Cell PRESS

The eukaryotic tree of life: endosymbiosis takes its TOL

Christopher E. Lane and John M. Archibald

The Canadian Institute for Advanced Research, Integrated Microbial Biodiversity Program, Department of Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, NS B3H 1X5, Canada

Resolving the structure of the eukaryotic tree of life remains one of the most important and challenging tasks facing biologists. The notion of six eukaryotic 'supergroups' has recently gained some acceptance, and several papers in 2007 suggest that resolution of higher taxonomic levels is possible. However, in organisms that acquired photosynthesis via secondary (i.e. eukaryote–eukaryote) endosymbiosis, the host nuclear genome is a mosaic of genes derived from two (or more) nuclei, a fact that is often overlooked in studies attempting to reconstruct the deep evolutionary history of eukaryotes. Accurate identification of gene transfers and replacements involving eukaryotic donor and recipient genomes represents a potentially formidable challenge for the phylogenomics community as more protist genomes are sequenced and concatenated data sets grow.

Evolution of the eukaryotic tree of life

A well-resolved and accurate phylogenetic reconstruction of life on Earth is the ultimate goal of systematics. Although the possibility of ever realizing this goal is a point of considerable debate in prokaryotes (see the review in this issue by McInerney *et al.* [1]), it might still be achievable for eukaryotes. A robustly supported eukaryotic tree of life (TOL) would provide an important foundation for research targeted at understanding the evolution of a myriad of traits in this domain, including adaptation to anaerobic environments and the origins of multicellularity and photosynthesis. Over the last 50 years, significant advances in taxon sampling and methodology have increased our understanding of eukaryotic evolutionary relationships and dramatically altered our view of high-level diversity. This is especially true of unicellular eukaryotes (protists), which are now recognized as being scattered throughout the tree rather than mostly belonging to a single clade (e.g. Refs [2,3]). Whereas early phylogenetic analyses relied primarily on the information contained within a single gene (i.e. that encoding 18S rRNA), and were thus limited in terms of their resolution (and in some cases positively misleading [4]), the genomics revolution has brought an ever-increasing amount of data to bear on the question of protist phylogeny and, more generally, the ancient divergences among eukaryotic lineages.

Research in the last decade has led to the hypothesis of six eukaryotic 'supergroups' (Figure 1), erected on the basis of an eclectic mix of morphological and molecular sequence data [5,6]. The strength of the evidence supporting these superassemblages (summarized in the accompanying Glossary) has been the subject of much debate [7] and the relationships between the supergroups are largely unknown. Indeed, whether molecular data can accurately resolve relationships between taxa that diverged ca. one billion years ago is unclear. In photosynthetic eukaryotes, the problem of ancient divergence times is confounded by

Glossary

Alveolata: composed of the apicomplexans, dinoflagellate algae and ciliates (e.g. *Tetrahymena*), a trio whose shared common ancestry is very well supported in nuclear gene phylogenies.

Amoebozoa: composed of lineages of unicellular organisms, generally well supported in molecular phylogenies; without clear-cut morphological apomorphies, although most lineages produce pseudopodia that are broad (lobose).

Archaeplastida: includes all primary plastid-containing lineages, and is well supported in both plastid and phylogenomic-scale nuclear gene phylogenies. Molecular evidence for the single origin of red algal, green algal and glaucophyte plastids is supported by the structure of plastid genomes and the light-harvesting complex.

Chromalveolata: predominantly unicellular collection of photosynthetic and nonphotosynthetic organisms united by the 'chromalveolate hypothesis' (Box 2), which states that the plastids of chromists and alveolates are the product of a single secondary endosymbiosis in the common ancestor of the two groups. Support for this group is based largely on plastid-related characters between subsets of its component lineages, with no single character or phylogeny that has been shown to unite all of its hypothesized members.

Chromists: include cryptophytes and haptophytes, two predominantly photosynthetic algal groups, and the stramenopiles, a group of unicellular (e.g. diatoms) and multicellular (e.g. kelp) algal species together with a diverse array of nonphotosynthetic free-living and parasitic lineages.

Endosymbiotic gene replacement (EGR): a specific case of EGT, where an endosymbiont gene is transferred to the host nucleus and replaces the function of a nuclear-encoded gene.

Endosymbiotic gene transfer (EGT): genes transferred from the genome of an endosymbiont to the nucleus of the host, where they can be degraded, assume novel functions, replace host genes (EGR) or acquire a targeting signal(s) so that their protein products are directed back to the endosymbiont compartment.

Excavata: encompass unicellular eukaryotes that share a distinctive ventral feeding groove and an array of cytoskeletal features (and their relatives as determined by molecular means). The common ancestry of the group as a whole is, at best, weakly supported by published molecular data.

Expressed sequence tag (EST): a single sequencing read of a cloned mRNA-derived cDNA, isolated from an organism. Multiple reads can be assembled to produce the entire sequence of a mature transcript, which can then be used in place of the genomic copy of the sequence in phylogenetic analyses.

Opisthokonta: includes animals, fungi and their unicellular relatives (such as *Capsaspora* and choanoflagellates), which share (or are derived from organisms with) a single posterior flagellum and are strongly supported by sequence data, including a unique shared insertion in the genes encoding Elongation Factor 1-α and enolase.

Rhizaria: a group united only by molecular phylogenies. Members include ecologically important and abundant organisms, such as foraminiferans and cercozoans, which are largely understudied.

Corresponding author: Lane, C.E. (c.lane@dal.ca).

0169-5347/$ – see front matter © 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.tree.2008.02.004 Available online 2 April 2008

the fact that the nuclear genome is a mosaic of genes with different evolutionary histories. The cyanobacterial progenitor of the plastid (chloroplast) in modern-day plants and algae donated thousands of prokaryotic genes to its eukaryotic host during the initial transition from endosymbiont to organelle [8], and plastids subsequently spread laterally by endosymbioses between two eukaryotic cells, a process known as 'secondary' endosymbiosis [9,10]. In the case of secondary endosymbiosis, the plastid acts as a genetic Trojan horse, bringing with it the nucleus of an unrelated eukaryotic endosymbiont whose genes meld with – and can replace – their counterparts in the host nuclear genome. The mixing and matching of eukaryotic genes that occurs in the context of secondary endosymbiosis seriously challenges our ability to accurately infer the evolutionary history of these organisms. At the same time, a comprehensive understanding of the impact of endosymbiosis on the structure and content of nuclear genomes has the potential to assist our efforts to resolve the structure of the eukaryotic tree of life.

The imprint of endosymbiosis

The process of endosymbiosis has been responsible for some of the most significant events in eukaryotic evolution. The most celebrated examples are the endosymbioses that gave rise to plastids and mitochondria. In the case of plastids, a wealth of biochemical and molecular data indicate that a prokaryotic relative of modern-day cyanobacteria was engulfed and retained by a heterotrophic eukaryote [11] and transformed into the photosynthetic organelle seen in the plants and algae of today (Box 1). Integration of a prokaryotic endosymbiont into the cellular machinery of a eukaryote is a complex process and involves substantial modifications to the genetic makeup of both cells [10]. All known organelles of endosymbiotic origin encode only a fraction of the genes present in their prokaryotic antecedents, meaning that scores of genes that were once essential to the free-living prokaryote, but obsolete in the context of intracellular life (e.g. genes for locomotion or certain metabolic pathways), are lost, either by transfer to the host nucleus or by genomic degradation [12]. Even the most gene-rich plastid genomes contain at most ~250 genes – an order of magnitude less than most free-living cyanobacteria [13]. As the genetic capacity of the prokaryotic endosymbiont is reduced during the transition from free-living cell to fully integrated organelle, the host cell becomes a repository of genetic information by way of endosymbiotic gene transfer (EGT) [8]. Many of the genes transferred to the nucleus acquire targeting signals, which allow their products to be shuttled back to the plastid to perform vital functions [10]. However, transferred genes can also assume novel functions in the eukaryotic cell, sometimes even replacing eukaryotic versions of the proteins they encode.

Comparative genomic studies of each of the three lineages of 'primary' plastid-containing eukaryotes (Box 1) have revealed just how significant endosymbiont-to-nucleus gene transfer is. A 2002 study of the nuclear genome of the land plant *Arabidopsis thaliana* suggested that an unexpectedly large fraction of its genes (1700/9368 = ~18%) are derived from the cyanobacterial progenitor of the plastid [8]. More recent studies using different approaches have produced lower estimates of EGT in *A. thaliana* [14,15] and proportionately fewer genes of cyanobacterial origin appear to reside in the nuclear genome of the green alga *Chlamydomonas reinhardtii* (6%) [16], the red alga *Cyanidioschyzon merolae* (12.7%) [14] and the glaucophyte alga *Cyanophora paradoxa* (10.8%) [17]. Regardless, it is clear that endosymbiont-derived genes have contributed substantially to the host cell nucleus. In the case of *A. thaliana*, more than half of the genes of putative cyanobacterial origin are predicted to perform functions unrelated to the plastid [8], indicating that these genes have either taken on new functions or replaced nuclear genes (termed endosymbiotic gene replacement; EGR). A specific example of EGR is phosphoglycerate kinase (PGK) [18], which occurs in two copies within land plants, with a plastid-encoded homolog functioning in the Calvin cycle and a nucleus-encoded copy involved in glycolysis in the cytosol. Phylogenetic analysis shows that both are derived from cyanobacteria, indicating that the plastid copy was transferred to the nucleus and duplicated before replacing the resident eukaryotic gene encoding PGK in the nucleus [18]. Genes transferred from the plastid genome to the nucleus are, for the most part, easily identified because of their phylogenetic affinity to

Box 1. The birth and spread of eukaryotic photosynthesis

Plastids (chloroplasts) are the light-harvesting organelles of photosynthetic eukaryotes and are derived from once free-living cyanobacteria by the process of endosymbiosis. It is generally accepted that plastids evolved from cyanobacteria only once during the history of life, in a common ancestor shared by green algae (and plants), red algae and glaucophyte algae (reviewed by Ref. [36]). There is, nevertheless, debate as to when this occurred (e.g. Refs [46,47]), and some authors (e.g. Refs [48–50]) believe that the notion of multiple primary endosymbiotic events should not be discounted.

Subsequent to the diversification of the three 'primary' plastid-containing lineages, plastids spread laterally across the tree of eukaryotes via the process of 'secondary' endosymbiosis, that is, the engulfment and retention of a eukaryotic phototroph by an unrelated nonphotosynthetic eukaryotic host. The number of times this has occurred is a topic of considerable debate, but most recognized hypotheses propose three or more events. Two unrelated lineages, the euglenids (Excavata) and chlorarachniophytes (Rhizaria), contain plastids of green algal ancestry [21], whereas plastids in the supergroup Chromalveolata are, with the exception of some dinoflagellates (below), descended from a red alga [51]. The morphological evidence for the eukaryotic origin of plastids in these three groups comes from the presence of three or four membranes surrounding secondary plastids (versus two membranes around primary plastids) and the remnant plastid-associated nuclei (nucleomorphs) that exist in chlorarachniophytes and cryptophytes (reviewed in Ref. [52]). Organisms harboring secondary plastids are thus the biological equivalent of nested Russian dolls – a cyanobacterium encased within a eukaryote, enveloped within a second eukaryote (Figure 2). To further complicate matters, some dinoflagellate algae have replaced their plastid with a 'tertiary' plastid, stolen from other chromalveolates including cryptophytes, diatoms and haptophytes, and even a green algal plastid in a case of serial secondary endosymbiosis (see Refs [37,53] and references therein). Therefore, although all plastids probably trace back to a single primary endosymbiotic event, they have been propagated throughout eukaryotic diversity multiple times by a similar process.

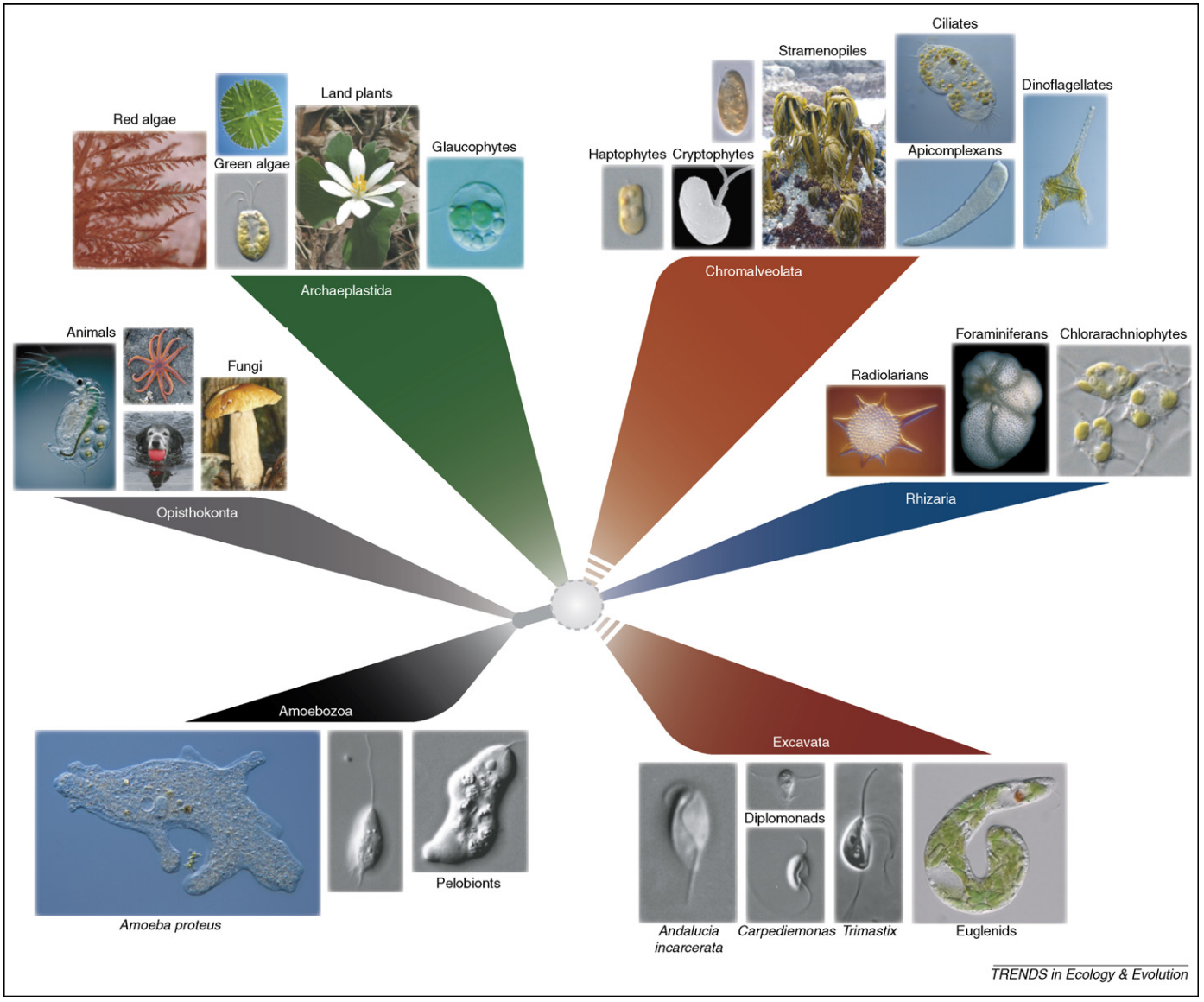

Figure 1. The six hypothesized supergroups of eukaryotes. There is currently no consensus regarding the root of eukaryotes or the branching order between these groups, except that recent studies indicate a sister relationship between Amoebozoa and Opisthokonta. Dashed lines at the base of Chromalveolata and Excavata indicate uncertainty regarding the monophyly of these assemblages. Images show representatives of some of the major lineages within each supergroup. Images are credited to the following: Chromalveolata: haptophyte and cryptophytes by the Archibald laboratory; stramenopile by G.W. Saunders; apicomplexan by B.S. Leander; ciliate and dinoflagellate from Visuals Unlimited (VU). Archaeplastida: red alga by C. Bates (http://www.coastalimageworks.com); green algae by the Archibald laboratory and from VU; land plant by J. Palmer; glaucophyte by D. Patterson, provided with permission by http://microscope.mbl.edu. Rhizaria: radiolarian and foraminiferan from VU; chlorarachniophyte by the Archibald laboratory. Excavata: euglenid by D. Patterson, provided with permission by http://microscope.mbl.edu; remaining excavates by A.G.B. Simpson. Amoebozoa: *Amoeba proteus* from VU; pelobionts from A.G.B. Simpson. Opisthokonta: Animals from W. F. Doolittle, C. Bates (http://www.coastalimageworks.com) and VU; fungus by G. Burger.

cyanobacteria. However, in eukaryotes that have acquired photosynthesis secondarily through the engulfment of a primary plastid-containing alga, the impact of EGT is much more subtle but no less significant.

The impact of secondary endosymbiosis on genome evolution

Three of the six eukaryotic supergroups (Figure 1), Rhizaria, Excavata and Chromalveolata (see Glossary), include organisms with secondary plastids (Box 1). Within the Rhizaria and Excavata, chlorarachniophytes and euglenids are the sole photosynthetic lineages, respectively, and both contain green algal secondary plastids [19,20]. As described in Box 1, secondary plastids have evolved on multiple occasions from both red and green algal endosymbionts. Whereas chlorarachniophytes and euglenids acquired plastids independent of one another and are members of predominantly nonphotosynthetic supergroups [21,22], most plastid-bearing members of the contentious supergroup Chromalveolata possess red algal-derived secondary plastids, which has been taken as a sign of monophyly of the group (the 'chromalveolate hypothesis').

Unlike primary endosymbiosis, secondary endosymbiosis involves the genetic integration of two eukaryotes. In this case, the host nucleus is bombarded with genes not only from the plastid genome but also from the nucleus of the endosymbiont, which in turn already contains cyanobacterial-derived genes that were previously transferred from the primary plastid (Figure 2). Indeed, clear cases

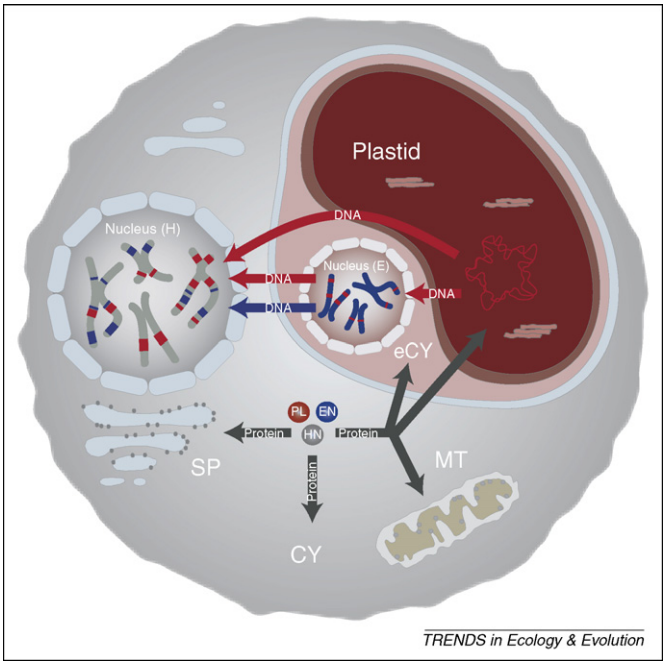

Figure 2. Endosymbiosis and the flow of genetic information in photosynthetic eukaryotes. A generic secondary plastid-containing alga with arrows indicating the movement of genes from the plastid genome to the endosymbiont nuclear genome, from the plastid genome to the nuclear genome of the host cell (H) and from the plastid and endosymbiont nuclear (E) genomes to the host nuclear genome. The result of this gene transfer is that the nuclear genome of the host cell is a mosaic of cyanobacterial (red), nucleomorph (blue) and host (gray) DNA. The protein products of genes derived from these three genomes (colored balls; PL = plastid; EN = endosymbiont; HN = host nucleus) have the potential to take on roles in various cellular compartments, including the cytosol (CY), mitochondrion (MT) and secretory pathway (SP) of the host cell, as well as the plastid and endosymbiont cytosol (eCY). Movement of genes associated with the mitochondria of the endosymbiont and host cell has been omitted for clarity, as have genes derived by lateral gene transfer (see main text).

of EGR involving nuclear genes for plastid-targeted proteins have provided some of the most compelling evidence in support of the chromalveolate hypothesis [23,24]. Genome-scale analyses of secondary plastid-containing organisms have also begun to reveal the extent of EGT. For example, expressed sequence tag (EST) data sets from the haptophyte *Emiliana huxleyi* and the dinoflagellate *Karenia brevis* have revealed 19 and 17 putative cases of EGT, respectively [25,26]. The nuclear genome sequence of the diatom *Thalassiosira pseudonana* [27] encodes 1057 genes (of 11 242 = 9.4%) that show sequence homology to only plants and algae and 271 genes with only cyanobacterial homologs. Additionally, an analysis of ESTs from *Euglena gracilis* (a photosynthetic excavate; Figure 1) found that 22% of 259 globally distributed genes showed phylogenetic affinity to primary plastid-containing lineages [28], suggesting that these genes represent cases of eukaryote–eukaryote EGT and, possibly, EGR. Instances of gene transfer can also point to a photosynthetic ancestry for nonphotosynthetic organisms, such as the parasitic stramenopile *Phytophthora ramorum* (the causative agent of the Irish potato blight), which does not contain a plastid. Strikingly, the identification of 855 genes of putative red algal or cyanobacterial ancestry in the *P. ramorum* nuclear genome [29] demonstrates that transferred genes can take on many functions in their new genomic context, unrelated to photosynthetic processes.

Therefore, with respect to broad-scale eukaryotic phylogenetics, the question is, how do we detect eukaryote-to-eukaryote EGT and EGR events and exclude them from our analyses so that only the genes that accurately reflect the history of the host cell lineage are included? Genes of prokaryotic origin are readily identified, assuming the time since their transfer has not erased the phylogenetic signal linking them to other prokaryotes. However, in cases of endosymbiont-to-host nuclear transfer or replacement, the eukaryotic origin of the foreign DNA makes its detection significantly more difficult. Unless the transfer is recent and the evolutionary position of the host nucleus in known, cases of eukaryote-to-eukaryote EGT or EGR are unlikely to be obvious. Additionally, a eukaryotic gene is presumably more likely to be expressed and to function if transferred to the host nucleus than is a prokaryotic gene, suggesting that the rate of successful eukaryotic EGTs and EGRs could be higher than those involving mitochondrion- or plastid-derived genes. If the number of 'foreign' eukaryotic genes in the genomes of secondary plastid-containing (or formerly containing) organisms is significant, but difficult to detect, then the evolutionary history of the host organism would be very difficult to resolve, because of the presence of genes with two (or more) evolutionary histories.

To make matters worse, detection of eukaryotic EGTs and EGRs should be even more difficult, owing to lineage-specific evolutionary rates or gene loss, when taxon sampling for a group of interest is low and the hypotheses being tested involve deep divergences. Eukaryote-wide phylogenies often include only a small number of representative taxa for lineages such as Rhizaria or some of the chromalveolate phyla. In addition, the taxa used to represent major groups in phylogenomic analyses are often highly reduced or simplified organisms, whose genome might not be an appropriate proxy for the majority of the lineage. This is both an artifact of targeting organisms that impact humans (e.g. pathogens) and the need to choose small genomes to make projects manageable. Although this situation is improving, and EST data can often be used in lieu of complete genomes, there is currently an extremely limited set of organisms available for comparison. In instances where the secondary plastid is of red algal origin, whole-genome comparisons must be made to the only red algal nuclear genome that has been completely sequenced, that of *Cyanidioschyzon merolae* [30]. Although a genome sequence from a second red alga is near completion (see the *Galdieria sulphuraria* Genome Project, http://genomics.msu.edu/galdieria/about.html), both of these genomes are from highly reduced unicellular organisms, adapted to life at high temperatures and belonging to a single order (Cyanidiales) among the poorly understood early-diverging subphylum Cyanidophytina [31]. Of the roughly 6000 recognized species of red algae, only four are classified in this subphylum. Therefore, through the combined effects of limited taxon sampling, lineage-specific variation in evolutionary rates and compositional biases in DNA and protein sequences [32], red algal genomic data currently available make the reliable detection of EGT in chromalveolate genomes a significant challenge.

Despite an overwhelming amount of evidence demonstrating that endosymbiont-derived genes readily establish themselves in the nuclear genome of their host, to our knowledge only a handful of studies have thoroughly investigated *bona fide* transfers of a eukaryotic gene from a secondary endosymbiont nucleus to a host nucleus. For example, two studies have convincingly shown that the cryptophyte nuclear genome possesses a gene encoding actin that is derived from the red algal nucleus that came in with the chromalveolate plastid [33,34]. In this case both the host and endosymbiont copies persist, but instances of complete EGR can often pass phylogenetic screening methods commonly used to detect anomalous genes in large data sets unless the issue of secondary EGR is specifically addressed [35].

The supergroup Chromalveolata – a playground for the evolution of photosynthesis

The supergroup Chromalveolata is composed of the 'chromists' and Alveolata whose evolutionary origins are a subject of active debate. Chromalveolates are a diverse assemblage of mostly unicellular lineages, and the complex distribution of plastids among chromalveolate taxa has fueled the debate regarding the monophyly of these organisms (Box 2). Of the chromists, cryptophytes and haptophytes are two predominantly photosynthetic algal groups, and the stramenopiles are a group of unicellular (e.g. diatoms) and multicellular (e.g. kelp) algal species together with a diverse array of nonphotosynthetic free-living and parasitic lineages (Figure 1). Of the Alveolata, dinoflagellates are microalgae that are notorious for producing red tides, but roughly half of dinoflagellates lack plastids. The Apicomplexa (a mostly parasitic assemblage, including the causative agent of malaria) have lost their photosynthetic capabilities, although species in such medically important genera as *Plasmodium* and *Toxoplasma* retain plastids in a highly reduced form (the apicoplast) that, on balance, appear to be of red algal origin [36]. By contrast, the ciliates are an exclusively plastid-free lineage. Although the common origin of chromalveolate plastids from a single endosymbiotic event is contentious, it is generally accepted that all but a few of the photosynthetic taxa that fall under this label contain plastids derived from a red algal endosymbiont. A handful of dinoflagellates have clearly taken on plastids from other chromalveolate lineages in tertiary endosymbiotic events (e.g. *Dinophysis*, *Karenia*, *Kryptoperidinium*), and the green algal plastid of *Lepidodinium* is an instance of serial secondary endosymbiosis (see Ref. [37] and references therein).

Shuffling supergroups

In 2007, a series of papers from independent research groups began to pull apart the chromists (cryptophytes + haptophytes + stramenopiles) and the chromalveolate concept as a whole or, at the very least, significantly complicate the hypothesis. Several studies [35,38,39] resolved topologies that break apart chromists, and instead unite stramenopiles with Alveolata with significant support. The finding is independently supported by an *rpL36* gene replacement in the plastid of both cryptophytes and haptophytes, suggesting these two lineages share a common ancestor [40]. Unexpectedly, however, in the analyses that included members of Rhizaria [35,39], this supergroup was resolved within chromalveolates, either as sister to stramenopiles + alveolates, to the exclusion of cryptophytes + haptophytes, or specifically sister to stramenopiles (Figure 3b). However, in one instance [35], a topology including a monophyletic origin of chromalveolates (excluding Rhizaria) could not be rejected by the data.

If chromalveolate taxa do indeed share a common ancestor (Box 2), the resolution of the supergroup Rhizaria within Chromalveolata not only complicates hypotheses regarding the evolution of secondary plastids (Figure 3a), it also demands that the ancestor of Rhizaria harbored a red algal-derived plastid. Specifically, the red algal plastid

Box 2. The chromalveolate controversy

Save for plastid-associated features, cellular characters that support the chromalveolate hypothesis (i.e. a monophyletic origin of Chromalveolata) are lacking, as is phylogenetic evidence based on nuclear DNA sequences. Molecular phylogenetic analyses of nuclear genes that are consistent with Chromalveolata have only included alveolates (ciliates, dinoflagellates and apicomplexans) and stramenopiles [54] and analyses of plastid genes, or nuclear-encoded genes for plastid-targeted proteins have provided varying degrees of support for the monophyly of chromists + dinoflagellates [47,53]. Critics of the chromalveolate hypothesis point out that such analyses only show a common ancestry of the plastid, not necessarily the host lineages in which they reside [55,56]. Thus, alternative models of plastid transfer among chromalveolate taxa have been suggested, specifically invoking a model of tertiary transfer of the haptophyte plastid [55,56]. This alternative hypothesis suggests that the red algal-derived plastids of 'chromalveolates' are united by horizontal transfer, not vertical inheritance from a common ancestor. Nevertheless, endosymbiotic gene replacements of nuclear-encoded genes for plastid-targeted or plastid-derived proteins have been described that unite photosynthetic chromalveolate taxa [23,24]. Although alternate explanations for the observed pattern of gene transfers have been proposed [55], the number of such cases continues to grow with more data.

Central to the debate of chromalveolate monophyly is the weight placed on plastid gain versus plastid loss. The main rationale for proposing chromalveolate monophyly is that it greatly reduces the number of secondary endosymbioses required to account for present-day photosynthetic diversity, a position that has previously been argued based on the presumed difficulty associated with the genetic integration of host and endosymbiont [51]. If plastid gain is much more difficult than plastid loss, it appears most plausible to hypothesize a single plastid acquisition in the common ancestor of all chromalveolates, and subsequent loss in the ciliates, some apicomplexans, and early-diverging members of many chromalveolate lineages (Figure 3a). An alternative view is that organelles are inherently more difficult to lose than gain [56,57], based on the fact that many nonphotosynthetic apicomplexan parasites retain an apicoplast (a remnant plastid used for cellular functions such as fatty acid, heme and amino acid biosynthesis) and the fact that derivatives of the mitochondrion have been found in every case where a putatively amitochondriate organism has been thoroughly investigated [4]. However, direct evidence for plastid loss from the apicomplexan *Cryptosporidium* and the stramenopile *Phytophthora* (e.g. Refs [29,41,58]) demands a reevaluation of this argument. Additionally, the description of a reduced plastid in the dinoflagellate *Perkinsus* [59], presumed to be homologous to the apicoplast, and the discovery of *Chromera velia*, a photosynthetic alveolate closely related to apicomplexans [42], suggest that the common ancestor of dinoflagellates and apicomplexans was a phototroph, further pushing back the plastid-bearing roots of chromalveolate lineages and indicating subsequent plastid loss.

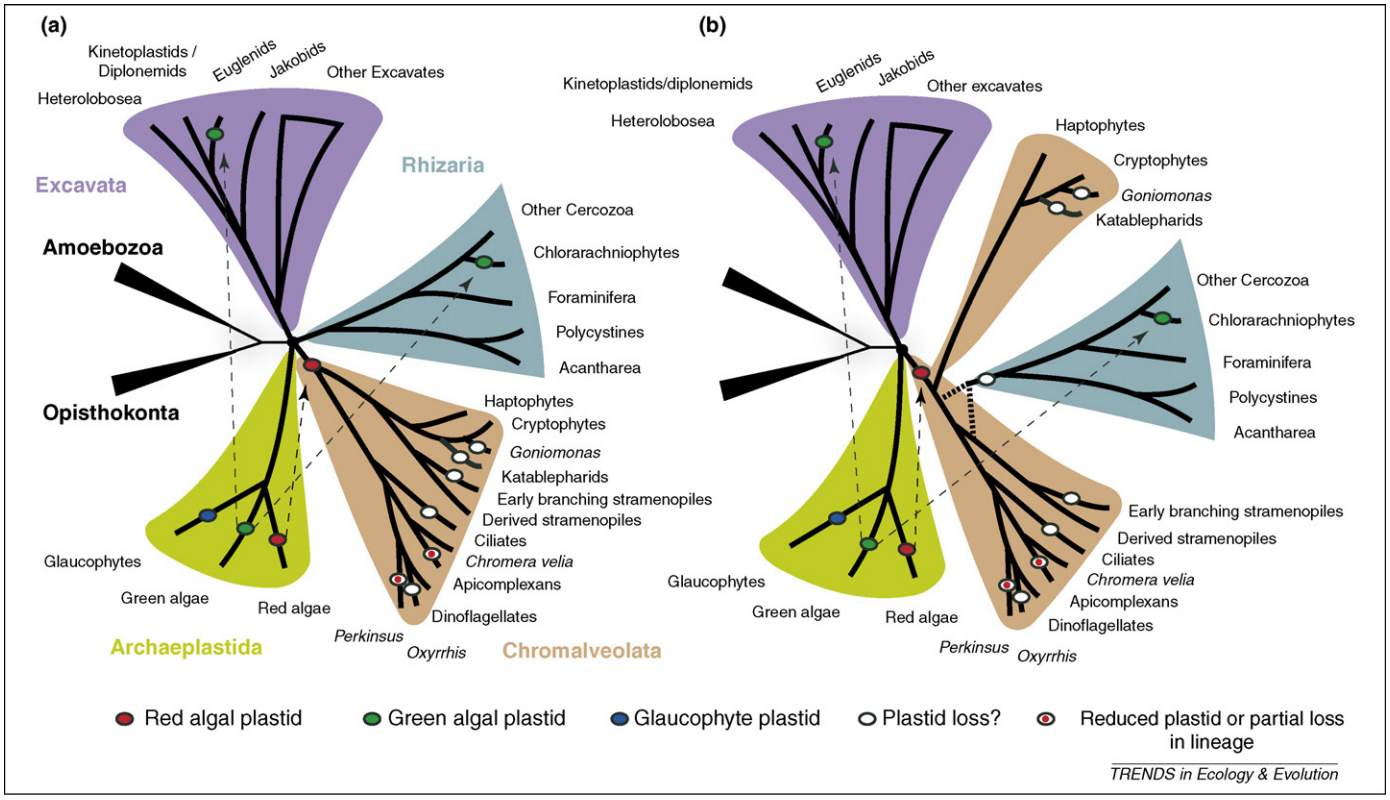

Figure 3. Two hypotheses to explain the distribution of secondary plastids, based on competing scenarios of eukaryotic evolution. A green algal-derived secondary plastid has been acquired by two separate lineages, in independent endosymbiotic events (thin dashed lines). **(a)** A single red algal endosymbiosis occurred in the common ancestor of Chromalveolata, necessitating multiple plastid losses at the base of the various nonphotosynthetic lineages. **(b)** If Rhizaria evolved from within chromalveolates, it is most parsimonious to assume that the red algal secondary plastid was lost before the diversification of this lineage. A green algal secondary plastid has been acquired by chlorarachniophytes more recently.

would have to have been acquired before the split between the haptophyte + cryptophyte clade from alveolates + -stramenopiles + Rhizaria (Figure 3b). Plastids would then presumably have been lost, independently, in Rhizaria, some stramenopiles, ciliates, early diverging dinoflagellates (e.g. *Oxyrrhis*) and many or most apicomplexans [39]. Finally, the subsequent uptake of a green algal endosymbiont in the ancestor of chlorarachniophytes would produce the distribution of plastids observed today (Figure 3b). Like the original chromalveolate hypothesis (Box 2), this scenario would require that plastid loss be far more common than gain. Although the prevalence of plastid loss (as opposed to loss of photosynthesis) among eukaryotes is unknown, the nuclear genomes of two *Phytophthora* species [29] (stramenopiles) and the apicomplexan *Cryptosporidium* [41] encode plastid-derived genes, despite these organisms lacking plastids, an indication of at least two instances of plastid loss in the ancestors of these different organisms. Additionally, the recently discovered photosynthetic eukaryote *Chomera velia* [42], which is closely related to apicomplexans, strongly indicates a shared photosynthetic ancestor of both Apicomplexa and dinoflagellates and subsequent loss in the plastid-lacking members of these groups.

If the new position of Rhizaria as a part of Chromalveolata reflects the true evolutionary history of this lineage, one would predict that genes of red algal ancestry might persist in the nuclear genomes of this group as remnants of the red algal genomes that were present in the rhizarian common ancestor. Interestingly, red algal-derived plastid genes were discovered in the nuclear genome of the green algal plastid-containing rhizarian *Bigelowiella natans* [43], and were interpreted as having been acquired by lateral gene transfer rather than vertically inherited from a red algal plastid-containing ancestor. A complete genome sequence for *B. natans* will soon be available (http://www.jgi.doe.gov) and will make it possible to test whether or not this red algal 'footprint' is (at least in part) the result of ancient endosymbiotic gene transfer. However, most Rhizaria are recalcitrant to laboratory experimentation, and significant amounts of sequence data from diverse members of this lineage will be slow in coming. At any rate, if analyses eventually show that two (or more) distinct plastids were harbored by the ancestors of extant organisms, as has been previously shown in some dinoflagellates (see Ref. [37]), then determining the organismal history of such eukaryotes might be even more difficult than currently appreciated.

Phylogenetic hope in light of EGT?

Although we have focused on chromalveolates and ignored the potentially significant role of lateral gene transfer in eukaryotic evolution (e.g. Ref. [44]), the reality of EGT and its phylogenetic implications can be extended to many of the eukaryotic supergroups. The relationships within and between chromalveolate and rhizarian taxa are not only important for understanding a major component of the tree of life but also for understanding organelle evolution and

the potential complications of EGT and EGR. On the surface, it would appear that taxonomic resolution at the deepest levels of eukaryotic divergence is possible if enough data are used. However, as biologists attempt to resolve deeper relationships within and between eukaryotic supergroups, phylogenetic artifacts are confounded by the significant divergence times between the lineages under study. Cases of gene transfers and replacements additionally complicate matters, and although phylogenomic data sets are routinely screened for genes with prokaryotic affinities, the enormous potential for eukaryotic EGRs in lineages that harbor (or have harbored) secondary endosymbionts is usually overlooked.

Recent methodological advances have made it possible to systematically and efficiently examine the phylogenetic signal of separate genes in a genome or in multigene data sets (Box 3). For example, the program Concaterpillar [45] separates genes in a multilocus data set based on pairwise comparisons of congruence (Box 3), potentially identifying instances of EGT and EGR. In a test case, when applied to a eukaryote-wide 60-gene data set of conserved translational proteins, Concaterpillar identified three data sets of 35, 15 and 10 genes, with the 15-gene data set recovering a sister relationship between stramenopiles and the red alga *Porphyra*, as would be predicted in the case of EGR [45]. If this pattern holds true on a larger scale, then the impact of EGT and EGR in lineages with photosynthetic ancestry might be far greater than currently appreciated, and empirical studies will be essential to determine the extent to which foreign genes have invaded the nuclear genomes of these organisms. Awareness of the extent of this phenomenon, however, will stimulate researchers to account for it.

Additionally, the constraints of taxon sampling, which have hampered in-depth exploration of EGT, are becoming less of a problem as new data come online. More red algal genome sequences will soon provide the data to better explore chromalveolate genomes for transferred genes, and a sharp increase in the number of chromalveolate nuclear sequences is making pairwise genomic comparisons realistic. Ironically, although their potential to introduce noise in phylogenetic studies is enormous, instances of EGR and EGT can be used as derived characters in support of common ancestry between two lineages (e.g. Refs [23,24,40]). It is imperative that the phylogenetics community continues to improve methods for the detection of foreign genes residing in nuclear genomes if resolution of ancient eukaryotic evolutionary divergences is to be achieved.

Box 3. Emerging EGT and EGR detection methods

Until recently, the only way to detect endosymbiotic gene transfers (EGTs) and replacements (EGRs) was to make individual gene trees for every gene of interest and manually compare their topologies. With modern data sets often exceeding 100 protein sequences per taxon, arranged end to end (concatenated) as one large sequence, this process can be extremely time consuming and laborious. For this reason, researchers have often had to rely on the assumption that the 'true' signal will overwhelm any noise in the data caused by the occasional protein with a discordant phylogenetic history. Fortunately, new methods have begun to emerge in response to increasing recognition that EGT and EGR are important and that there is a critical need to detect and understand conflicting signals in data sets.

PhyloSort [16] is an automated method for investigating genome-scale data sets on a gene-by-gene basis for anomalous phylogenetic patterns. The program can be used to screen the output of phylogenomics 'pipelines' such as PhyloGenie [60] for genes that show a phylogenetic pattern congruent with a user-defined monophyletic group. By searching a collection of single-gene input trees for a relationship of interest, users can rapidly focus on only those genes that support a particular phylogenetic hypothesis.

The program Concaterpillar is a likelihood ratio-based method designed to test for congruence among loci in concatenated data sets using hierarchical clustering [45]. Concaterpillar performs pairwise likelihood ratio comparisons for all loci in an iterative fashion, combining the proteins with the lowest likelihood ratio at each step and considering them as a single locus for subsequent steps. In this way, the program builds data sets of proteins based on the similarity of their phylogenetic signal until all loci have been considered or a comparison exceeds the user-defined likelihood ratio α level. The resulting data sets can then be analyzed independently, giving the user the potential to tease apart the various competing phylogenetic signals in the data, rather than settling for a tree based on a mix of conflicting information. In sum, phylogenetic tools such as Concaterpillar and PhyloSort are paving the way for researchers to systematically address questions of EGT/EGR using different approaches and types of data.

Acknowledgements

We thank A. Simpson, E. Kim and G. Sanchez Perez for helpful comments on this manuscript and the useful input of four anonymous reviewers. We apologize to those whose important work could not be cited owing to space constraints. Research in the Archibald laboratory is supported by the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada, the Dalhousie Medical Research Foundation and the Canadian Institute for Advanced Research (CIFAR). J.M.A. is a Scholar of the CIFAR Program in Integrated Microbial Biodiversity.

References

1 McInerney, J.O. *et al.* (2008) The prokaryotic tree of life: past, present...and future? *Trends Ecol. Evol.* 23, 276–281
2 Margulis, L. and Schwartz, K. (1988) *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth,* W.H. Freeman
3 Whittaker, R.H. (1969) New concepts of kingdoms of organisms. *Science* 163, 150–160
4 Embley, T.M. and Martin, W. (2006) Eukaryotic evolution, changes and challenges. *Nature* 440, 623–630
5 Simpson, A.G. and Roger, A.J. (2004) The real 'kingdoms' of eukaryotes. *Curr. Biol.* 14, R693–R696
6 Keeling, P.J. *et al.* (2005) The tree of eukaryotes. *Trends Ecol. Evol.* 20, 670–676
7 Wegener Parfrey, L. *et al.* (2006) Evaluating support for the current classification of eukaryotic diversity. *PLoS Genet.* 2, e220
8 Martin, W. *et al.* (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12246–12251
9 Archibald, A.M. and Keeling, P.J. (2002) Recycled plastids: a 'green movement' in eukaryotic evolution. *Trends Genet.* 18, 577–584
10 Bhattacharya, D. *et al.* (2007) How do endosymbionts become organelles? Understanding early events in plastid evolution. *BioEssays* 29, 1239–1246
11 Deschamps, P. *et al.* (2008) Metabolic symbiosis and the birth of the plant kingdom. *Mol. Biol. Evol.* 25, 536–548
12 Martin, W. *et al.* (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393, 162–165
13 Stoebe, B. and Kowallik, K.V. (1999) Gene-cluster analysis in chloroplast genomics. *Trends Genet.* 15, 344–347
14 Sato, N. *et al.* (2005) Mass identification of chloroplast proteins of endosymbiotic origin by phylogenetic profiling based on organism-optimized homologous protein groups. *Genome Informat.* 16, 56–68

15 Richly, E. and Leister, D. (2004) An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice. *Gene* 329, 11–16
16 Moustafa, A. and Bhattacharya, D. (2008) PhyloSort: a user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to algal nuclear genomes. *BMC Evol. Biol.*, DOI: 10.1186/1471-2148-1188-1186 (http://www.biomedcentral.com/bmcevolbiol)
17 Reyes-Prieto, A. *et al.* (2006) Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr. Biol.* 16, 2320–2325
18 Brinkmann, H. and Martin, W. (1996) Higher-plant chloroplast and cystolic 3-phosphoglycerate kinases: a case of endosymbiotic gene replacement. *Plant Mol. Biol.* 30, 65–75
19 Cavalier-Smith, T. (2003) Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 109–134
20 Takahashi, F. *et al.* (2007) Origins of the secondary plastids of euglenophyta and chlorarachniophyta as revealed by an analysis of the plastid-targeting, nuclear-encoded gene *psbO*. *J. Phycol.* 43, 1302–1309
21 Cavalier-Smith, T. (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa. *Int. J. Syst. Evol. Micro.* 52, 297–354
22 Rogers, M.B. *et al.* (2007) The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol. Biol. Evol.* 24, 54–62
23 Harper, J.T. and Keeling, P.J. (2003) Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol. Biol. Evol.* 20, 1730–1735
24 Patron, N.J. *et al.* (2004) Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot. Cell* 3, 1169–1175
25 Li, S.L. *et al.* (2006) Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. *Mol. Biol. Evol.* 23, 663–674
26 Nosenko, T. and Bhattacharya, D. (2007) Horizontal gene transfer in chromalveolates. *BMC Evol. Biol.* 7, 173
27 Armbrust, E.V. *et al.* (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306, 79–86
28 Ahmadinejad, N. *et al.* (2007) Genome history in the symbiotic hybrid *Euglena gracilis*. *Gene* 402, 35–39
29 Tyler, B.M. *et al.* (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313, 1261–1266
30 Matsuzaki, M. *et al.* (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428, 653
31 Yoon, H.S. *et al.* (2006) Defining the major lineages of red algae (Rhodophyta). *J. Phycol.* 42, 482–492
32 Rodríguez-Ezpeleta, N. *et al.* (2007) Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst. Biol.* 56, 389–399
33 Stibitz, T.B. *et al.* (2000) Symbiotic origin of a novel actin gene in the cryptophyte *Pyrenomonas helgolandii*. *Mol. Biol. Evol.* 17, 1731–1738
34 Tanifuji, G. *et al.* (2006) Diversity of secondary endosymbiont-derived actin-coding genes in cryptomonads and their evolutionary implications. *J. Plant Res.* 119, 205–215
35 Hackett, J.D. *et al.* (2007) Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of Rhizaria with chromalveolates. *Mol. Biol. Evol.* 24, 1702–1713
36 Palmer, J.D. (2003) The symbiotic birth and spread of plastids: how many times and whodunit? *J. Phycol.* 39, 4–11
37 Archibald, J.M. (2005) Jumping genes and shrinking genomes – probing the evolution of eukaryotic photosynthesis using genomics. *IUBMB Life* 57, 539–547
38 Patron, N.J. *et al.* (2007) Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr. Biol.* 17, 887–891
39 Burki, F. *et al.* (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* 2, e790
40 Rice, D.W. and Palmer, J.D. (2006) An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol.* 4, 31
41 Huang, J. *et al.* (2004) Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biol.* 5, R88
42 Moore, R.B. *et al.* (2008) A photosynthetic alveolate closely related to apicomplexans parasites. *Nature* 451, 959–963
43 Archibald, J.M. *et al.* (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7678–7683
44 Rogers, M.B. *et al.* (2007) A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer. *BMC Evol. Biol.* 7, 89
45 Leigh, J.W. *et al.* (2008) Testing congruence in phylogenomic analysis. *Syst. Biol.* 57, 4–15
46 Nozaki, H. and Iseki, M. (2007) Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Mol. Biol. Evol.* 24, 1592–1595
47 Yoon, H.S. *et al.* (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21, 809–818
48 Larkum, A.W.D. *et al.* (2007) Shopping for plastids. *Trends Plant Sci.* 12, 189–195
49 Stiller, J.W. (2007) Plastid endosymbiosis, genome evolution and the origin of green plants. *Trends Plant Sci.* 12, 391–396
50 Stiller, J.W. *et al.* (2003) A single origin of plastids revisited: convergent evolution in organellar genome content. *J. Phycol.* 39, 95–105
51 Cavalier-Smith, T. (1999) Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* 46, 347–366
52 Archibald, J.M. (2007) Nucleomorph genomes: structure, function, origin and evolution. *BioEssays* 29, 1239–1246
53 Yoon, H.S. *et al.* (2005) Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol. Biol. Evol.* 22, 1299–1308
54 Rodríguez-Ezpeleta, N. *et al.* (2007) Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr. Biol.* 17, 1420–1425
55 Bodyl, A. (2005) Do plastid-related characters support the chromalveolate hypothesis? *J. Phycol.* 41, 712–719
56 Bodyl, A. and Moszczynski, K. (2006) Did the peridinin plastid evolve through tertiary endosymbiosis? A hypothesis. *Eur. J. Phycol.* 41, 435–448
57 Bodyl, A. *et al.* (2007) The intracellular cyanobacteria of *Paulinelia chromatophora*: endosymbionts or organelles? *Trends Microbiol.* 15, 295–296
58 Huang, J. *et al.* (2004) A first glimpse into the pattern and scale of gene transfer in Apicomplexa. *Int. J. Parasitol.* 34, 265–274
59 Teles-Grilo, M.L. *et al.* (2007) Is there a plastid in *Perkinsus atlanticus* (Phylum Perkinsozoa)? *Eur. J. Protistol.* 43, 163–167
60 Frickey, T. and Lupas, A.N. (2004) PhyloGenie: automated phylome generation and analysis. *Nucleic Acids Res.* 32, 5231–5238

Reproduction of material from Elsevier articles

Interested in reproducing part or all of an article published by Elsevier, or one of our article figures? If so, please contact our *Global Rights Department* with details of how and where the requested material will be used. To submit a permission request online, please visit:

www.elsevier.com/locate/permissions