REVIEW

After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl c

Beverley R. Green

Abstract The chromalveolate hypothesis proposed by Cavalier-Smith (J Euk Microbiol 46:347–366, 1999) suggested that all the algae with chlorophyll c (heterokonts, haptophytes, cryptophytes, and dinoflagellates), as well as the ciliates, apicomplexans, oomycetes, and other non-photosynthetic relatives, shared a common ancestor that acquired a chloroplast by secondary endosymbiosis of a red alga. Much of the evidence from plastid and nuclear genomes supports a red algal origin for plastids of the photosynthetic lineages, but the number of secondary endosymbioses and the number of plastid losses have not been resolved. The issue is complicated by the fact that nuclear genomes are mosaics of genes acquired over a very long time period, not only by vertical descent but also by endosymbiotic and horizontal gene transfer. Phylogenomic analysis of the available whole-genome data has suggested major alterations to our view of eukaryotic evolution, and given rise to alternative models. The next few years may see even more changes once a more representative collection of sequenced genomes becomes available.

Keywords Algae · Chloroplasts · Chlorophyll c · Chromalveolate hypothesis · Molecular evolution · Phylogeny · Plastids · Secondary endosymbiosis

B. R. Green (✉)
Botany Department, University of British Columbia, Vancouver, B.C V6T 1Z4, Canada
e-mail: brgreen@interchange.ubc.ca

Abbreviations
EGT   Endosymbiotic gene transfer
EST   Expressed sequence tag (cDNA sequence)
HGT   Lateral or horizontal gene transfer
LHC   Light-harvesting complex

Introduction

It is now generally accepted that the chloroplasts of all photosynthetic eukaryotes had a common ancestor: a cyanobacterial endosymbiont that became completely integrated into the cell of its host, a heterotrophic eukaryote. This primary endosymbiosis gave rise to three photosynthetic lineages: the green line (green algae and plants), the red line (rhodophyte algae), and the glaucophyte algae (Fig. 1). This concensus is based on the highly conserved nature of the photosynthetic apparatus, sequence similarities between the 100–250 genes remaining on chloroplast genomes and their cyanobacterial homologs, and most multi-gene phylogenetic trees of nuclear-encoded proteins (e.g. Rodriguez-Ezpeleta et al. 2005). However, the three lineages diverged so long ago that the order in which they branched off is still not resolved (Deschamps and Moreira 2009; Inagaki et al. 2009).

In addition to the three major lineages of photosynthetic eukaryotes with "primary" chloroplasts, there are many algal groups with more complex evolutionary histories. These include all the algae with chlorophyll c, which make up a large part of the marine photosynthetic biota. They are distinguished by having Chl a/c light-harvesting antenna complexes (LHCs), in contrast to the Chl a/b LHCs of the green lineage and the Chl a LHCs of the red lineage. In contrast to plant chloroplasts, there is no differentiation
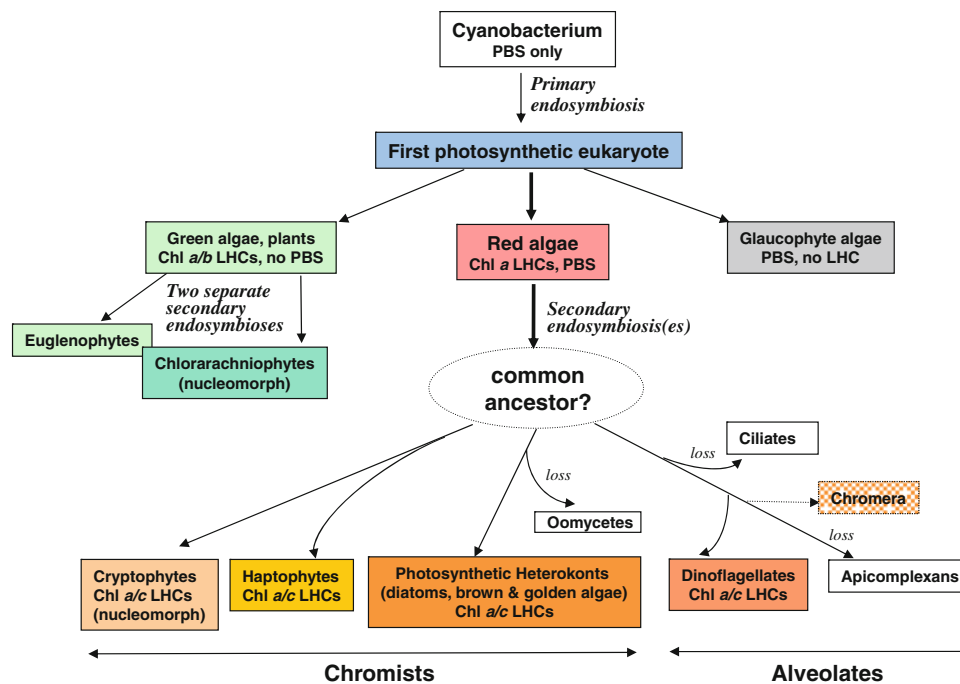
Fig. 1 Overview of the major eukaryotic algal groups and their non-photosynthetic relatives. LHCs (membrane-intrinsic light-harvesting complexes) and PBS (phycobilisomes) are the major pigment-proteins traditionally used to classify algal groups. The cryptophytes and dinoflagellates have additional unique antenna complexes in the thylakoid lumen: phycobiliprotein tetramers in cryptophytes and the peridinin–Chl *a* complex in dinoflagellates (Green 2004). According to the chromalveolate hypothesis (Cavalier-Smith 1999) a single secondary endosymbiosis ("common ancestor", *dotted oval*) gave rise to all the algae with Chl *c* and many of their (now) non-photosynthetic descendants, such as oomycetes, ciliates, and apicomplexans (indicated by "loss"). There is considerable support for the alveolates (dinoflagellates, apicomplexans, ciliates) being a monophyletic group, but there is no similar consensus for the chromists (cryptophytes, haptophytes, heterokonts). Note that all the chromalveolate branches include non-photosynthetic taxa, although only the major ones are diagrammed here. Two completely separate secondary endosymbiotic events involving green algal endosymbionts gave rise to the Euglenophytes and Chlorarachniophytes

into stacks of thylakoids (grana) interconnected by single thylakoids, and the thylakoid membranes are simply appressed along most of their length, mostly in groups of two or three (reviewed in Green et al. 2003). This review addresses the complicated evolutionary relationships of the algae with Chl c and their non-photosynthetic relatives. More information can be found in several excellent reviews by Archibald (2009), Elias and Archibald (2009), Keeling (2009, 2010), and Lim and McFadden (2010).

The algae with Chl c belong to four phylogenetically distinct groups: the heterokonts (stramenopiles), the haptophytes, the cryptophytes, and the dinoflagellates (Fig. 1, Table 1). The first three groups are collectively termed the Chromista (Cavalier-Smith 1981). The heterokonts include a great diversity of algal groups, ranging in size from nanoplankton to multicellular seaweeds many meters in length. Their name comes from the fact that they have two unequal flagella, one of which carries distinctive tripartite hairs. Both the heterokonts and the haptophytes use fucoxanthin and its derivatives as their major photosynthetic carotenoids, but the haptophytes are distinguished from the heterokonts on the basis of flagellar structure and the possession of a haptonema, a microtubule-based projection which may be involved in obtaining prey. One group of haptophytes is covered with calcium carbonate scales (coccoliths) that reflect enough sunlight during large blooms to influence climate. The cryptophytes have different carotenoids and are the only algal group with distinctive phycobilin tetramers in the thylakoid lumen. Their most distinguishing characteristic is the possession of a miniature second nucleus (nucleomorph) located between the inner and outer pair of membranes surrounding the chloroplast.

The dinoflagellates have a variety of morphological features that separate them from the other algae with Chl c. Some of these features, in particular the possession of cortical alveoli underneath the plasma membrane, show that the dinoflagellates are actually related to the ciliates and apicomplexans rather than to the chromists. Phylogenetic trees of ribosomal RNA genes (Gajadhar et al. 1991) were the first to support the idea that dinoflagellates, ciliates, and apicomplexans are indeed each others closest relatives (now termed the Alveolata), and the monophyly of the Alveolata has been confirmed in many subsequent

Table 1 The Chromalveolates

	Characteristics	Examples
Chromists		
Heterokonts-photosynthetic	Algae with two unequal flagella, one with tripartite hairs, at some stage of life cycle. Chl *a/c* LHCs with fucoxanthin as major carotenoid. "Seaweeds" are multicellular	Diatoms (*Thalassiosira, Phaeodactylum*), brown and chrysophyte algae (*Fucus, Ectocarpus, Heterosigma*), pelagophytes (*Aureococcus*)
Heterokonts-heterotrophic	Common plant pathogens. No plastid. Unique sexual reproduction	Oomycetes (*Phytophthora*)
Haptophytes	Unicellular algae with two flagella and a haptonema. Chl *a/c* LHCs with fucoxanthin and its derivatives as major carotenoids. Some armored with $CaCO_3$ plates (coccoliths)	*Emiliania, Pavlova, Phaeocystis*
Cryptophytes	Unicellular algae distiguished by a nucleomorph (relict nucleus) in periplastidal space and unique phycobiliproteins in thylakoid lumen. Chl *a/c* LHCs with alloxanthin and other carotenoids.	*Guillardia, Rhodomonas, Cryptomonas, Hemiselmis*
Alveolates		
Dinoflagellates	Unicellular. Phototrophs have Chl *a/c* LHCs with peridinin as major carotenoid and a unique lumenal peridinin-Chl *a* antenna. About 50% are heterotrophs. Others have non-peridinin plastids with various pigments, acquired by tertiary endosymbiosis	*Heterocapsa, Amphidinium, Symbiodinium*
Apicomplexans	Non-photosynthetic intracellular parasites distinguished by microtubular apical complex. Some have a relict plastid (apicoplast)	*Plasmodium, Toxoplasma, Cryptosporidium*
Ciliates	Non-photosynthetic, mostly phagotrophic protists with multiple rows of cilia and two types of nuclei. No evidence for a relict plastid	*Paramecium, Tetrahymena*

More detail and many excellent illustrations can be found in Graham and Wilcox (2000)

multi-gene analyses (Rodriguez-Ezpeleta et al. 2007; Burki et al. 2008). No ciliates are genuinely photosynthetic, although there are a few ciliates that can acquire chloroplasts from their prey (kleptoplasty) and one that has an intact green algal symbiont (Johnson 2010). The apicomplexans are non-photosynthetic intracellular parasites that are responsible for serious human and animal diseases such as malaria (caused by *Plasmodium* species), toxoplasmosis, and African cattle diseases. In contrast to the ciliates, many apicomplexans retain a relict plastid with a reduced genome encoding a small number of genes for non-photosynthetic functions (Lim and McFadden 2010).

A newly discovered organism has recently provided a photosynthetic link between the dinoflagellates and the apicomplexans (Moore et al. 2008). *Chromera velia* is a small golden-brown alga that lives as a coral symbiont but can be cultured in the laboratory. Its plastids are surrounded by four membranes and the thylakoids are appressed in threes. Phylogenetic trees of nuclear and plastid rRNA show conclusively that this organism is a basal apicomplexan, supporting the idea that dinoflagellates and apicomplexans had a common photosynthetic ancestor (Moore et al. 2008; Janouškovec et al. 2010).

Secondary endosymbiosis

The concept of secondary endosymbiosis underlies all current thinking about the evolution of the eukaryotic supergroups. During the 1960s, Sarah Gibbs did an extensive ultrastructural investigation of all the major groups of algae, that showed that the plastids of algae with Chl *c* (then referred to as chromophytes) were surrounded by either three or four membranes rather than the two membranes of plant and green algal plastids (Gibbs 1970). She also noted that plastids of the green alga *Euglena* had three membranes. This eventually led her to propose that these algae were the result of secondary endosymbiosis, in which a non-photosynthetic eukaryote acquired a plastid by engulfing an alga that already had a primary plastid (Gibbs 1978, 1981). The outermost of the two extra membranes surrounding the plastids of crytophyte, haptophyte, and heterokont algae would be the remains of the endocytotic vacuole membrane of the host, and the innermost would have been derived from the plasma membrane of the engulfed endosymbiont. The plastids of *Euglena* and the dinoflagellates could have lost one membrane, probably the host's vacuolar membrane, leaving them free in the host cytoplasm. Based on

pigmentation, *Euglena* would not have obtained its plastid from the same source(s) as the chromophytes, but from a green alga. An elegantly written account of the insight that led to the secondary endosymbiosis theory can be found in Gibbs (2006).

The strongest piece of evidence in favor of the concept of secondary endosymbiosis was the existence in the cryptophyte algae of the second nucleus (nucleomorph) in between the chloroplast envelope (two innermost membranes) and the two outermost membranes (Gibbs 1978, 1981; Gillott and Gibbs 1980; Ludwig and Gibbs 1985). The periplastidal space between the two pairs of membranes also had its own unique ribosomes that resembled cytoplasmic rather than plastid ribosomes (Gillott and Gibbs 1980). All the cryptophyte nucleomorphs examined so far have three miniature chromosomes (Lane et al. 2006; Archibald 2007). The genomes of two of them have now been sequenced (Douglas et al. 2001; Lane et al. 2007) and phylogenetic analysis supports the idea that the nucleomorph represents the remains of a red algal nucleus. However, these analyses do not establish what sort of red alga it belonged to, nor whether the chromists originated from a single secondary endosymbiosis or several different ones involving different red algae and/or different heterotrophic hosts.

The chlorarachniophytes are a completely separate group of algae with nucleomorphs, whose chloroplasts appear to have been obtained from a green algal endosymbiont (Ludwig and Gibbs 1989; Rogers et al. 2007a). Their initial discovery supported the idea that many algae may have acquired their chloroplasts from eukaryotic endosymbionts (Ludwig and Gibbs 1989). The nucleomorph of one of them, *Bigellowiella natans*, has been sequenced, and it has three miniature chromosomes like those of the cryptophyte nucleomorphs (Gilson et al. 2006). However, it has retained a different set of genes, which are clearly related to nuclear genes of green algae, and there is no doubt that a different secondary endosymbiosis was involved.

The chromalveolate hypothesis

In the 1990s it was discovered that the apicomplexan *Plasmodium* had two organellar genomes, one of which was a small circular plastid genome with rRNAs, tRNAs, and genes for several cyanobacterial-type RNA polymerase subunits and a few other housekeeping proteins (Wilson et al. 1996). It had a rather divergent set of ribosomal protein genes, but they were organized in the canonical bacterial operon order, as in all other plastids (Stoebe and Kowallik 1999). All the genes required for photosynthesis were missing. Similar plastid-like genomes were found in other apicomplexans, and they were localized to a small membrane-bound body near the nucleus that had puzzled

microscopists for years (McFadden et al. 1996). Although these plastids (now called apicoplasts) cannot fix carbon or convert light energy into other forms, at some stages of the parasite's life cycle they provide a site for at least three important cellular functions: fatty acid synthesis, isoprenoid synthesis, and Fe–S cluster assembly, all of which are potential drug targets (Wilson 2004; Lim and McFadden 2010). The heme biosynthesis pathway, which is almost exclusively chloroplastic in higher plants, is divided among three cellular compartments in apicomplexans: the cytosol, the mitochondrion and the apicoplast (Sato et al. 2004; Ralph et al. 2004).

The chromalveolate hypothesis proposed by Cavalier-Smith (1999) suggested that the chromists and the alveolates had a common photosynthetic ancestor, and that they are all descended from a single secondary endosymbiotic relationship involving a red algal endosymbiont (Fig. 1). The rationale was based on the parsimony principle: that establishing a stable endosymbiotic relationship with all the required control systems and the mechanisms for transporting proteins, lipids, and carbohydrates back and forth across four organellar membranes was too unlikely to have happened multiple times. In the same article, Cavalier-Smith also proposed that a second secondary endosymbiosis involving a green algal endosymbiont gave rise to both the photosynthetic euglenids and the chlorarachniophytes, but molecular phylogenetic analysis of "host" nuclear genes soon showed that there must have been two separate events involving different green algal endosymbionts (Cavalier-Smith 2002). The sequence of a chlorarachniophyte plastid genome confirmed that it was derived from a different group of green algae than the *Euglena* plastid (Rogers et al. 2007a).

One difficulty with the concept of a common chromalveolate ancestor is that there is an enormous diversity in morphology and lifestyle among its putative descendants (Table 1). Many of them are heterotrophs. Ciliates show no sign of a relict plastid. The same is true of the oomycetes (e.g. the plant pathogen *Phytophthora*) which are the closest relatives to the photosynthetic heterokonts. Many dinoflagellates and some cryptophytes are not photosynthetic. The only way this could be explained is if the chromalveolate ancestor had a plastid but many of its descendants subsequently lost their plastids and all trace of having had one.

It should be pointed out that many happily photosynthesizing algae, as well as their heterotrophic relatives, are able to supplement their nutritional income by various types of mixotrophy, i.e. absorbing dissolved organic compounds (osmotrophy) or engulfing bacteria or other protists (phagotrophy). Phagotrophy also provides a means for the acquisition of new or replacement plastids (Johnson 2010). Dinoflagellates are particularly good at this: various species have replaced or augmented their peridinin-type

plastids with new ones acquired by tertiary endosymbiosis from diatoms, haptophytes, cryptophytes, cyanobacteria, and a green alga (Hackett et al. 2004; Archibald 2009). Given that losses and acquisitions appear to have been common, do we really need to propose a common ancestor for all the chromalveolate plastids (Fig. 1)? To answer this question, we need to look at the plastid-encoded genes, the nuclear-encoded genes for plastid proteins acquired along with the plastid, as well as the other nuclear genes which are presumably representative of the "host" genomes.

What plastid genomes tell us

If all the chromalveolate plastids were acquired from a red algal endosymbiont, or even from different red algal endosymbionts, it should be possible to determine the relationships among them by phylogenetic analysis of their chloroplast genes. Complete chloroplast genome sequences are now available for five red algae, three diatoms and several other heterokonts, three cryptophytes and one haptophyte (NCBI Organelle Genome Resources 2010). In current practice, the sequences of all the genes or proteins of a plastid genome are concatenated into one long sequence, and these concatenated sequences are used for phylogenetic analysis. This is more rigorous than relying on gene trees of a small number of proteins (e.g. Yoon et al. 2002, 2004) because differences in evolutionary rates between proteins, or between different sites in a protein, are more likely to be cancelled out. In some studies, the fastest evolving sites or proteins are eliminated from analysis since they are the most likely to show long branches, or to give misleading results due to mutational saturation (Iida et al. 2007; Khan et al. 2007; Sanchez-Puerta et al. 2007).

Chloroplast multi-gene trees support the red algal origin of the chromist plastid, probably from within the more advanced groups of red algae (Bachvaroff et al. 2005; Rodriguez-Ezpeleta et al. 2005; Khan et al. 2007; Sanchez-Puerta et al. 2007; Wang et al. 2008; Le Corguillé et al. 2009; Janouškovec et al. 2010). The monophyly of chromist plastids is also usually supported, but the branching order of the three chromist groups depends on the methods used and the corrections applied to avoid problems due to rapidly evolving sequences and compositional bias (Iida et al. 2007; Sanchez-Puerta et al. 2007). Some trees show the haptophytes and heterokonts as closest relatives, with cryptophytes basal (Yoon et al. 2004), whereas others group the haptophytes and cryptophytes together, with heterokonts at the base (Iida et al. 2007; Khan et al. 2007; Sanchez-Puerta et al. 2007; Le Corguillé et al. 2009). The sister relationship of haptophytes and cryptophytes is supported by a rare horizontal transfer of a bacterial 50S

ribosomal protein gene (rpl36) gene into the plastid genomes of these two groups, to the exclusion of all other photosynthetic eukaryotes (Rice and Palmer 2006). It is also supported by some nuclear gene trees (see below).

The situation is not so clear if dinoflagellate plastid sequences are included. The dinoflagellate plastid genomes have been shattered into a collection of minicircular chromosomes, carrying between one and four genes, and many of the usual chloroplast genes have been either lost or transferred to the nucleus (Zhang et al. 1999). The 17–20 genes that remain are those that encode ribosomal RNAs, a few tRNAs and the core proteins of the photosynthetic apparatus (Green 2004; Nelson et al. 2007; Howe et al. 2008). Even these conserved proteins have diverged so much that they form long branches on phylogenetic trees (Zhang et al. 1999, 2000; Bachvaroff et al. 2005; Sanchez-Puerta et al. 2007; Wang et al. 2008). Divergent or rapidly evolving sequences tend to cluster with each other on phylogenetic trees whether or not they are the most closely related, and do little to establish the relationship between dinoflagellate plastids and those of any other group. This is the "long-branch attraction artefact", one of the major causes of misleading results in phylogenetic analysis (Felsenstein 1978). Another complication is compositional bias, which can have a serious effect on single-gene trees even when dinoflagellates are excluded (Iida et al. 2007; Sanchez-Puerta et al. 2007).

According to the chromalveolate hypothesis, the dinoflagellates and apicomplexans should have shared a common photosynthetic ancestor (Fig. 1). The *Plasmodium* apicoplast genome is almost the complement of the dinoflagellate plastid genome in terms of gene content, although it maps as one circle of about 45 kb and has not been broken up into minicircles (Wilson et al. 1996). It lacks all photosynthesis genes but has retained genes for RNA polymerase, some ribosomal proteins, and several other proteins, all of which appear to have been lost by the dinoflagellate plastid. Here again, the sequences are so divergent it is difficult to make valid phylogenetic trees, although 23S rRNA sequences do support a relationship between the dinoflagellate and apicomplexan plastids (Zhang et al. 2000). Several other apicoplast genomes have been sequenced, and their sequences are equally divergent. The best evidence for a common ancestor comes from the existence of two basal apicomplexans that still have photosynthetic plastids: *Chromera velia* and the as-yet unnamed CCMP3315 (Moore et al. 2008; Janouškovec et al. 2010). Their plastid genomes encompass all the genes found in either dinoflagellate or apicomplexan plastids, and their nuclear-encoded Type II Rubisco is clearly closely related to that of dinoflagellates. Separate phylogenetic trees including either dinoflagellate or apicoplast homologs support a red algal origin, with heterokonts as closest

relatives. This suggests that plastid genes were lost independently in the two lineages, as their plastids evolved to support different life-styles of their hosts.

In summary, there is little doubt that all the chromalveolate plastids had a red algal plastid ancestor, although whether it was acquired in one event by a common chromalveolate ancestor or in several separate secondary endosymbiotic events by different hosts cannot be resolved by looking at plastid genes. A few researchers are still considering the possibility of a green algal ancestor for the apicoplast (Lau et al. 2009), but the weight of evidence supports a red algal contribution (Lim and McFadden 2010). To find out more, we need to look at nuclear genes, particularly those that support chloroplast function.

Where did Chl *c* come from?

Since Chl *c* as accessory pigment is what unites the chromists and the dinoflagellates, the genes required for its synthesis should provide some insight into chloroplast ancestry. If the chromalveolates had a common ancestor, the ability to synthesize Chl *c* could have arisen before the lineages diverged. Alternatively, it could have arisen in one group and spread by horizontal transfer. On the other hand, if the chromists and the dinoflagellates acquired red plastids through several independent secondary endosymbiotic events, that would probably require that at least some red algae were able to make Chl *c*. However, Chl *c* has never been found in any extant red alga.

It is an astonishing fact that absolutely nothing is known about how Chls c_1 and c_2 are synthesized! On paper, all that should be needed is the introduction of a double bond into the propionyl side-chains of mono- and di-vinyl prochlorophyllide, respectively (Fig. 2). One suggestion is an oxidation at 17^1 followed by dehydration to form the double bond (Rüdiger and Grimm 2006), but this possibility has not investigated experimentally. Chl c_1

inhibits the light-dependent protochlorophyllide oxidoreductase, suggesting a reason why the porphin ring is not reduced further (Helfrich et al. 2003). The other difference with Chl *a* synthesis is that no phytyl tail is added at the last step, although in some haptophytes Chl c_2 has a monogalactosyl diacylglyerol tail (Zapata et al. 2006). To my knowledge, no Chl *c*-deficient mutants have been reported, although there are some heterokonts, e.g. the eustigmatophytes and the chrysophyte *Ochromonas,* that make only Chl *a*. This is one of those surprising gaps in scientific knowledge that suddenly appear like pot-holes in an otherwise well-studied highway. I hope this review will stimulate some work on this embarrassing gap in our knowledge of such a large fraction of the earth's photosynthetic eukaryotes.

Endosymbiotic gene transfer to the nucleus

It has been estimated that several thousand cyanobacterial genes were relocated to the nucleus of the host during the primary endosymbiosis (Martin et al. 2002). Some of them acquired targeting sequences that allowed their encoded proteins to be translocated back into the plastid to keep it functioning; others replaced or augmented the nuclear gene repertoire. This is referred to as "endosymbiotic gene transfer" (EGT), to distinguish it from "horizontal gene transfer" (HGT) which implies transfer between whole organisms rather than within one. Secondary endosymbiosis in its turn would have involved significant transfer of endosymbiont nuclear genes to the secondary host nucleus. Some of these genes would encode plastid-targeted proteins and facilitate the establishment of the endosymbiotic relationship, but others would function in other cell compartments and contribute to genomic diversity. For example, both diatom nuclear genomes have numerous genes of red algal origin, less than half of which are predicted to be plastid-targeted (Armbrust et al. 2004; Bowler et al. 2008).

Fig. 2 Proposed pathway for the synthesis of the two major Chls *c* from divinyl- and monovinylprotochlorophyllide by the formation of a double bond in the propionyl side-chain. The "17^1 oxidase" is hypothetical and may represent several linked reactions. Chl c_3 is 7-methoxycarbonyl-Chl c_2. There is no phytyl tail, although some haptophytes have Chl c_2 with a monogalactosyldiacyl glycerol side-chain (Zapata et al. 2006)

Regardless of their current role, they represent "footprints" of the endosymbiont ancestor that can be used in phylogenetic analysis.

The availability of nuclear genome sequences from several algae and a number of plants, as well as cDNA sequences from a large number of other protist lineages, has led to a rapid increase in the amount of data available to study evolutionary origins. Single-gene trees utilizing cDNA sequences supported a substantial contribution of red algal nuclear genes to the nuclear genomes of haptophytes, cryptophytes, and in some cases dinoflagellates (Hackett et al. 2004; Waller et al. 2006; Li et al. 2006). EST sequences are particularly important for dinoflagellate nuclear genes, because the dinoflagellates have enormous nuclear genomes: even bigger than the human genome (LaJeunesse et al. 2005), so we are unlikely to see a genome project any time soon. In a study using 20 nuclear-encoded proteins of the dinoflagellate *Alexandrium* that are plastid-encoded in all other groups, the grouping of dinoflagellate with red algal and chromist sequences had strong support, although the dinoflagellate branch was very long, indicating a high rate of divergence (Wang et al. 2008).

It should be pointed out that the flow of genes from the plastid to the nucleus did not stop with secondary endosymbiosis. Chromist plastid genomes have 140–180 genes, compared to 200–250 in red algal plastids, and some of the "lost" red plastid genes have been found in the nuclear genomes of the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* or the haptophyte *Emiliania huxleyii* (Oudot-Le Secq et al. 2007). Gene transfer appears to happen independently in each lineage. For example, we found several genes on the *Phaeodactylum* chloroplast genome that have been lost from the *Thalassiosira* chloroplast genome (Oudot-Le Secq et al. 2007). Another gene (*psb28*) is present on both chloroplast and nuclear genomes in *Thalassiosira*, suggesting it is in the process of being transferred (copied) to the nucleus. Therefore, plastid gene transfer is still a "work in progress", although on an evolutionary (very long) time scale.

One of the problems with the chromalveolate hypothesis is the inclusion of two large groups of non-photosynthetic organisms, the oomycetes and the ciliates. The oomycetes are definitely heterokonts, even though many of them are plant pathogens and have a fungal-like life-style. None of them has anything resembling a relict plastid. If the common ancestor of photosynthetic and non-photosynthetic heterokonts had a "red" plastid that was lost in the oomycete lineage, we might still expect to find a few red footprints in the oomycete nuclear genome. These would likely be genes that did not have anything to do with plastid function. Out of the 171 red algal genes with very high bootstrap support (>85%) in one or both of the diatom genomes, 11 were shared with oomycetes (Bowler et al.

2008), confirming an earlier report (Tyler et al. 2006). However, questions have been raised about whether the number of red algal genes is above the background that might be expected due to the occasional acquisition of genes from prey by phagocytic ancestors (Stiller et al. 2009; Elias and Archibald 2009).

In the case of ciliates, whole-genome sequences of *Paramecium* (Aury et al. 2006) and *Tetrahymena* (Eisen et al. 2006) showed no evidence that either of them ever had an endosymbiont with a plastid. A detailed phylogenomic study brought to light 16 genes most closely related to algal and plant genes, but not specifically to red algal genes (Reyes-Prieto et al. 2008). However, it has been pointed out that 16 out of more than 25,000 genes might simply represent a background level of HGT, especially since certain ciliates have been shown to have acquired bacterial genes from their environment (Elias and Archibald 2009). On the other hand, if the secondary endosymbiotic relationship was initiated shortly before the ciliates separated from the line leading to the dinoflagellates and apicomplexans, and the ciliate plastid was lost soon afterwards, it might have left few signs of its temporary residence. The same argument might be applied to the oomycetes. Fortunately, on-going studies of organisms that might represent transitional forms between photosynthetic and non-photosynthetic sister groups may provide more conclusive data (Keeling 2010).

HGT and the mosaic nature of eukaryote genomes

Horizontal gene transfer (HGT) has been an important factor in the evolution of metabolic pathways in eukaryotes, and it is becoming evident that eukaryotic genomes are mosaics (Keeling and Palmer 2008; Keeling 2009). Although the majority of nuclear genes may be acquired by vertical inheritance and EGT, in many cases there has been a substantial input from other organisms, particularly bacteria. One of the first examples of this was found in the chlorarachniophyte *Bigelowiella*, an amoeboflagellate with a green (Chl *a/b*) secondary chloroplast (Fig. 1). Analysis of cDNA sequences obtained from this organism showed that a number of them were more related to red algal or bacterial homologs rather than to the expected green algal homologs (Archibald et al. 2003). Since most chlorarachniophytes have an ameboid stage (and the one that engulfed the green algal endosymbiont undoubtedly did), it is not too surprising that over the eons the chlorarachniophytes picked up some genes from their food. "You are what you eat", as Ford Doolittle famously said in explaining HGT in prokaryotes (Doolittle 1998). The same argument can be applied to any organism that was able to acquire an endosymbiont by phagocytosis at some point in its evolutionary history.

Until recently, HGT events (as opposed to EGT events) were regarded as involving occasional transfers here and there, for example the heme biosynthesis enzyme porphobilinogen deaminase which is of proteobacterial rather than cyanobacterial origin in all photosynthetic eukaryotes (Oborník and Green 2005). However, phylogenomic analysis of the two finished diatom genomes has recently shown that they acquired a substantial number of genes from a number of different bacterial phyla (Bowler et al. 2008). Genes with cyanobacterial affinity would be expected, but there were three times as many genes of proteobacterial origin. Only 10% were shared with *Phytophthora*, suggesting that most of these acquisitions occurred after the divergence of the two lineages. More work will be necessary to determine whether other unicellular photosynthetic eukaryotes have experienced this level of HGT, and what impact (if any) it has had on the evolution of photosynthesis.

To further confuse the issue, there have been three studies suggesting Chlamydiae bacteria contributed more than a few genes to the common ancestor of the red and green lineages, and that some of them have plastid (but not photosynthetic) function (Huang and Gogarten 2007; Becker et al. 2008; Moustafa et al. 2008). It will be interesting to see if these results are confirmed when a more representative sampling of red algal and glaucophyte genome sequences become available.

A number of other factors complicate studies of the evolution of photosynthesis. Calvin cycle enzymes have very complex phylogenetic histories that include multiple isoforms, gene duplications, retargeting, HGT, and independent losses (Martin and Schnarrenberger 1997; Rogers and Keeling 2004; Grauvogel et al. 2007; Rogers et al. 2007b; Reyes-Prieto and Bhattacharya 2007; Teich et al. 2007). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is particularly complicated (Rogers et al. 2007b). Only two cases clearly support the chromalveolate hypothesis, and in neither case is there any evidence for a red algal ancestor. The phosphoribulokinase of dinoflagellates and chromists appears to have been obtained from a green alga (Petersen et al. 2006) but the chloroplast-targeted Class IIA fructose-1,6-bisphosphate aldolase, which also unites dinoflagellates and chromists, is of uncertain origin (Patron et al. 2004). A study of cyanobacteria-like genes in non-photosynthetic protists shows the difficulty of separating out HGT contributions from vertical inheritance of nuclear genes (Maruyama et al. 2009).

What is becoming clear from the research done in the last few years is that the nuclear genomes of eukaryotes are mosaics, particularly with respect to metabolic pathways (Whitaker et al. 2009). This complicates the job of untangling the evolutionary story of the major eukaryotic lineages, particularly those that have or once had chloroplasts.

I have chosen not to discuss the many pitfalls in the construction of phylogenetic trees, but interested readers should consult the writings of H. Philippe, Y. Inagaki and A.J. Roger for insights into these problems (e.g. Delsuc et al. 2005; Hampl et al. 2009; Inagaki et al. 2009 and references therein). Fortunately, as increasing amounts of whole-genome data become available, methods for analyzing large amounts of complex data are also increasing in number and sophistication.

Phylogenomics approaches and the tree of life

Large-scale data mining and analysis of the massive amounts of sequence data now obtainable is referred to as "phylogenomics". At the simplest level, sequences from the organism of interest (e.g. the translated EST sequences or all the protein models of a genome) are used as "bait" to search all available genomes and sequence databases to obtain all the sequences that have at least a certain similarity score and a significant level of overlap with the query sequence. Protein sequences are used, because they are much more conserved than DNA sequences. Sometimes, sequences of all the genomes are blasted against all the others to select a collection of reciprocal best hits. The results are then tabulated as a crude measure of genes shared with several other organisms or groups of organisms, usually shown as a Venn diagram (e.g. Armbrust et al. 2004; Bowler et al. 2008). The results may be divided up further according to functional categories or shared domains. This "first pass" through a mass of data is found in most eukaryote genome papers, and serves as starting point for more in-depth analyses.

At the next level of phylogenomic analysis, each collection of putative homologs from the initial database search is filtered to try to ensure that the sequences are true orthologs rather than paralogs, and those that pass are subjected to automated phylogenetic tree construction. The output from this step undergoes another bioinformatic filtering step to determine the number of trees that fit the desired criteria and have reasonable statistical support. The relatively small number of candidates that survive this filter are screened by human annotators and analyzed in more depth with different phylogenetic methods and substitution matrices. This is how the bacterial genes in the diatom genome were detected (Bowler et al. 2008).

Another analysis of the diatom genomes using a similar approach produced a surprising result: there were several times more genes closely related to green algae than to the red alga *Cyanidioschyzon merolae* (Moustafa et al. 2009). The largest number of genes appear to have been acquired from prasinophytes, a basal group of small green algae that are a prominent part of the oceanic biota. The authors

suggested that the common ancestor of the chromists might have had a green algal endosymbiont whose plastid was later replaced by a red algal plastid. However, there were several problems with this analysis, not least of which was that only one small red algal genome was available for analysis, compared to four prasinophyte genomes and two chlorophyte genomes (Dagan and Martin 2009; Elias and Archibald 2009). This is indeed a very interesting finding if it is confirmed by more in-depth studies, such as the careful examination of individual gene trees that showed that several enzymes of carotenoid biosynthesis were of green algal origin (Frommolt et al. 2008).

The most rigorous type of phylogenomic analysis involves the construction of "supertrees" based on concatenated sequences of a large number of proteins. A set of sequences that pass the initial phylogenetic pipeline are screened further for adequate overlap, taxon sampling, and absence of paralogs. The much smaller number of sequences that pass these steps are concatenated and assembled into a multi-protein alignment. Then, phylogenetic trees are generated with several different methods and models of evolution. Theoretically, this approach is very powerful and should eventually result in the much sought-after "Tree of Life".

A 16-protein dataset from 46 taxa gave good support for the haptophytes and cryptophytes being sister groups, and for the heterokonts and alveolates being each other's closest relatives (Hackett et al. 2007). In this case, genes that had something to do with plastid function were deliberately omitted in order to concentrate on the evolution of the host lineages. A much larger dataset (102 genes) making use of an EST survey from the photosynthetic cryptophyte *Guillardia theta* also provided strong support for these two relationships (Patron et al. 2007). Two large datasets with a different collection of taxa also grouped the heterokonts with the alveolates and quite separate from the haptophytes and cryptophytes (Burki et al. 2007, 2008). Yet another dataset (143 genes, 48 taxa), analyzed in a somewhat different way, also placed the heterokonts and alveolates together and separate from the haptophytes (Rodriguez-Ezpeleta et al. 2007; Hampl et al. 2009), with good statistical support provided that long-branch (very divergent) sequences were first removed (Hampl et al. 2009). None of these analyses supported monophyly of the three chromist groups, and only the analysis of the smallest dataset supported the chromalveolate hypothesis (Hackett et al. 2007).

The sister relationship of cryptophytes and haptophytes is now considered so well-supported by some researchers that a new supertaxon named Hacrobia has been proposed (Okamoto et al. 2009; Elias and Archibald 2009). However, this grouping is not supported by a complicated statistical approach designed to detect signals negating the chromalveolate hypothesis, although the grouping of

stramenopiles (heterokonts) with alveolates is supported (Baurain et al. 2010).

The real objective behind these large studies had nothing to do with the evolution of Chl *c* or the plastids that make it. It was to understand the evolutionary relationships of a new eukaryotic supergroup, named the Rhizaria. The Rhizaria include foraminiferans, filose amoebae and cercozoans. The only cercozoans with plastids are the chlorarachniophytes such as *Bigelowiella*, which got their plastids from a chlorophyte alga by secondary endosymbiosis (Fig. 1). In each of the analyses above, the Rhizaria formed a well-supported clade with the heterokonts plus the alveolates (the SAR clade), but not including the haptophytes and cryptophytes. Going further, the two most recent studies suggest a gigantic "megagroup" that includes the SAR clade and the primary photosynthetic lineages as well as the haptophytes and the cryptophytes (Burki et al. 2008; Hampl et al. 2009). One interpretation of this data is that the red algae were the ancestral group in the entire assemblage (Nozaki et al. 2009). If these ideas are borne out in subsequent studies, it will have a major impact on how we view the evolution of photosynthesis, and the evolution of eukaryotes as a whole.

The most serious problem with all the current phylogenomic studies is biased taxon sampling. At the current rate of improvement in sequence speed and analysis, this problem may diminish significantly in a few years but until then major revisions of the Tree of Life should be treated with healthy skepticism. Most of the sequenced eukaryotic genomes are those of Opisthokonts (animals plus fungi). The only red algal genomes available for the super-trees were those of *Cyanidioschyzon merolae* and *Galdieria sulfuraria*, two primitive reds with small genomes which are unlikely to be representative of the ancestral red algal endosymbiont (Dagan and Martin 2009), and the only cryptophyte sequences were from EST collections. However, two more red algal genomes are in the pipeline (*Chondrus crispus* and *Porphyra umbilicalis*) as well as the genomes of the cryptophyte *Guillardia theta* and the glaucophyte *Cyanophora paradoxa*. Once these are available, we will have a much better dataset for evaluating the chromalveolate hypothesis and determining the true level of HGT in chromist and alveolate genomes. In the meantime, new models are being proposed to account for some of the inconsistencies in the available data.

A look to the future and back to the past

One thing we can say for sure is that our views of eukaryotic phylogeny are in a state of flux. There is more and more evidence that nuclear genomes are a mosaic of genes acquired via vertical descent, endosymbiotic gene

transfer and horizontal gene transfer, as well as originating via endogenous processes such as gene duplication and recombination. Along with this there have been many independent losses of genes and organelles. The chromalveolate hypothesis is supported (or at least not contradicted) by a considerable amount of data, but some of the newer analyses suggest that it may be too simple a story. Sanchez-Puerta and Delwiche (2007) have presented an alternative hypothesis involving serial transfer of secondary plastids from one lineage to another. In their model, the common ancestor of cryptophytes and haptophytes was the one that acquired a red algal endosymbiont. At some later time, the secondary plastid and many nuclear genes were transferred to heterokont and alveolate lineages in one or more tertiary endosymbioses.

At least so far as the dinoflagellates are concerned, this seems very reasonable, especially considering the many ways in which some dinoflagellates have acquired plastids from other algal groups by tertiary and serial secondary endosymbiosis (reviewed by Archibald 2009). If the common ancestor of the apicomplexans and dinoflagellates had the same "tastes", it could have been the one that acquired a plastid by tertiary endosymbiosis. In that case, it would not be necessary to propose that ciliates ever had a plastid, although they could have acquired many genes by HGT from their food or from endosymbionts.

Another aspect that deserves some consideration is the possibility of multiple endosymbiotic relationships in the history of any major lineage. As mentioned above, there are indications of both green and red ancestors in the diatoms, apicomplexans, and *Bigelowiella*. A number of other examples are reviewed by Bodył et al. (2010) and Archibald (2009). We know that there have been many losses of plastids, so it would not be surprising if a host lineage went through a series of partially successful relationships followed by losses, before finally establishing a well-integrated secondary plastid. Throwing in some bacterial endosymbionts and intracellular parasites here and there, multiple endosymbioses could explain quite a bit of the mosaic nature of modern genomes, a model referred to as the "shopping bag model" (Larkum et al. 2007).

Could there even have been a photosynthetic ancestor at the origin of most eukaryotic lineages (Burki et al. 2008; Nozaki et al. 2009)? These are exciting times for everyone interested in the Tree of Life. There will undoubtedly be many modifications to its branching pattern over the next few years as key algal genomes are sequenced.

References

Archibald JM (2007) Nucleomorph genomes: structure, function, origin and evolution. BioEssays 29:392–402

Archibald JM (2009) The puzzle of plastid evolution. Curr Biol 19:R81–R88

Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. Proc Natl Acad Sci USA 100:7678–7683

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D et al (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86

Aury JM, Jaillon O et al (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature 444:171–178

Bachvaroff TR, Sanchez-Puerta MV, Delwiche CF (2005) Chlorophyll *c*-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. Mol Biol Evol 22:1772–1782

Baurain D, Brinkmann H, Petersen J, Rodriguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H (2010) Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes and stramenopiles. Mol Biol Evol 27:1698–1709

Becker B, Hoef-Embden K, Melkonian M (2008) Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. BMC Evol Biol 8:203

Bodył A, Mackiewicz P, Milanowski R (2010) Did trypanosomatid parasites contain a eukaryotic alga-derived plastid in their evolutionary past? J Parasitol 96:463–475

Bowler C et al (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244

Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland Å, Nikolaev SI, Jakobsen KS, Pawlowski J (2007) Phylogenomics reshuffles the eukaryotic supergroups. PloS ONE 2:e790

Burki F, Shalchian-Tabrizi K, Pawlowski J (2008) Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett 4:366–369

Cavalier-Smith T (1981) Eukaryote kingdoms: seven or nine? BioSystems 14:461–481

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 Euk Microbiol 46:347–366

Cavalier-Smith T (2002) Chloroplast evolution: secondary symbiogenesis and multiple losses. Curr Biol 12:R62–R64

Dagan T, Martin W (2009) Seeing green and red in diatom genomes. Science 324:1651–1652

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nature Rev Genet 6:361–375

Deschamps P, Moreira D (2009) Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. Mol Biol Evol 26:2745–2753

Doolittle WF (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet 14:307–311

Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu XN, Reith M, Cavalier-Smith T, Maier UG (2001) The highly reduced genome of an enslaved algal nucleus. Nature 410:1091–1096

Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M et al (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. PLoS Biol 4:e286

Elias M, Archibald JM (2009) Sizing up the genomic footprint of endosymbiosis. BioEssays 31:1273–1279

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

Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier UG, Grossman AR, Bhattacharya D, Lohr M (2008) Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. Mol Biol Evol 25:2653–2667

Gajadhar AA, Marquardt WC, Hall R, Gunderson J, Ariztia-Carmona EV, Sogin ML (1991) Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. Mol Biochem Parasitol 45:147–154

Gibbs SP (1970) The comparative ultrastructure of the algal chloroplast. Ann NY Acad Sci 175:454–473

Gibbs SP (1978) The chloroplasts of *Euglena* may have evolved from symbiotic green algae. Can J Bot 56:2883–2889

Gibbs SP (1981) The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. Ann NY Acad Sci 361:193–208

Gibbs SP (2006) Looking at life: from binoculars to the electron microscope. Annu Rev Plant Biol 57:1–17

Gillott MA, Gibbs SP (1980) The cryptomonad nucleomorph: its ultrastructure and evolutionary significance. J Phycol 16:558–568

Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI (2006) Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc Natl Acad Sci USA 103:9566–9571

Graham LE, Wilcox LW (2000) Algae. Prentice-Hall. Upper Saddle River, NJ, USA

Grauvogel C, Brinkmann H, Petersen J (2007) Evolution of the glucose-6-phosphate isomerase: the plasticity of primary metabolism in photosynthetic eukaryotes. Mol Biol Evol 24:1611–1621

Green BR (2004) The chloroplast genome of dinoflagellates—A reduced instruction set? Protist 155:23–31

Green BR, Anderson JM, Parson WW (2003) Photosynthetic membranes and their light-harvesting antennas. In: Green BR, Parson WW (eds) Light-harvesting antennas in photosynthesis. Kluwer Academic Press, Dordrecht, pp 1–28

Hackett JD, Anderson DM, Erdner DL, Bhattacharya D (2004) Dinoflagellates: a remarkable evolutionary experiment. Am J Bot 91:1523–1534

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

Hampl V, Hug V, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ (2009) Phylogenomic analyses support the monophyly of excavata and resolve relationships among eukaryotic "super-groups". Proc Natl Acad Sci USA 106:3859–3864

Helfrich M, Bommer B, Oster U, Klement H, Mayer K, Larkum AWD, Rüdiger W (2003) Chlorophylls of the *c* family: absolute configuration and inhibition of NADPH:protochlorophyllide oxidoreductase. Biochim Biophys Acta 1605:97–103

Howe CJ, Nisbet RER, Barbrook AC (2008) The remarkable chloroplast genome of dinoflagellates. J Exp Bot 59:1035–1045

Huang J, Gogarten JP (2007) Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol 8:R99

Iida K, Takishita K, Ohshima K, Inagaki Y (2007) Assessing the monophyly of chlorophyll-*c* containing plastids by multi-gene phylogenies under the unlinked model conditions. Mol Phylogen Evol 45:227–238

Inagaki Y, Nakajima Y, Sato M, Sakaguchi M, Hashimoto T (2009) Gene sampling can bias multi-gene phylogenetic inferences: the relationship between red algae and green plants as a case study. Mol Biol Evol 26:1171–1178

Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc Natl Acad Sci USA 107:10949–10954

Johnson MD (2010) The acquisition of phototrophy: adaptive strategies of hosting endosymbionts and organelles. Photosynth Res. doi:10.1007/s11120-010-9546-8

Keeling PJ (2009) Functional and ecological impacts of horizontal gene transfer in eukaryotes. Curr Opin Genet Dev 19:613–619

Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc B 365:729–748

Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryote evolution. Nat Rev Genet 9:605–618

Khan H, Parks N, Kozera C, Curtis B, Parsons B, Bowman S, Archibald J (2007) Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. Mol Biol Evol 24:1832–1842

LaJeunesse CT, Lambert G, Anderson RA, Coffroth MA, Galbraith DW (2005) *Symbiodinium* (Pyrrhophyta) genome sizes (DNA content) are smallest among dinoflagellates. J Phycol 41:880–886

Lane CE, Khan H, MacKinnon M, Fong A, Theophilou S, Archibald JM (2006) Insight into the diversity and evolution of the cryptomonad nucleomorph genome. Mol Biol Evol 23:856–865

Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM (2007) Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. Proc Natl Acad Sci USA 104:19908–19913

Larkum AWD, Lockhart PJ, Howe CJ (2007) Shopping for plastids. Trends Plant Sci 12:189–195

Lau AOT, McElwain TF, Brayton KA, Knowles DP, Roalson EH (2009) *Babesia bovis*: a comprehensive phylogenetic analysis of plastid-encoded genes supports green algal origin of apicoplasts. Exp Parasitol 123:236–243

Le Corguillé G, Pearson G, Valente M, Viegas C, Gschloessl B, Corre E, Bailly X, Peters AF, Jubin C, Vacherie B, Cock JM, Leblanc C (2009) Plastid genomes of two brown algae, *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal derived plastids. BMC Evol Biol 9:253

Li S, Nosenko T, Hackett JD, Bhattachary D (2006) Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. Mol Biol Evol 23:663–674

Lim L, McFadden GI (2010) The evolution, metabolism and functions of the apicoplast. Philos Trans Roy Soc B 365:749–763

Ludwig M, Gibbs SP (1985) DNA is present in the nucleomorph of cryptomonads: further evidence that the chloroplast evolved from a eukaryotic endosymbiont. Protoplasma 127:9–20

Ludwig M, Gibbs SP (1989) Evidence that the nucleomorphs of *Chlorarachnion reptans* (Chlorarachniophyceae) are vestigial nuclei: morphology, division and DNA-DAPI fluorescence. J Phycol 25:385–394

Martin W, Schnarrenberger C (1997) The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr Genet 32:1–18

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of

cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA 99: 12246–12251

Maruyama S, Matsuzaki M, Misawa K, Nozaki H (2009) Cyanobacterial contribution to the genomes of the plastid-lacking protists. BMC Evol Biol 9:197

McFadden GI, Reith ME, Munholland J, Lang-Unnasch N (1996) Plastid in human parasites. Nature 381:482

Moore RB, Oborník M, Janouškovec J, Chrudimsky T, Vancová M, Green DH, Wright SW, Davies NW, Bolch CJS, Heimann K, Šlapeta J, Hoegh-Guldberg O, Logsdon JM, Carter D (2008) A photosynthetic alveolate closely related to apicomplexan parasites. Nature 451:959–963

Moustafa A, Reyes-Prieto A, Bhattacharya D (2008) Chlamydiae has contributed at least 55 genes to plantae with predominantly plastid functions. PloS ONE 3:e2205

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science 324:1724–1726

NCBI Organelle Genome Resources (2010) http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid

Nelson MJ, Dang Y, Filek E, Zhang A, Yu VWC, Ishida K, Green BR (2007) Identification and transcription of transfer RNA genes in dinoflagellate plastid minicircles. Gene 392:291–298

Nozaki H, Maruyama S, Matsuzaki M, Nakada T, Kato S, Misawa K (2009) Phylogenetic positions of Glaucophyta, green plants (Archeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. Mol Phylogenet Evol 53:872–880

Oborník M, Green BR (2005) Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. Mol Biol Evol 22: 2343–2353

Oborník M, Janouškovec J, Chrudimsky T, Lukeš J (2009) Evolution of the apicoplast and its hosts: from heterotrophy to autotrophy and back again. Int J Parasitol 39:1–12

Oudot-Le Secq MP, Grimwood J, Shapiro H, Armbrust EV, Bowler C, Green BR (2007) Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage. Mol Genet Genomics 277:427–439

Patron NJ, Rogers MB, Keeling PJ (2004) Gene replacement of fructose-1, 6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. Euk Cell 3:1169–1175

Patron NJ, Inagaki Y, Keeling PJ (2007) Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. Curr Biol 17:887–891

Petersen J, Teich R, Brinkmann H, Cerff R (2006) A "green" phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts, and dinoflagellates. J Mol Evol 62:143–157

Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, Tonkin CJ, Roos DS, McFadden DI (2004) Metabolic maps and functions of the *Plasmodium falciparum* apicoplast. Nat Rev Microbiol 2:1–15

Reyes-Prieto A, Bhattacharya D (2007) Phylogeny of Calvin cycle enzymes supports Plantae monophyly. Mol Phylogenet Evol 45:384–391

Reyes-Prieto A, Moustafa A, Bhattacharya D (2008) Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. Curr Biol 18:956–962

Rice D, Palmer J (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

Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, Bohnert HJ, Philippe H, Lang BF (2005) Monophyly of primary photosynthetic eukaryotes: green plants red algae and glaucophytes. Curr Biol 15:1325–1330

Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Burger G, Roger AJ, Gray MW, Philippe H, Lang BF (2007) Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. Curr Biol 17:1420–1425

Rogers MB, Keeling PJ (2004) Lateral transfer and recompartmentalization of Calvin cycle enzymes of plants and algae. J Mol Evol 58:367–375

Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ (2007a) 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

Rogers MB, Watkins RF, Harper JT, Durnford DG, Gray MW, Keeling PJ (2007b) A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer. BMC Evol Biol 7:89

Rüdiger W, Grimm B (2006) Chlorophyll metabolism, an overview. In: Grimm B, Porra RJ, Rüdiger W, Scheer H (eds) Chlorophylls and Bacteriochlorophylls: biochemistry, biophysics, functions and applications. Springer, Dordrecht, pp 133–146

Sanchez-Puerta MV, Delwiche CF (2007) A hypothesis for plastid evolution in chromalveolates. J Phycol 44:1097–1107

Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF (2007) Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids. Mol Phylogenet Evol 44:885–897

Sato S, Clough B, Coates L, Wilson RJM (2004) Enzymes for heme biosynthesis are found in both the mitochondrion and plastid of the malaria parasite *Plasmodium falciparum*. Protist 155: 117–125

Stiller JW, Huang J, Ding Q, Tian J, Goodwillie C (2009) Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses? BMC Genomics 10:484

Stoebe B, Kowallik KV (1999) Gene-cluster analysis in chloroplast genomics. Trends Genet 15:344–347

Teich R, Zauner S, Baurain D, Brinkmann H, Petersen J (2007) Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses. Protist 158:263–276

Tyler BM et al (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261–1266

Waller RF, Patron NJ, Keeling PJ (2006) Phylogenetic history of plastid-targeted proteins in the peridinin-containing dinoflagellate *Heterocapsa triquetra*. Int J Syst Evol Microbiol 56:1439–1447

Wang Y, Joly S, Morse D (2008) Phylogeny of dinoflagellate plastid genes recently transferred to the nucleus supports a common ancestry with red algal plastid genes. J Mol Evol 66:175–184

Whitaker JW, McConkey GA, Westhead DR (2009) The transferome of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. Genome Biol 10:R36

Wilson RJM (2004) Plastid functions in the Apicomplexa. Protist 155:11–12

Wilson RJM, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW, Williamson DH (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. J Mol Biol 261: 155–172

Yoon HS, Hackett JD, Pinto G, Bhattacharya D (2002) The single ancient origin of chromist plastids. Proc Natl Acad Sci USA 99:15507–15512

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809–818

Zapata M, Garrido JL, Jeffrey SW (2006) Chlorophyll *c* pigments: current status. In: Grimm B, Porra RJ, Rüdiger W, Scheer H (eds) Chlorophylls and Bacteriochlorophylls: biochemistry, biophysics, functions and applications. Springer, Dordrecht, pp 39–53

Zhang Z, Green BR, Cavalier-Smith T (1999) Single gene circles in dinoflagellate chloroplast genomes. Nature 400:155–159

Zhang Z, Green BR, Cavalier-Smith T (2000) Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. J Mol Evol 51:26–40