuclear phylogenies are less supportive (Moreira et al. 2000; Rodriguez-Ezpeleta et al. 2005; Nozaki et al. 2007, 2009), but this supergroup is generally widely recognized.

**SAR**

This vast assemblage is the most recently recognized supergroup and, contrary to the other supergroups, its existence is exclusively supported by molecular data (i.e., phylogenomic analyses (Burki et al. 2007, 2008; Hackett et al. 2007; Rodriguez-Ezpeleta et al. 2007a) and a derived RAB1 paralog (Elías et al. 2009). It was originally named SAR as an acronym of its constituents: stramenopiles, alveolates, and Rhizaria (Burki et al. 2007). Stramenopiles (also known as heterokonts) embrace a very large diversity of protists, including, for example, ecologically important algal groups such as diatoms or large multicellular seaweeds (e.g., kelps), as well as heterotrophic, often parasitic species such as oomycetes and *Blastocystis*. Stramenopiles are typically characterized by two opposing flagella with tripartite flagellar hairs (mastigonemes) on the forward flagellum, but also include many lineages that lost one or both flagella (Cavalier-Smith 1986). They are well supported by molecular data (Ben Ali et al. 2002; Cavalier-Smith and Chao 2006; Riisberg et al. 2009). Alveolates are also extremely diverse, notably including the dinoflagellate algae, but also apicomplexan parasites such as the malaria agent *Plasmodium* or ciliate protozoans (e.g., *Tetrahymena*). The cortical alveoli, a system of vesicles supporting the plasma membrane, constitute a morphological synapomorphy for the alveolates (Cavalier-Smith 1991). Molecular phylogenies strongly support the monophyly of alveolates (e.g., Fast et al. 2002; Harper et al. 2005). Rhizaria is based on molecular characters only (e.g., Keeling 2001; Archibald et al. 2003; Nikolaev et al. 2004; Bass et al. 2005; Burki and Pawlowski 2006; Burki et al. 2010; Brown et al. 2012; Sierra et al. 2013). It includes naked and testate amoeboid, heterotrophic protists with filose or reticulose pseudopods such as foraminifers and radiolarians. Some rhizarians are photosynthetic, including lineages in the euglyphids (e.g., *Paulinella*) and chlorarachniophytes (e.g., *Bigelowiella*). A large diversity of free-living flagellates and amoeboflagellates, sorocarpic (e.g., *Guttulina*), and parasite protists also belong to this group (see Pawlowski and Burki 2009 for a review). One enigmatic parasite, *Mikrocytos mackini*, which infects and kills oysters, was recently shown to belong to Rhizaria and attracted attention because it likely possesses highly reduced mitochondrion-derived organelle (Burki et al. 2013).

**THE RISE OF SAR**

Before being members of SAR, stramenopiles and alveolates were part of another supergroup, chromalveolates, which has played a central role in shaping our understanding of eukaryotic evolution, particularly the origin and spread of secondary plastids of red algal origin (Keeling 2009). In addition to stramenopiles and alveolates, the four original constituent chromalveolate lineages also included two important
protist groups: haptophytes and cryptomonads. Although many of these lineages are nonphotosynthetic or lack plastid altogether, the rational for grouping them into a monophyletic entity was based on the idea that the plastids of the chromalveolate taxa, which all share chlorophyll c, can be traced back to a single endosymbiotic event with a red alga (Cavalier-Smith 1999). Under this hypothesis, the secondary red plastid origin is unique and took place in the chromalveolate ancestor, and the numerous nonphotosynthetic lineages scattered among the chromalveolate tree were inferred to have lost their plastid and/or photosynthesis (Cavalier-Smith 1999). However, because of the absence of integrative molecular evidence supporting it, the chromalveolates have long been a controversial supergroup and it was recently remodeled in such a way that it has disappeared from the current consensus of the eukaryotic tree (Fig. 1) (Adl et al. 2012; Keeling 2013; Pawlowski 2013).

In phylogenetic terms, one condition of the chromalveolate hypothesis is that both the plastid and host (i.e., nuclear) trees must be consistent in showing the monophyly of alveolates, stramenopiles, haptophytes, and cryptomonads. From the plastid side, the monophyly has usually been recovered for three of these groups (Yoon et al. 2002; Khan et al. 2007). Alveolates, however, have proven nearly impossible to fit into this molecular framework because their plastid genomes are generally highly reduced, providing only few genes for comparative analyses (Köhler et al. 1997; Green 2004). But this was before the recent unexpected discovery of deep-branching relatives of apicomplexans (e.g., Chromera, Vitrella) (Moore et al. 2008), members of alveolates, which possess more gene-rich plastid genomes (Janouskovec et al. 2010). When the genomes of these species were compared to those of other chromalveolates, the tree that emerged showed a robust union between the alveolate and stramenopile plastids, and recovered the global monophyly of the red plastids also including haptophytes and cryptomonads (which were most closely related to each other), albeit with lower support (Janouskovec et al. 2010). Early nuclear phylogenies based on single genes also recovered a close association between alveolates and stramenopiles (Van de Peer et al. 1996; Harper et al. 2005). Phylogenomics largely confirmed this observation and, at the same time, produced trees in which haptophytes and cryptomonads shared a common ancestor, congruent with the plastid topology (Hackett et al. 2007; Patron et al. 2007; Burki et al. 2009). This latter association was also supported by a unique shared insertion of a laterally transferred rpl36 gene into the plastids of haptophytes and cryptomonads (Rice and Palmer 2006), which led some to propose the name Hacrobia to accommodate the body of evidence in favor of a common origin between these two groups (Okamoto et al. 2009).

However, what appeared to be a consistent scenario rapidly became challenged by the accumulating genetic data from diverse species as well as evidence against the monophyly of its original constituents. First, several phylogenomic studies showed that Rhizaria branched together with alveolates and stramenopiles (the SAR group) to the exclusion of haptophytes and cryptomonads, whose exact positions remained unresolved (Burki et al. 2007, 2008, 2009; Hackett et al. 2007; Rodriguez-Ezpeleta et al. 2007a). This was significant because Rhizaria include mostly heterotrophic groups and only two known photosynthetic lineages (chlorarachniophytes and Paulinella), neither of which possesses plastids of red algal origin. Thus, under the chromalveolate hypothesis, the SAR relationships imply that the ancestor of Rhizaria had a red-algal-derived plastid, which was lost before their diversification. At first glance, this might not substantially alter the chromalveolate hypothesis: Regardless of how many species belong to Rhizaria, only one more loss of an ancestral red plastid (out of many genuinely assumed) is required to explain the current plastid distribution, provided that SAR is closely related to Hacrobia.

In addition to Rhizaria, however, phylogenetic analyses showed that several other heterotrophic lineages were linked to the chromalveolates, further challenging the hypothesis.
Specifically, telonemids, centrohelids, katablepharids, and picobiliphytes (now Picozoa) (Sectonivasan et al. 2013) have all been inferred as sister to either haptophytes or cryptomonads (Shalchian-Tabrizi et al. 2006; Not et al. 2007; Burki et al. 2009), but these associations are generally poorly resolved, with the exception of katablepharids, which is robustly related to cryptomonads (Okamoto and Inouye 2005; Burki et al. 2012b). Nevertheless, the addition of new plastid-lacking lineages compromised both the hacrobian and chromalveolate hypotheses because both are based on a common plastid origin (Cavalier-Smith 1999; Okamoto et al. 2009). Proponents of the chromalveolate hypothesis, however, can always put forward the argument of plastid loss to explain these relationships, an argument that is difficult to refute until the prevalence of plastid loss among eukaryotes is better understood.

More problematic is evidence calling into question the existence of Hacrobia; in contrast to earlier reports, a recent phylogenomic study did not recover the association between haptophytes and cryptomonads, but instead showed haptophytes branching closer to SAR, whereas cryptomonads were sister to Archaeplastida (Burki et al. 2012b). These relationships were generally weakly supported, in particular, the position of cryptomonads, and thus require further testing before one can safely dismantle the Hacrobia hypothesis. But, by recovering cryptomonads distantly related to SAR and haptophytes, this study shifted attention on the position of this group as key to infer red plastid evolution (Burki et al. 2012b). Indeed, if the cryptomonad-Archaeplastida grouping is confirmed, it would not only invalidate the condition of a shared origin between all chromalveolate lineages in both nuclear and plastid phylogenies, it would also conflict with plastid phylogenies that strongly group haptophytes and cryptomonads (Janouškovec et al. 2010). Furthermore, a study evaluating the phylogenomic signal across the three genomic compartments (nuclear, plastid, and mitochondrial) in chromalveolate taxa reported discrepancies too high to be explained by a common origin of both the plastid and host lineages (Baurain et al. 2010). Altogether, these observations have forged the basis for alternative scenarios to the chromalveolate hypothesis: scenarios in which red plastids spread across the tree not by means of vertical inheritance, but through more complex serial eukaryote-to-eukaryote endosymbioses (Lane and Archibald 2008; Sanchez Puerta and Delwiche 2008; Archibald 2009; Bodily et al. 2009; Baurain et al. 2010; Dorrell and Smith 2011).

**DIGGING FOR THE ROOT**

In addition to improving the resolution of the eukaryotic tree, knowing precisely where the root lies is essential to give directionality to evolution and go beyond the classical star-like representation of the supergroups. This fundamental issue has proven extremely difficult to tackle because it refers to a prohibitively ancient event: the last eukaryotic common ancestor. The position of the root is generally considered unresolved, but a few hypotheses have emerged. The most straightforward approach to root a phylogenetic tree is to use an outgroup, which often consists of one or several lineages known to stem from the base of the group of interest. In the case of eukaryotes, one possible outgroup is to be found within prokaryotes. This is problematic because even the closest prokaryotic outgroup (Archaebacteria) represents a considerable evolutionary distance from eukaryotes, which could potentially lead to the LBA artefact (Felsenstein 1981). If fact, analyses including prokaryotes usually produced suspicious phylogenies in which the fastest evolving eukaryotes branched off close to the base of the tree, attracted to the distant outgroup (Philippe et al. 2000b; Ciccarelli et al. 2006; Williams et al. 2012).

To reduce the outgroup-to-ingroup distance, some researchers have used genes derived from the α-proteobacterium endosymbiotic progenitor of mitochondria (Derelle and Lang 2012). This analysis placed the eukaryotic root between unikonts and everything else (often referred to as bikonts) (Cavalier-Smith 2002), supporting what is perhaps the most prevailing view for the original bifurcation in the eukary-
otic tree. Two rare genomic changes have also supported the basal unikont–bikont split: (1) unikonts have ancestral, bacterial-like dihydrofolate reductase (DHFR) and thymidylate synthase (TS) genes, whereas bikonts have derived and fused version of these two genes (Philippe et al. 2000b; Stechmann and Cavalier-Smith 2002, 2003); and (2) unikonts have a unique glycine insertion to myosin class II, whereas this paralog is absent from bikonts (Richards and Cavalier-Smith 2005). Following a parsimonious argumentation, these two signatures were taken as evidence for the monophyly of unikonts and bikonts, and a root position outside of both groups. The unikont–bikont bifurcation was proposed to coincide with a fundamental difference of the flagellar apparatus: unikonts have retained the ancestral organization with one basal body anchoring one flagellum, whereas bikonts are ancestrally biflagellated with two basal bodies (Cavalier-Smith 2002; Stechmann and Cavalier-Smith 2002, 2003).

However, the sequencing of new enigmatic lineages has repetitively challenged the unikont–bikont bifurcation. For example, although the protistean apuzomonads (e.g., *Thecamonas*) are by essence bikonts (with two basal bodies and two flagella) and possess the fused TS-DHFR, they were shown to be related to unikonts (Richards and Cavalier-Smith 2005). Following a parsimonious argumentation, these two signatures were taken as evidence for the monophyly of unikonts and bikonts, and a root position outside of both groups. The unikont–bikont bifurcation was proposed to coincide with a fundamental difference of the flagellar apparatus: unikonts have retained the ancestral organization with one basal body anchoring one flagellum, whereas bikonts are ancestrally biflagellated with two basal bodies (Cavalier-Smith 2002; Stechmann and Cavalier-Smith 2002, 2003).

However, the sequencing of new enigmatic lineages has repetitively challenged the unikont–bikont bifurcation. For example, although the protistean apuzomonads (e.g., *Thecamonas*) are by essence bikonts (with two basal bodies and two flagella) and possess the fused TS-DHFR, they were shown to be related to unikonts (Kim et al. 2006; Brown et al. 2012; Derelle and Lang 2012). Similarly, the breviate amoeba (e.g., *Breviata*), which harbors two basal bodies (but only one flagellum), turned out to be a deep-branching unikont (Minge et al. 2009; Roger and Simpson 2009; Brown et al. 2013). These data suggest that the true unikonts, in addition to some protists not fitting with the original description of a monoflagellated ancestry and separate TS and DHFR genes, may form a clade, but it seems increasingly untenable that this group was ancestrally unikont with one basal body. To formalize this taxon without reflecting the controversial ancestral state, the name Amorphea was recently introduced (Adl et al. 2012).

Other analyses have further exacerbated the uncertain position of the eukaryotic root. Recently, a study looking at the taxonomic distribution of the endoplasmic reticulum (ER)-mitochondria encounter structure, which tethers mitochondria to the ER membrane, showed that the root is most consistently placed between Amorphea + excavates and all other eukaryotes (Wideman et al. 2013). Researchers investigating a new class of rare genomic changes involving multiple, conserved amino acid residues inferred the initial split in eukaryote evolution between Archaeplastida and everything else (Rogozin et al. 2009). A different approach showed that by minimizing the number of gene duplications and loss across 20 genes, the most parsimonious root was found to lie between opisthokonts and the rest of eukaryotes (He et al. 2014). The list of hypotheses goes on, and it may be some time before the root is dug out, but when its position is revealed with more consistency, it will have a fundamental impact on our understanding of how the eukaryotic supergroups relate to one another.

**DEEP RELATIONSHIPS AMONG EUKARYOTES**

In an attempt to best reflect the current view of eukaryotic relationships, I present a rooted tree with the origin of eukaryotes between Amorphea and everything else, but also indicate five alternative rooting positions (Fig. 1). Amorphea includes opisthokonts, Amoebozoa, and several protist lineages of uncertain phylogenetic placement, such as apusomonads, ancyromonads, and the breviate amoebae. (Heiss et al. 2011; Katz et al. 2011; Brown et al. 2013). The rest of eukaryotes comprises excavates and the recently erected mega-clade Diaphoretickes...
(Adl et al. 2012), which recognizes the SAR and Archaeplastida clade (Burki et al. 2008) and corresponds to the corticates of Cavalier-Smith (2010a). Other incertae sedis lineages, such as haptophytes, cryptomonads, katablepharids, teylonemids, centrohelids, Rappemonads, and Palpitomonas, may also belong to Diaphoretickes, as compiled from several phylogenetic results (Rodriguez-Ezpeleta et al. 2007a; Burki et al. 2009; Hampl et al. 2009; Yabuki et al. 2010; Kim et al. 2011; Brown et al. 2012). The branching pattern within Diaphoretickes is generally poorly resolved, but haptophytes may be sister to SAR (Burki et al. 2012b). The position of cryptomonads, although ambiguous, has recently shown affinities to Archaeplastida (Burki et al. 2012b). The other lineages have failed to show reliable phylogenetic placement and remain enigmatic.

With this evolutionary framework in mind, one group, Collodictyon, stands out as particularly remarkable. Collodictyon is an omnivorous amoeboflagellated cell with a mysterious position in the tree of eukaryotes. It possesses an eclectic mix of cellular features, such as a ventral feeding groove typical of excavates and amoebozan-like pseudopods, which make pinpointing its phylogenetic position difficult but desirable (Klaveness 1995; Brugerolle et al. 2002). When its position was investigated using a large phylogenomic data set, Collodictyon branched off early in eukaryote evolution, close to the Amorphea-bikont bifurcation (Fig. 1), suggesting that its morphological characteristics may represent some of the ancestral conditions of the eukaryotic cell (Zhao et al. 2012). With the potential exception of the contentious excavate Malawimonas, which has proven impossible to unambiguously assign to any large group of eukaryotes and showed affinities to Collodictyon, this branch of the eukaryotic tree is known to include only one other genus (Diphyllleia), a view supported by observations of very low genetic diversity of Collodictyon-like sequences in environmental surveys (Zhao et al. 2012). Given a root as in Figure 1, this group may represent the first confirmed case of a lineage that diverged after the Amorphea-bikont split, but before the subsequent diversification of these two groups.

CONCLUDING REMARKS

Over the last 10 years, phylogenomics has led to important refinements of the global tree of eukaryotes. Most of the large building blocks of the tree (the supergroups) predating the genomic era have been reinforced by the analyses of larger data sets, but some were also shuffled into different arrangements. The increased phylogenetic power of phylogenomics helped resolve the relationships between the supergroups, providing new hypotheses for the deep backbone of the tree. It also allowed the placement of orphan taxa that had eluded proper classification until more data became available. At the same time, new challenges have appeared and several key lineages remain of unknown evolutionary origin. It is clear that the dawn of eukaryote evolution and subsequent diversification will not be fully understood until these enigmatic lineages find a home and the root of the tree is characterized. What’s more, as we explore incertae sedis taxa, classical microscopy as well as next-generation environmental surveys and single-cell genomics will undoubtedly reveal new essential protist taxa. But the exciting news is: Technological advancements such as sequencing preparation from nano-quantities of material now allow one to tackle these taxa from a genomic perspective, even when cell cultures cannot be established (Yoon et al. 2011).

Resolving the tree of eukaryotes will necessitate continuing the integrative approach blending morphology, single-gene phylogeny, and phylogenomics including all diversity. It will also require more reliable ways of assembling genomic data and the development of new methods to extract the phylogenetic information contained in these genomes. Importantly, the pervasiveness of HGT in eukaryotes, in particular, the genes transferred from plastids (endosymbiotic gene transfer), will need to be systematically evaluated. Presently, the fierce debate over the global impact of HGT on eukaryote evolution and its harmful consequences on phylogenomics is far from a consensus (Moustafa et al. 2009; Stiller 2011; Burki et al. 2012a; Chan et al. 2012; Deschamps and Moreira 2012). Eukaryote evolution started off more
than a billion years ago and reconstructing the twists and turns that led to the diversity we observe today is a tremendously difficult task, but a task that is both fascinating and on its way to be deciphered.

ACKNOWLEDGMENTS

I express my deepest gratitude to Yana Eglit, Thierry Heger, Jan Pawlowski, David Smith, and Barbara Yerman for critical and constructive reading of an earlier version of this manuscript. My research is supported by a grant from the Tula Foundation to the Centre for Microbial Diversity and Evolution.

REFERENCES


Burki F, Okamoto N, Pombert J-F, Keeling PJ. 2012b. The evolutionary history of haptophytes and cryptophytes:


Klaveness D. 1995. *Collodictyon triciliatum* H.J. Carter (1865)—A common but fixation-sensitive algivorous fla-
The Tree of Eukaryotes


The Tree of Eukaryotes


The Eukaryotic Tree of Life from a Global Phylogenomic Perspective

Fabien Burki

Cold Spring Harb Perspect Biol 2014; doi: 10.1101/cshperspect.a016147

Subject Collection  The Origin and Evolution of Eukaryotes

The Persistent Contributions of RNA to Eukaryotic Gen(om)e Architecture and Cellular Function

Jürgen Brosius

Green Algae and the Origins of Multicellularity in the Plant Kingdom

James G. Umen

The Archaean Legacy of Eukaryotes: A Phylogenomic Perspective

Lionel Guy, Jimmy H. Saw and Thijs J.G. Ettema

Origin and Evolution of the Self-Organizing Cytoskeleton in the Network of Eukaryotic Organelles

Gáspár Jékely

On the Age of Eukaryotes: Evaluating Evidence from Fossils and Molecular Clocks

Laura Eme, Susan C. Sharpe, Matthew W. Brown, et al.

Origin of Spliceosomal Introns and Alternative Splicing

Manuel Irimia and Scott William Roy

Protein and DNA Modifications: Evolutionary Imprints of Bacterial Biochemical Diversification and Geochemistry on the Provenance of Eukaryotic Epigenetics

L. Aravind, A. Maxwell Burroughs, Dapeng Zhang, et al.

The Eukaryotic Tree of Life from a Global Phylogenomic Perspective

Fabien Burki

Eukaryotic Origins: How and When Was the Mitochondrion Acquired?

Anthony M. Poole and Simonetta Gribaldo

Bacterial Influences on Animal Origins

Rosanna A. Alegado and Nicole King

Missing Pieces of an Ancient Puzzle: Evolution of the Eukaryotic Membrane-Trafficking System

Alexander Schlacht, Emily K. Herman, Mary J. Klute, et al.

The Neomuran Revolution and Phagotrophic Origin of Eukaryotes and Cilia in the Light of Intracellular Coevolution and a Revised Tree of Life

Thomas Cavalier-Smith

Protein Targeting and Transport as a Necessary Consequence of Increased Cellular Complexity

Maik S. Sommer and Enrico Schleiff

How Natural a Kind is "Eukaryote?"

W. Ford Doolittle

What Was the Real Contribution of Endosymbionts to the Eukaryotic Nucleus?

Insights from Photosynthetic Eukaryotes

David Moreira and Philippe Deschamps

Bioenergetic Constraints on the Evolution of Complex Life

Nick Lane

For additional articles in this collection, see http://cshperspectives.cshlp.org/cgi/collection/

Copyright © 2014 Cold Spring Harbor Laboratory Press; all rights reserved