First published online as a Review in Advance on June 28, 2007

The Origin and Establishment of the Plastid in Algae and Plants

Adrian Reyes-Prieto,[1] Andreas P.M. Weber,[2] and Debashish Bhattacharya[1]

[1]Department of Biological Sciences and Roy J. Carver Center for Comparative Genomics, University of Iowa, Iowa City, 52242-1324; email: adrian-reyes@uiowa.edu, debashi-bhattacharya@uiowa.edu

[2]Department of Plant Biochemistry, Heinrich-Heine-University, 40225 Düsseldorf, Germany; email: andreas.weber@uni-duesseldorf.de

Annu. Rev. Genet. 2007. 41:147–68

The *Annual Review of Genetics* is online at http://genet.annualreviews.org

This article's doi: 10.1146/annurev.genet.41.110306.130134

0066-4197/07/1201-0147$20.00

Key Words

cyanobacterium, endosymbiosis, endosymbiotic gene transfer, organelle, protein import, solute transporter

Abstract

The establishment of the photosynthetic organelle (plastid) in eukaryotes and the diversification of algae and plants were landmark evolutionary events because these taxa form the base of the food chain for life on our planet. The plastid originated via a putative single, ancient primary endosymbiosis in which a heterotrophic protist engulfed and retained a cyanobacterium in its cytoplasm. Once successfully established, this plastid spread into other protist lineages through eukaryote-eukaryote (secondary and tertiary) endosymbioses. This process of serial cell capture and enslavement explains the diversity of photosynthetic eukaryotes. Recent genomic and phylogenomic approaches have significantly clarified plastid establishment in the first algae, plastid genome evolution, the movement of endosymbiont genes to the "host" nuclear genome (endosymbiotic gene transfer), and plastid spread throughout the eukaryotic tree of life. Here we review these aspects of plastid evolution with a focus on understanding early events in plastid endosymbiosis.

INTRODUCTION

The Eukaryotic Tree of Life as Backdrop for Plastid Origin

Protist: microbial eukaryote not including plants and fungi

Algae: photosynthetic eukaryotes (protists) not including plants

'Chromalveolata': putative monophyletic group descended from a protist common ancestor that captured a red alga and maintained it as a secondary endosymbiont

Multigene phylogenetics and genome data from microbial eukaryote (protist) lineages have provided a renewed impetus to resolving the eukaryotic tree of life (e.g., 11, 71, 90), culminating recently in a formal classification of eukaryotes into 6 "supergroups" (3, 44). These supergroups (see **Figure 1**) contain the protistan roots of all multicellular eukaryotes and are currently defined as 'Opisthokonta' (e.g., animals, fungi, choanoflagellates), 'Amoebozoa' (e.g., lobose amoebae, slime molds), 'Archaeplastida' or 'Plantae' [red, green (including land plants), and glaucophyte algae], 'Chromalveolata' (e.g., diatoms, ciliates, giant kelps), 'Rhizaria' (e.g., cercomonads, foraminifera), and 'Excavata' (e.g., diplomonads, parabasalids). Although the supergroups broadly capture the diversity of eukaryotes, there are in fact only two that currently have robust support from molecular phylogenetic analyses, the 'Opisthokonta' and the 'Amoebozoa' (71). Therefore in this review all supergroups are marked with ' ' to denote their provisional nature. Of the remaining lineages, the 'Plantae' is gaining the most support from multigene trees (83) and features associated with the photosynthetic organelle (plastid) in these taxa (e.g., 63, 78, 99). This group is very likely to be monophyletic, a key feature that plays an important role in understanding plastid evolution. The 'Rhizaria' includes photosynthetic amoebae (chlorarachniophytes and *Paulinella chromatophora*) and receives

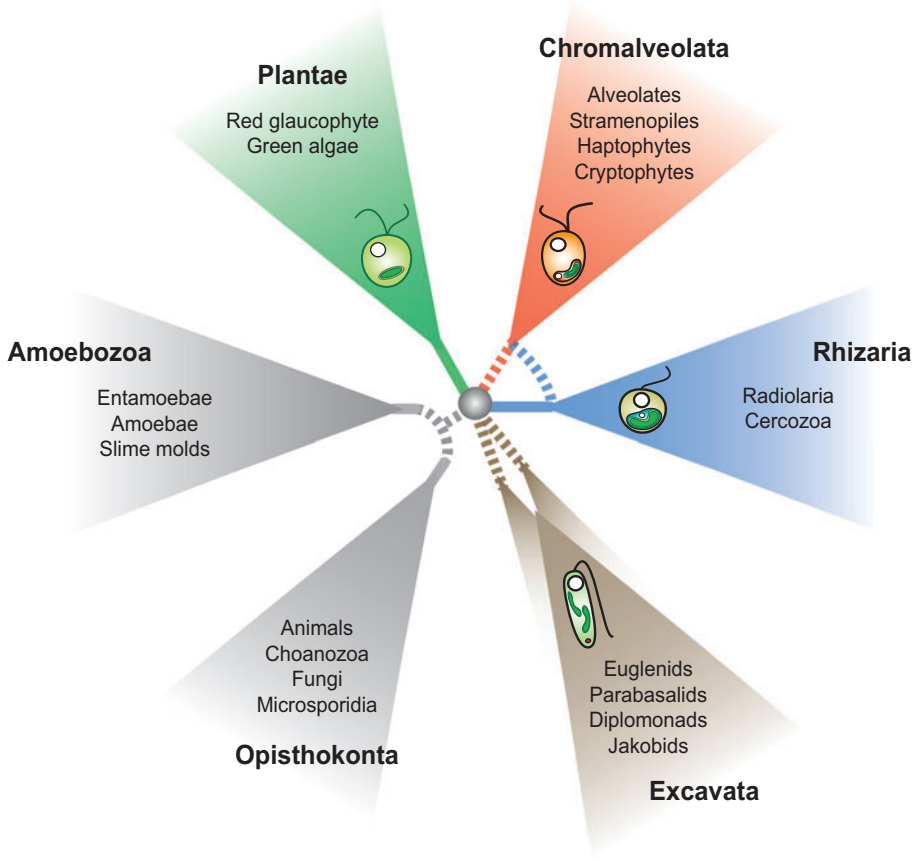

Figure 1

Schematic view of the eukaryotic tree of life showing the putative six supergroups. The broken lines denote uncertainty of branch positions in the tree. For example, the 'Rhizaria' are likely monophyletic but may branch within chromalveolates and the 'Excavata' may comprise at least two distinct lineages. The presence of plastid-containing taxa in the supergroups is shown with the cartoon of an alga.

moderate support in different studies but a broad taxon sampling that uses multigene methods has not yet been utilized for this supergroup (71). The 'Chromalveolata' and the 'Excavata' are currently the most controversial supergroups with no robust support from any study of nuclear genes for the monophyly of these groups, despite often extensive sampling of both taxa and loci (71, 89). The Excavata includes one important algal group, the euglenids (**Figure 1**). And finally, the future addition to phylogenies of uncultured environmental samples or poorly studied taxa such as amoebae and heterotrophic flagellates may affect supergroup membership and their interrelationships in ways that are currently difficult to predict.

In spite of these uncertainties, the tree of life is an important enterprise in molecular systematics and the overall phylogeny has started to take shape. This is critical because a well-sampled and resolved eukaryotic tree of life is invaluable for many reasons including the generation of hypotheses regarding plastid endosymbiosis and "host" cell evolution. The two most outstanding examples in this regard dealt with here are the 'Plantae' and the 'Chromalveolata' (**Figure 1**). If the 'Plantae' are monophyletic as most investigators in the field believe (but see 93a), then the initial cyanobacterial capture and enslavement occurred in the common ancestor of this lineage. Algal members of the 'Plantae' should therefore be outstanding models for understanding plastid establishment and the evolution of host-endosymbiont integration. The 'Chromalveolata' contains chromist and alveolate protists that are postulated to have shared a plastid of red algal origin in their common ancestor (16). If this hypothesis is true then we can study chromalveolate genomes for clues to eukaryotic plastid integration including gene transfer from the multiple genomes of the captured eukaryote. Here we address several key issues in plastid endosymbiosis including frequency of events and plastid donors, early events in plastid establishment, evolution of plastid protein import, intracellular gene transfer from the endosymbiont to the host nucleus (i.e., endosymbiotic gene transfer, EGT), and a discussion of the key features that characterize and differentiate permanent plastids (organelles) from temporary symbionts or endosymbionts.

THE ANCIENT PRIMARY PLASTID ENDOSYMBIOSIS

Evidence for a Single Plastid Origin in the 'Plantae' Ancestor

The eukaryotic plastid originated through endosymbiosis whereby a single-celled protist (host) engulfed and retained a free-living photosynthetic cyanobacterium (12, 13, 19, 28, 55a, 63, 70). Over time, the prokaryote was reduced to a double membrane-bound plastid and vertically transmitted to subsequent generations. A potential scenario for plastid origin is that cyanobacteria were engulfed through phagocytosis as a prey item countless times by the 'Plantae' ancestor, and in some of these cells, the cyanobacterium was not digested in the food vacuole but rather maintained as an endosymbiont (**Figure 2**). This is an astoundingly rare event because despite the many times that such a scenario of phagotrophy must have played out during eukaryotic evolution, only a single, ancient primary plastid endosymbiosis has persisted. The primary evidence for 'Plantae' monophyly comes from molecular phylogenetic and other comparative analyses of plastid and nuclear genes (12, 13, 19, 28, 63, 67, 70, 83) and genes involved in plastid function such as plastid protein import; i.e., members of the Tic-Toc translocon (62, 64), plastid-targeted solute transporters (99), and enzymes involved in plastid-localized biochemical reactions such as the Calvin cycle (78). The only other bona fide primary endosymbiosis resulting in the gain of photosynthesis that we know of (for details see below) occurred relatively recently in the filose amoeba *Paulinella chromatophora*, which harbors a plastid (cyanelle) derived from a *Prochlorococcus-Synechococcus*-type

'Plantae': putative monophyletic group containing the red, green (including land plants) and glaucophyte algae

Endosymbiosis: the uptake and retention of a foreign cell and its conversion into a cell organelle

EGT: endosymbiotic gene transfer

Phagocytosis: uptake of particles by the cell membrane and its internalization as a food vacuole

Primary plastid: PLASTID originating from the primary endosymbiosis of the cyanobacterium in the 'Plantae' ancestor

Tic: translocon of the inner chloroplast (plastid) envelope membrane

Toc: translocon of the outer chloroplast (plastid) envelope membrane

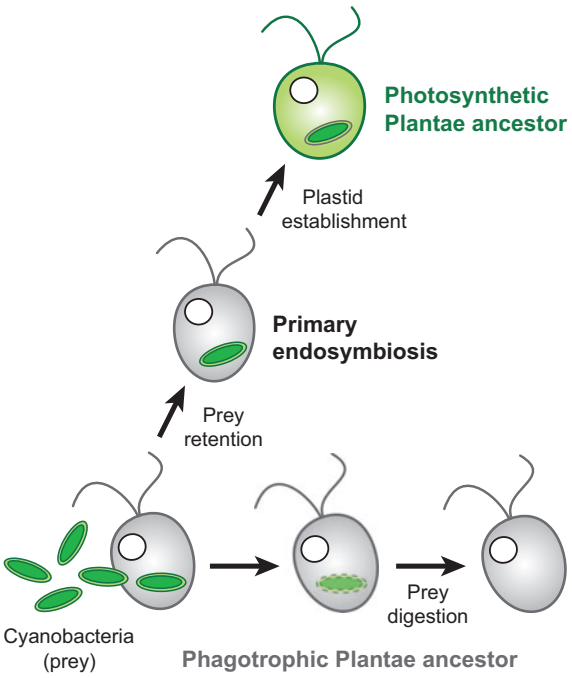

Figure 2

Hypothetical model showing the primary endosymbiotic origin of the plastid in the 'Plantae' common ancestor.

cyanobacterium (10, 56, 103). The sister of *Paulinella chromatophora*, *Paulinella ovalis*, lacks a plastid but is an active predator of cyanobacteria that are localized in food vacuoles in the cytoplasm (43). This observation provides some support for the phagotrophic origin of the ancient plastid. Molecular clock analyses using multigene data sets and "relaxed clock" approaches (e.g., penalized likelihood, Bayesian methods) that do not assume strict chronometric behavior of genes under study suggest that the 'Plantae' primary endosymbiosis is an ancient event in eukaryotic evolution. Although still controversial (23), recent analyses suggest that the primary plastid was established ca. 1.5 billion years ago in the Mesoproterozoic (e.g., 14, 35, 59, 101).

Antiporter: an integral membrane protein that couples the active transport of two different molecules in opposite directions across the membrane, as in the plastid triosephosphate/phosphate antiporter

ER: endoplasmic reticulum

Early Events in 'Plantae' Plastid Evolution

We have hypothesized that a crucial early step in endosymbiosis must have been the establishment of a reliable connection between the host cell and the ancestral plastid to allow the controlled exchange of metabolic intermediates between the symbiotic partners. Regulated exchange is important because the unfettered flux of metabolites between the host and plastid would have had detrimental effects on the metabolism of both partners and thereby lowered the evolutionary fitness of the symbiosis. A diverse set of metabolite antiporters that are embedded in the inner membrane of current day plastids allows the controlled exchange of solutes between cellular compartments (100). This antiport function is dependent on the presence of a suitable counterexchange substrate on the *trans*-site of the membrane. It was recently shown that the plastid triosephosphate and related sugar transporters were established in the common ancestor of the red and green algae (and likely all 'Plantae', supporting their monophyly), allowing this first alga to profit from cyanobacterial carbon fixation (99). This evolutionary step likely rendered irreversible the association between the plastid and the host cell. The ancestral plastid antiporter evolved from an existing metabolite translocator in the host cell that had evolved due to the pre-existence of mitochondria and an endomembrane system and was likely transferred to the ancient plastid via membrane fusion. This hypothesis of a close interaction between plastid envelope membranes and the host endomembrane system is supported by the observation that the outer leaflet of the outer membrane of extant primary plastids consists mainly of ER-derived phospholipids (17, 21). A more recent analysis of 83 annotated plastid solute transporters from *Arabidopsis thaliana* shows that the majority of these genes that have a resolved phylogeny in this taxon and in other 'Plantae', including all carbohydrate transporters, originated from co-option (i.e., gene duplication followed by retargeting to the plastid) of host genes, whereas about a quarter are of cyanobacterial (i.e., endosymbiont) provenance. The nuclear origin of many transporters is supported by their absence in

currently sequenced cyanobacterial genomes, suggesting (not proving) that these genes were not originally of prokaryotic origin and were replaced over evolutionary time by host-derived homologs. This suggests that early plastid evolution was essentially a host-driven process (D.B. & A.P.M.W. unpublished data).

Once the endosymbiont was established in the 'Plantae' ancestor then more complex processes could play out. This includes gene transfer from the endosymbiont to the host nucleus and the import of now nuclear encoded proteins to the plastid. This latter problem was solved in the long-term in all canonical plastids with the Tic-Toc protein import system that required the evolution of a "transit" sequence (about 24–100 amino acids in length (e.g., 26) at the N terminus of nuclear-encoded plastid-targeted proteins. This extra sequence likely came about through mutations that extended the original open reading frame at its 5′ end and the encoded proteins were selected for due to their targeting capacity or the genes integrated downstream of existing host genes and were originally translated as chimeric proteins (see below). Once the sequences containing these transit peptides were established in a few genes, then they may have spread through exon shuffling into other photosynthetic genes (15, 58). Comparative analyses demonstrate the conservation of many aspects of the plastid import machinery in 'Plantae', again providing support for a single plastid origin in the common ancestor of this supergroup (62, 64).

PAULINELLA AND CLUES TO PLASTID ESTABLISHMENT

Evidence for an Independent Primary Endosymbiosis in *Paulinella chromatophora*

Paulinella chromatophora is a filose amoeba in the 'Rhizaria' that contains blue chromatophores and is distantly related to the green plastid containing chlorarachniophyte amoebae in this supergroup (10). *Paulinella* was first described in 1895 by Robert Lauterborn (51) and, although understudied in the past century, has played a prominent role in the field of plastid endosymbiosis. This species is the only known case of an independent acquisition of photosynthetic capacity through primary (cyanobacterial) endosymbiosis (45, 56, 65, 84), making *Paulinella* an important model for understanding plastid establishment. The *Paulinella* plastid (also referred to as the cyanelle) retains key cyanobacterial features such as peptidoglycan and carboxysomes, but can be considered a bona fide organelle for the following reasons. (*a*) The cyanelle is no longer bound by a vacuolar membrane and lies free in the cytoplasm. (*b*) The cyanelle number is regulated (i.e., two cyanelles in each mature host cell), implying genetic integration. (*c*) The cyanelle cannot be cultured without the host (43, 45, 46).

Organelle: a differentiated membrane-enclosed structure within a cell originating from endosymbiosis (i.e., plastid and mitochondrion)

Part of the *Paulinella* mystery was recently clarified. In this study, a robust phylogenetic positioning of the plastid in *P. chromatophora* was generated that provided key insights into its genome evolution (103). Two phage inserts containing 9.4-kb (kilobases) and 4.3-kb fragments of the *Paulinella* plastid genome were compared with homologous regions in sequenced cyanobacterial genomes. The *Paulinella* fragments showed significant colinearity with *Prochlorococcus-Synechococcus* species, with the strongest conservation of plastid gene order to *Synechococcus* sp. WH5701. The 9.4-kb fragment encoded a number of genes that have been transferred to the nucleus in other algae and plants (e.g., *psb*O), demonstrating the essential cyanobacterial nature of the *Paulinella* endosymbiont. Finally, a multigene phylogeny using *Paulinella* plastid proteins confirmed the phylogenetic affinity to *Prochlorococcus-Synechococcus* species suggested by the gene order data (103).

These results, although preliminary in nature, provide a foothold into understanding a recent primary plastid establishment. They

suggest most importantly that key insights into early events in primary endosymbiosis such as control of organelle division and carbon translocation (99) will likely come from analysis of the *Paulinella* nuclear genome rather than analysis of its recent endosymbiont. We postulate that genes crucial to primary plastid establishment such as those required for organelle division (e.g., *ftsZ*) have been transferred to the nucleus. In addition, it is expected that *Paulinella* has devised a way of translocating fixed carbons from the cyanelle to the host cytoplasm. Whether this has occurred through the co-option and retargeting of existing host endomembrane transporters to the plastid as hypothesized for the 'Plantae' ancestor (99) remains to be determined. For these reasons, *Paulinella chromatophora* is a prime target for a complete genome sequencing project.

Organelle or Endosymbiont?

The argument has been made (95) that the *Paulinella* inclusions do not rise to the rank of a true organelle because the existence of a Tic-Toc-type protein import system has not yet been demonstrated in this system. In the view of Theissen & Martin (95), the critical difference between endosymbionts and organelles is protein import because all (or most) of the cytosolic proteins in an endosymbiont are encoded in its own genome, whereas, as we have discussed above, most organellar proteins are nucleus-encoded, translated in the host cytosol, and targeted to the organelle using a protein import apparatus (20, 91). This issue begs one key question: Is there evidence of plastid protein import that is independent of the canonical Tic-Toc machinery and therefore could have facilitated this important function in the "pre-Toc-Toc" world of plastids, and therefore possibly also in *Paulinella*? It is clear, however, that we will not find the de novo origin of a Tic-Toc translocon in *Paulinella* or in any either case of independent primary plastid acquisition.

Protein Import through the Endomembrane System

Until recently, our knowledge of the proteomes of plastids and mitochondria was mostly based on bioinformatic analysis of the deduced proteomes of a relatively small number of plant species (1, 60). Computational tools such as TargetP (24) were used to identify organellar-targeting signals reasoning that the presence of such targeting signals would indicate localization in a particular organelle. This approach was inherently biased toward identifying organellar proteins that followed established protein import pathways, because known proteins following these canonical import routes were used to train the corresponding prediction programs. Recent progress in proteomics has, however, made it possible to generate comprehensive inventories of the proteomes of subcellular compartments and thus provide direct experimental evidence for subcellular localization of proteins. Several recent proteomics studies have analyzed the proteomes of various plastid subtypes and of mitochondria (7, 27, 34, 47, 72, 88, 98). With respect to protein import, two surprising discoveries were made: (*a*) a relatively large number of proteins are apparently targeted to both mitochondria and plastids (dual-targeting) and (*b*) the plastid stroma contains a sizeable number of proteins that do not carry plastid-targeting signals (41, 66). For example, in a comprehensive analysis of the total *Arabidopsis* chloroplast proteome, Kleffmann et al. (47) found that of 604 nuclear-encoded, plastid-localized proteins that were identified with high confidence, only 376 (62%) were predicted to have plastid-targeting signals using the program TargetP. Even more surprisingly, 49 proteins not featuring a plastid-targeting signal were predicted by TargetP to have endoplasmic reticulum (ER) signal sequences (8% of the total). When compared to the total *Arabidopsis* proteome, this subset of putatively non-canonically targeted proteins was significantly enriched in proteins that have their closest

ortholog in cyanobacteria (7.2% versus 4.4%). In addition, the genes encoding these proteins in many cases were expressed at relatively low levels, making it unlikely that the proteins in questions represented highly abundant contaminants from other cellular fractions. Together with the low level of contamination of the chloroplast fraction with abundant proteins from other cellular organelles such as mitochondria, this provides strong evidence for noncanonical targeting of a considerable share of the plastid proteome. In a more recent proteomics study of rice etioplasts, 240 plastid proteins were identified with very high confidence, of which 224 are encoded by the nuclear genome (the remainder being plastid-encoded; 98). Whereas 168 (75%) of the 224 nuclear-encoded proteins were predicted by TargetP to carry plastid-targeting sequences, 10 (4.5%) were predicted to localize to the secretory pathway (i.e., they feature an ER signal sequence; 98).

Of course, classification of an N-terminal extension as ER signal sequence by TargetP cannot serve as conclusive evidence for noncanonical targeting of a particular protein. In addition, it is possible that some sequences classified as ER signals by TargetP are actually recognized by the chloroplast protein import complex and thus follow the canonical import pathway. Nevertheless, proteomics indicates that targeting of proteins to chloroplasts might be more complex and might involve more pathways than previously assumed (42, 74, 91). A recent study by Villarejo et al. (97) provided direct experimental evidence for routing of an α-carbonic anhydrase (CAH1) to the *Arabidopsis* chloroplast via the secretory pathway. Similar to secreted carbonic anhydrases, CAH1 carries a short N-terminal extension that is predicted by TargetP as an ER signal sequence. However, using a specific antiserum, CAH1 was localized to the chloroplast fraction in subcellular localization studies and a carboxy-terminal CAH1-GFP-fusion protein was localized to the chloroplasts in transiently transformed *Arabidopsis* protoplasts. Whereas these experiments confirm that CAH1 is indeed plastid-localized, they do not show that the protein is routed through the secretory pathway. However, when the 40 N-terminal amino acid residues of CAH1 were fused to the N terminus of a GFP-protein carrying a KDEL ER-retention signal at its C terminus, the resulting chimeric protein was retained in the ER, indicating that the N terminus of CAH1 represents a functional ER signal sequence. Full-length in vitro transcribed and translated CAH1 was not imported into isolated chloroplasts, but into microsomes. These experiments indicated that CAH1 was entering the chloroplast through a pathway that is independent of the Tic-Toc complex, likely by vesicular transport through the secretory pathway. Indeed, routing of CAH1 through the secretory pathway was demonstrated by showing that authentic chloroplast CAH1 is *N*-glycosylated and that CAH1-GFP transport to the chloroplast in transiently transformed protoplasts was inhibited by Brefeldin A, an inhibitor of Golgi-mediated vesicular traffic. Taken together, this study provided for the first time conclusive evidence for protein transport to the chloroplast via the secretory pathway. In addition, the authors demonstrated that the *Arabidopsis* chloroplast stroma contains a surprisingly large number of additional *N*-glycosylated (i.e., fucosylated) proteins. Since fucose epitopes are added to the protein backbone within the Golgi apparatus (53), this indicates that routing of proteins to the plastid via the secretory pathway is not restricted to CAH1 (97).

GFP: green fluorescent protein

In a fascinating recent study, Andersson et al. (4) used a combination of optical tweezers and confocal microscopy to demonstrate for the first time a physical interaction of ER-membranes with the surface of the chloroplast in living cells. It was shown that distinct domains of the ER localized to the chloroplast surface and that a force of 400 pN was not sufficient to remove these ER-patches from the chloroplast surface. Whereas these membrane contact sites between the ER and chloroplasts were discussed in the context of lipid

Secondary plastid: plastid originating from a secondary endosymbiosis, as in the red algal-derived secondary plastid in chromalveolates

trafficking between the ER and chloroplast, it might well be that this close physical interaction between the ER and chloroplast is also involved in routing proteins to the chloroplast.

The findings outlined above are important with respect to plastid establishment during endosymbiosis because they conclusively demonstrate that import of proteins into the plastid does not exclusively require a functional Tic-Toc import apparatus. Based on these recent findings and other evidence, as outlined above, it is tempting to hypothesize that sorting of proteins to the evolving plastid initially occurred via the host protein secretion system. A more efficient sorting system evolved later, recruiting parts of the cyanobacterial protein secretion apparatus (75–77), pre-existing components of the mitochondrial protein import complex (68), as well as novel proteins. This hypothesis is consistent with the fact that protein targeting to chromalveolate plastids always involves routing through the secretory pathway, thus recapitulating the process that occurred during establishment of the primary plastid in the 'Plantae' ancestor (18). This ancient noncanonical pathway for targeting proteins to the evolving chloroplasts was critical for plastid targeting of solute translocators, most of which have evolved from existing endomembrane host transporters (99; D.B. & A.P.M.W., unpublished results). Even more strikingly, the triosephosphate/phosphate transporter was adopted by the chromalveolates by horizontal transfer from the nuclear genome of the red algal endosymbiont, followed by expansion of the chromalveolate gene family, and acquisition of ER-targeting signals, thus recapitulating this critical step in endosymbiont establishment (99).

Given these findings and ideas, we posit that organelle genesis does not need, and should not be defined by, the evolution of a complex protein import machinery. It is far more likely that the ancient primary plastid was established using existing tools available in the host cell (see **Figure 3**) and became an organelle long before the evolution of the canonical Tic-Toc system. This initial phase of protein import via the secretory pathway not only made permanent the endosymbiosis but also allowed time for the gradual development of more complex traits such as large-scale EGT, plastid-nuclear genome integration, and evolution of the Tic-Toc protein import machinery. In the next section, we briefly review how the complex machinery for photosynthesis developed by the 'Plantae' was then passed on to other protists via secondary endosymbiosis.

SECONDARY PLASTIDS

Origin of Secondary Plastids

Soon after the split of red and green algae, it is hypothesized that a member of the red lineage was engulfed by a nonphotosynthetic protist giving rise to the pigmented ancestor of the 'Chromalveolata' (16). This supergroup was originally defined as the algal lineages cryptophytes, haptophytes, stramenopiles, and dinoflagellates and the nonphotosynthetic ciliates and apicomplexans. Recent analyses of aquatic biodiversity suggests that katablepharid and telonemid protists also belong in this supergroup (69a, 86a). In separate, more recent endosymbioses, green algae were independently engulfed by the common ancestor of the chlorarachniophyte amoebae ('Rhizaria') and of the euglenids ('Excavata'), giving rise to two distinct lines of green secondary plastids (85) (see **Figure 4*a***). Evidence for a red algal plastid in chromalveolates comes from plastid gene trees, phylogenies inferred from nuclear-encoded plastid-targeted proteins, and the occurrence of unique gene duplications and protein-retargeting events in this lineage (e.g., 13, 25, 33, 54, 69, 102). As mentioned above, the branching order of chromalveolates and the overall monophyly of this supergroup remain in question. However, for our purposes, it is well documented that most plastid-localized proteins in photosynthetic chromalveolates are of red algal origin. Given this observation then, it is clear that

gene transfer was also rampant in chromalveolates but in this case from the nucleus of the secondary endosymbiont (red alga) to that of the host. In all chromalveolate groups except the cryptophytes, which retain a remnant of the red algal nucleus (30), the endosymbiont nucleus has been eliminated, indicating that all genes necessary to control the plastid have been transferred to the host nucleus (22). A typical tree inferred from a multiprotein

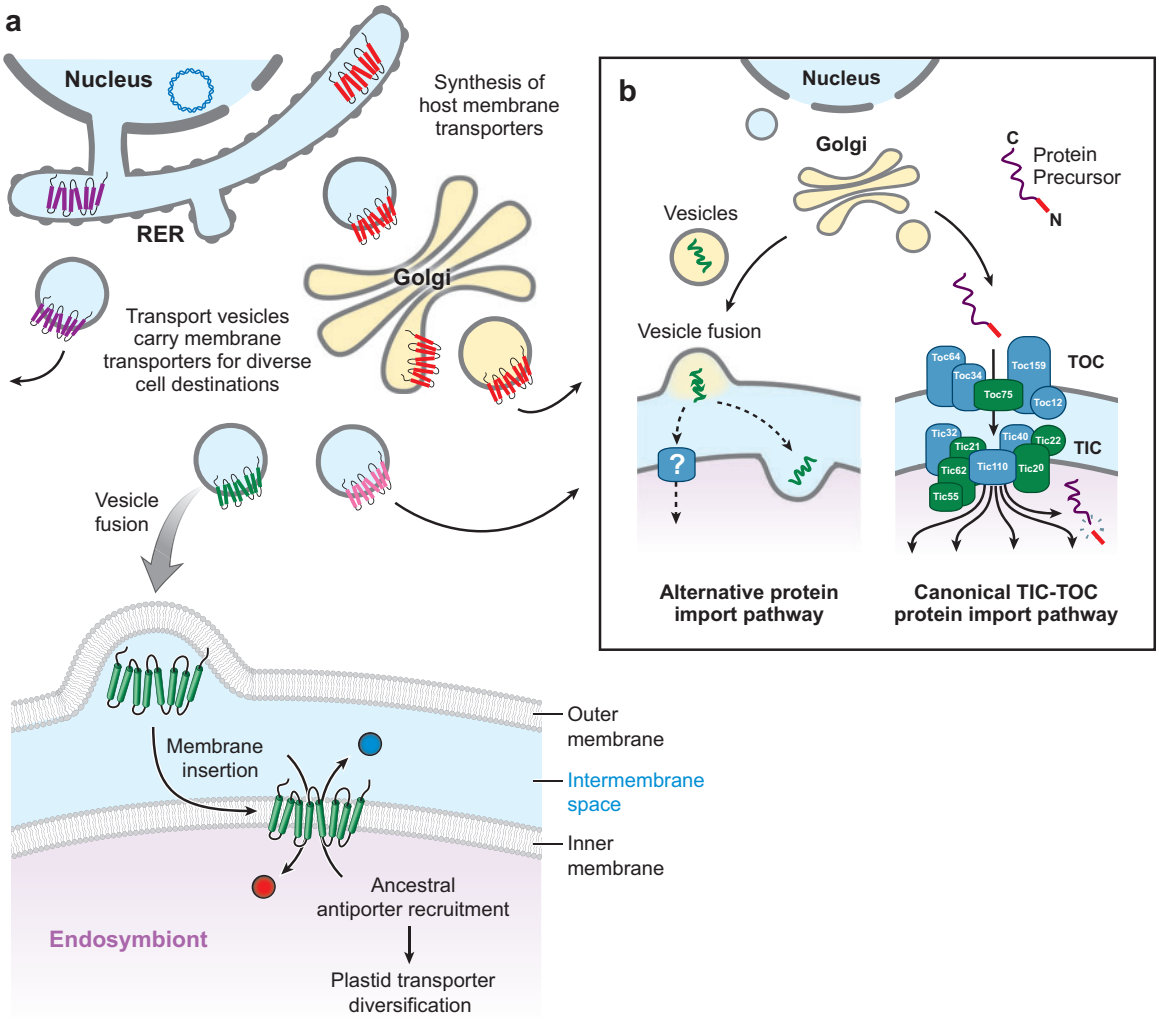

Figure 3

Hypothetical models for the origin of plastid-targeted transporters (*a*) and a Tic-Toc independent plastid protein import system (*b*) in the common ancestor of 'Plantae'. (*a*). Under this scenario, vesicles arising from the rough endoplasmic reticulum (RER) that carry membrane transporters to diverse cell locations fuse with the plastid outer membrane, delivering the first solute transporters to the nascent primary plastid. Metabolites being exchanged between the host and endosymbiont are represented as red and blue filled circles. (*b*). The initial protein import system in plastids was independent of the Tic-Toc system and resulted from vesicle fusion with the outer plastid membrane.

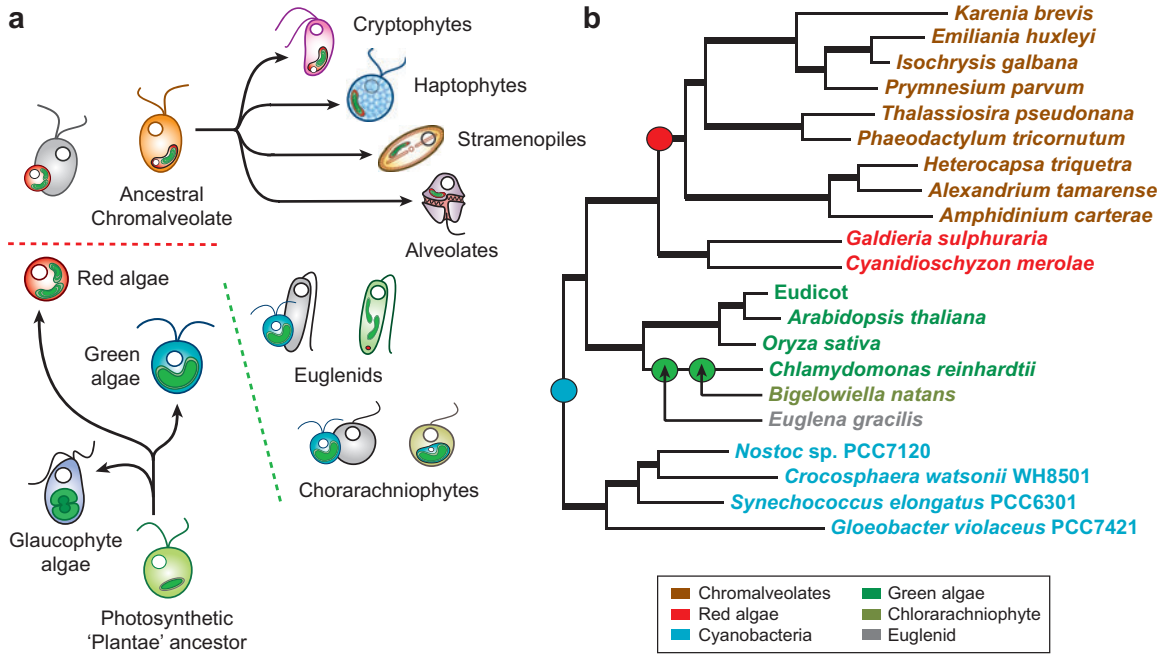

Figure 4

The origin(s) of plastids in photosynthetic eukaryotes. (*a*) Multiple lines of evidence (see text) support the single origin of the primary plastid in the 'Plantae' common ancestor. The plastid in red and green algae was then transferred to chromalveolates, euglenids, and chlorarachniophyte amoebae via independent secondary endosymbioses. (*b*) Phylogenetic tree based on maximum likelihood analysis of a data set of 6 nuclear-encoded plastid-targeted proteins that shows the origin of the primary plastid in 'Plantae' from a cyanobacterial source (*blue circle*), the secondary origin of the red algal plastid (*red circle*) in chromalveolates, and the independent origins of the green algal plastid (*green circles*) in euglenids, and chlorarachniophytes (see text for details). These latter two groups are not part of the phylogenetic analysis and have been simply added to the tree.

analyses of nuclear-encoded plastid-targeted proteins that supports the monophyly of chromalveolate plastids is shown in **Figure 4*b***. The separate origins of the chlorarachniophyte and euglenid green plastids that was inferred from analysis of plastid genomes from these taxa (85) have been added to this tree. The potential power offered by phylogenetics is exemplified by **Figure 4*b*** in which we can trace in one framework the origin of prokaryotic genes in eukaryotic nuclear genomes via primary endosymbiosis (filled blue circle) and the subsequent transfer of these genes from one or more red algae to the chromalveolates via secondary endosymbiosis (filled red circle). This type of analysis has also provided direct evidence for tertiary endosymbiosis in which an alga containing a secondary plastid was itself engulfed and retained by another protist (13, 40, 69). Although not discussed in detail here, this phenomenon is until now limited to dinoflagellates that are the masters of serial endosymbiosis (31).

Case Study: The Peculiar Path of Dinoflagellate Peridinin Plastid Evolution

The most common type of plastid in dinoflagellates contains peridinin as the major carotenoid. This pigment, although similar in structure to fucoxanthin, is unique to this group. Three membranes surround the peridinin-containing plastid, which is not

within the lumen of the ER as in other chromalveolate algae and the sister group of dinoflagellates, the apicomplexans. The plastid genome in peridinin plastids is remarkably different from that of other photosynthetic eukaryotes. Plastids generally contain a circular genome that, while varying in size and genetic content, is about 150 kb and encodes from 100–200 genes. In comparison, a free-living cyanobacterium typically has a genome of about 4000–5000 kb. Even the plastid genomes of nonphotosynthetic eukaryotes (e.g., *Plasmodium falciparum*, *Epifagus virginiana*, *Euglena longa*) are a single circular molecule with reduced gene content, primarily lacking genes involved in photosynthesis. In contrast, the plastid genome of peridinin-containing dinoflagellates is highly reduced and broken up into single gene minicircles. Currently, over a dozen proteins are encoded on these minicircles, in addition to the large (LSU) and small (SSU) subunits of the plastid ribosomal RNA, tRNAs, and "empty" minicircles or those encoding pseudogenes (8, 9, 36, 49, 50, 105, 106). These sequences include the core subunits of the photosystem (*atp*A, *atpB*, *pet*B, *pet*D, *psa*A, *psa*B, *psb*A-E) and four other proteins (*ycf*16, *ycf*24, *rpl*28, and *rpl*23). The remaining genes required for photosynthesis have been lost from the plastid and presumably moved to the nucleus. The migration of the plastid genome to the nucleus was recently documented for three dinoflagellates (*Alexandrium tamarense*, *Amphidinium carterae*, and *Lingulodinium polyedrum*; 6, 32). We analyzed a comprehensive set of 6480 unique cDNAs from *Alexandrium* and found that 15 genes (among others) that are encoded in the plastid in all other photosynthetic eukaryotes have been moved to the nucleus in this species (32). The majority of these nuclear genes encode a typical dinoflagellate plastid-targeting sequence (32). It is unknown what forces set into motion the remarkable movement of plastid genes to the nucleus in peridinin dinoflagellates and how these taxa have overcome the barriers to gene transfer that maintain many plastid genes in this organelle. Unlike eukaryotes that have reduced plastid genomes due to the loss of photosynthetic capacity or the evolution of a parasitic lifestyle, most peridinin dinoflagellates are free-living photoautotrophs or mixotrophs.

EST: expressed sequence tag

ENDOSYMBIOTIC GENE TRANSFER

Primary EGT

Another fundamental hurdle that was crossed in both primary and secondary plastid endosymbiosis was EGT from the captured cell to the host nucleus. The genes that remain in the plastid in both types of endosymbiosis are primarily involved in photosynthesis or transcription and translation of plastid genes, whereas most genes needed to maintain the plastid are encoded in the nucleus. Recent bioinformatic analyses are starting to unravel the quantity and quality of EGT from both primary and secondary endosymbionts. In an important analysis of the *Arabidopsis* genome, it was suggested that 18% of the nuclear genome (4500 genes) of this land plant may be of cyanobacterial origin (60), with about one half of these transferred genes postulated to be involved in plastid-independent functions. This result was recently tested using EST data from the glaucophyte *Cyanophora paradoxa*. Bioinformatic analysis of these algal data provided a different view of primary EGT. Out of 3576 *Cyanophora* nuclear genes that were analyzed, 1226 had significant hits to sequences in GenBank. Of the latter set, only 10.8% were of cyanobacterial origin and one ninth of these had nonplastid functions (79). Assuming that *Cyanophora* contains 12,000–15,000 genes, the cyanobacterial component is ca. 1500 genes in this species. These results indicate that, unlike plants, early diverging algal groups may retain a smaller number of endosymbiont genes in their nucleus with only a minor proportion of these recruited for nonplastid functions. Although surprising, these numbers are consistent with more recent analyses by Sato et al. (86) of cyanobacterial genes that are

HGT: horizontal gene transfer

nupDNA: nuclear DNA of plastid origin

Myr: million years

plastid-targeted in *Arabidopsis* (1192/25,500 genes = 4.7%) and the red alga *Cyanidioschyzon merolae* (676/5331 gene = 12.7%) using phylogenetic profiling. Richly & Leister (81) calculated a value of 880/25500 = 3.45% plastid-targeted genes of cyanobacterial origin in *Arabidopsis*. These authors also found that out of 857 plastid-targeted proteins of cyanobacterial ancestry that are shared between *Arabidopsis* and *Oryza*, about 650 constitute the minimal core set of endosymbiotic proteins required for angiosperm plastid function. The remainder of the plastid-targeted proteins in plants (and algae) is derived from the host, the protomitochondrial genome, or from horizontal gene transfer. For example, there are many cases of the retargeting of existing host proteins that either add to or replace the cyanobacterial homologs [e.g., Shikimate pathway see (80), Calvin cycle, see (78)]. Future analyses of other algal and plants genomes promise to more clearly explain the early history of primary EGT in 'Plantae', secondary EGT in chromalveolates, and differences in gene retention or recruitment over millions of years through duplication or horizontal gene transfer (HGT).

Recent EGT Events in Green Algae and Land Plants

An important contribution to understanding ancient EGT in 'Plantae' comes from analysis of modern-day plants. In flowering plants (angiosperms), the widespread presence of nuclear copies of plastid is well described (5, 73, 96). These findings were significantly accelerated by comprehensive analysis of the complete genome sequences of *Arabidopisis* and *Oryza sativa*. The nuclear-located fragments of plastid DNA (nupDNA) origin range from 1–131 Kb in length (61). Similarity searches at different BLAST cutoff values in the genomes of *Arabidopsis*, *Oryza*, and *Chlamydomonas reinhardtii* show the unicellular green alga has markedly less nupDNA than its angiosperm cousins (82). The fact that *Chlamydomonas* contains a single plastid could constrain the likelihood of DNA release to the cytoplasm (i.e., resulting in cell death; 55, 57, 82).

An excellent example of the frequency of plastid DNA integration is given by the 131-Kb nupDNA in chromosome 10 of *Oryza* (var. *japonica*). The fragment, which corresponds to ca. 97% of the complete *Oryza* plastid genome, is apparently absent in both *Oryza* (var. *indica*) and *Oryza nivara* (39). The *Oryza* (var. *japonica*) 131-Kb nupDNA shows evidence of recombination events, deletions, and insertions in comparison to the three *Oryza* spp. plastid genomes (39). Using molecular clock approaches, Huang et al. (39) estimated a time between 74,000–296,000 years ago for the integration event. The transferred fragment is undergoing a random mutation process (nonsynonymous substitutions number twice synonymous substitutions, and there are numerous nonsense mutations) and is destined for inactivation (39). A recent analysis suggests that ca. 0.2% (0.9 Mb distributed in 701 potential nupDNAs; BLAST E value $<10^{-10}$) of the rice genome corresponds to multiple nupDNAs. The integration of large fragments occurs mainly in the chromosome pericentromeric regions (61). These regions are rich in transposable elements and gene-poor and could be a potential haven for the emergence of "new genes" (29, 48, 61, 82, 94). In addition to the 131-Kb insertion, the rice genome contains 11 other nupDNA insertions greater than 10 Kb. They have undergone multiple rearrangements (e.g., inversions, deletions) but still retain high similarity (>99%) to the plastid genome, suggesting that once inserted, nupDNA is rapidly fragmented.

Analyses of the IR (inverted repeat) region of the plastid genome and in nupDNAs indicate that plastid DNA transfers have occurred repeatedly during rice evolution (61). Matsuo et al. (61) estimated that the majority of rice nupDNAs was acquired in the last million years (Myr). If we assume constant rates of transfer, then more than 90% of nupDNAs vanish after 2 Myr. The main

conclusion from this study is the identification of a dynamic equilibrium between the repeated transfer of plastid DNA and their rapid elimination from nuclear chromosomes (61). Therefore it appears that despite repeated transfers of nupDNA in plants, virtually all are destined for inactivation.

We suggest this latter observation reflects an advanced state of primary endosymbiosis in which nuclear-plastid genome integration is highly derived and large-scale EGT (plastid gene loss) is no longer a pervasive phenomenon. Clearly, a low level of primary plastid EGT is an ongoing process, particularly in plants (e.g., 58) and has taken an extreme turn in peridinin dinoflagellates (32). However, the large number of conserved, resident genes across most algal/plant plastid genomes (ca. 45–50 genes; e.g., 83) that have been maintained since the split of the 'Plantae' lineages over a billion years ago suggests that stasis has largely been achieved with regard to EGT. In comparison, it is approximated that ca. 1000 cyanobacterial genes were transferred to the nucleus in different 'Plantae', most of these prior to the separation of the red, green, and glaucophyte algae. For example, in our analysis of *Cyanophora* we found that of the nuclear genes of cyanobacterial origin in this species, ca. 80% were found as nuclear-encoded homologs in green and/or red algae (79). This result suggests that EGT was a significant force in the early evolution of this supergroup. This burst of EGT likely reflected strong selection to relocate plastid genes to the nucleus, for example, to escape the deleterious effects of Muller's ratchet in nonrecombining organelle genomes (58) and to establish nuclear control of plastid gene expression.

Experimental Evidence for EGT

In recent years, plastid transformation experiments in angiosperms have been used to estimate the rate of gene transfer from plastids to the nucleus. The design of these experiments takes advantage of the maternal inheritance of plastids in plants and the availability of antibiotic resistance genes (ARG) that can be actively expressed in the nucleus or in the plastid. In a key study, Huang et al. (37) transformed tobacco chloroplasts with a construct that included the ARG genes for kanamycin resistance (neomycin phosphotranspherase, *neo*STLS2) that used a plant-viral promoter and for spectinomycin resistance (aminoglycoside adenyltransferase, *aad*), that is suitable for selection in plastids. After biolistic transformation and crossing experiments, two plant lines were identified that were resistant to spectinomycin (with the *aad* gene successfully integrated in the plastid genome), but not for kanamycin (with the *neo*STLS2 gene not integrated in the nuclear genome, but in the plastid). Additional screening and selection were used to identify lines with both antibiotic-resistant genes resident in the plastid genome. Thereafter, the screens produced one kanamycin-resistant seedling (with the *neo*STLS2 gene active in the nucleus, transferred from the plastid). The conclusive screening of 250,000 test-crosses resulted in 16 plants that were kanamycin-resistant, providing evidence for successful plastid *neo*STLS2 transfer (37). Using the same strategy of antibiotic selection in tobacco, Stagemann & Bock (92) transformed cells in culture with a single plasmid containing ARG *nptII* (kanamycin-resistant and driven by a mosaic virus promoter that would be active in the nucleus) and *aadA* (controlled by a plastid promoter) to obtain cells with the plasmid integrated in the plastid genome. Thereafter, selection of transformed leaf cells resistant to kanamycin identified plant lines with the *nptII* gene active in the nucleus. Using this strategy, Stagemann & Bock (92) showed that successful gene transfer from the plastid to the nucleus in tobacco occurred in 1 out of 5 million somatic cells (93). Remarkably, in all of the kanamycin-resistant plants (i.e., that contain the *nptII* gene transferred from the plastid) both the active *nptII* and *aadA* genes were detected in the same genomic vicinity (ca. 1 Kb). Given that ARGs *nptII* and *aadA* were under the control of

different promoters, it is evident that flanking sequences were simultaneously transferred. This result suggests a DNA-based mechanism for DNA transfer and argues against a mRNA or cDNA intermediate (93).

In a similar line of experiments, Stagemann & Bock (92) addressed the question of how a plastid (i.e., prokaryotic) gene becomes active in the nucleus. These authors used the previously generated tobacco plant lines with the ARG *nptII* and *aadA* integrated in the nucleus (93). The assumption was that the inserted *aadA* gene would behave as a prokaryote-like acquired sequence. Through subsequent crosses of the plants, they selected lines with the *aadA* cassette integrated, but inactive (full spectinomycin sensitivity), in the nucleus and absent in the plastid genome (92). With exhaustive antibiotic screening, a new set of plants that are spectinomycin-resistant was identified. The segregation of the phenotype and RNA hybridization tests demonstrated that the *aadA* gene became active in the nucleus of eight plant lines. The accumulated *aadA* transcripts were of different lengths. This result indicated each gene transfer underwent molecular rearrangements. A fundamental conclusion from this study is that

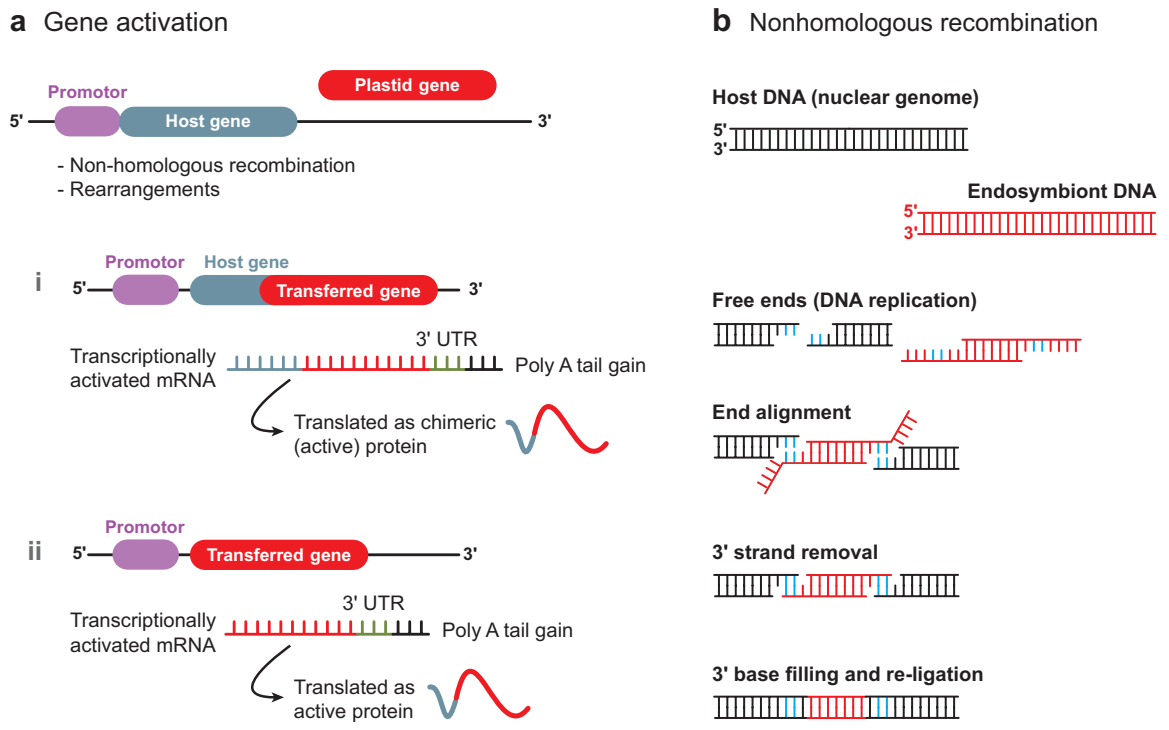

Figure 5

Endosymbiotic gene transfer and activation of nupDNA in the host nuclear genome. (*a*) Under one scenario (*i*), the gene of plastid origin can be activated if it integrates downstream of an expressed host gene by co-option of existing host promoters of transcription. This would result initially in a chimeric protein of host-plastid origin. The nupDNA could integrate via DNA-based molecular mechanisms (e.g., nonhomologous recombination) followed by sequence rearrangements (deletions, insertions, shuffling). Under a second scenario (*ii*), the nupDNA could integrate just downstream of an active host promoter of transcription and thereby be expressed. Gain of the polyadenylation signal under both scenarios does not involve major innovations at the sequence level (see text for details). (*b*) Model of nupDNA integration in nuclear DNA. Current evidence implicates nonhomologous DNA repair mechanisms such as nonhomologous end joining (NHEJ) as a potential mechanism for nupDNA integration.

once a plastid gene is located in the nuclear genome, transcriptional activation can occur in different ways. In all cases described here, the *aadA* genes was de novo activated through deleterious events (e.g., deletions, point mutations, and insertions) in the upstream *nptII* gene in the construct (inactive in all cases), but did not affect the mosaic virus promoter. In other words, the "new" gene was transcriptionally activated using the upstream preexisting promoter (92). Consistent with these experimental results, studies of rice EST libraries suggest some transferred plastid sequences have gained promoters in the nucleus. Moreover, some of them (*psb*E, *rpl*22, *inf*A) are apparently transcribed in both the nucleus and in the plastid (87).

Another important result from these experiments is that they provide insights into the acquisition of polyadenylation (polyA) signals in the transferred genes. The presence of a 3′ polyadenine extension is related to the stability of the mRNA in eukaryotes and may be a requirement for successful gene expression. All of the transferred *aadA* genes acquired their own polyA signal as an outcome of the intrinsic high AT-content of the plastid genomes. The original *aadA* construct included the 3′ UTR of the plastid gene *psb*A. Once nuclear transcribed, the *aadA* UTR (AT-rich region) matched randomly the typical consensus polyA signal in plants, which is also rich in AT (AU). This demonstrates that polyA signal gain does not pose a significant hurdle for the activation of nupDNA (92). Taken together, the results described above demonstrate that given selection for gene activation in nupDNA, EGT is a demonstrable mechanism for the transfer and successful expression (see **Figure 5*a***) of plastid genes. The integration results are most easily explained by DNA-directed processes and indicate that RNA-cDNA-mediated mechanisms (e.g., reverse transcription) likely contributed marginally to plastid EGT in land plants. The latter mechanism has, however, played an important role in the EGT of plant mitochondrial genes (e.g., 2).

Most studies suggest that plastid (and mitochondrial) DNA integration can be explained by nonhomologous recombination (38, 52, 92) such as nonhomologous end joining (NHEJ) DNA repair (**Figure 5*b***). When a DNA molecule undergoes a double-strand (DS) break, it is not possible to use a complementary strand to repair the damage. DS-breaks may occur during cell stress and DNA replication and transcription (52). NHEJ can facilitate DNA repair with chain ends lacking perfect complementarity (104), and this process is consistent with the possible incorporation of plastid (or any) DS DNA into the nuclear genome through the repair process.

UTR: untranslated region

CONCLUSIONS

In this review, we have discussed the origin and spread of the plastid in eukaryotes with a focus on clarifying early events in plastid establishment such as the origin of solute transporters and protein import. These issues, although of central importance to the field of plastid endosymbiosis, are nevertheless very difficult to address because of their ancient occurrence. This challenge is being met with large-scale genomic and phylogenomic analyses from an ever-growing list of protist genomes and by taking advantage of discoveries in well-established models such as *Arabidopsis* and *Chlamydomonas*. Given these extensive data sets, we can now frame reasonable hypotheses for plastid endosymbiosis, many of which can be tested through comparative analysis of different genes and genomes. For example, we can now with some confidence postulate that the ancient primary endosymbiosis occurred a single time in the 'Plantae' ancestor (i.e., excluding the *Paulinella* example in the 'Rhizaria') and that the host played a central role in plastid establishment by providing many of the solute transporters for reaping immediate benefits from the endosymbiont and the pre-Tic-Toc protein import apparatus via the secretory pathway. This latter development allowed the first algae to import both host-derived and

endosymbiont-derived proteins (that are now nuclear localized) into the organelle. The Tic-Toc protein import machinery was a later development in 'Plantae' evolution that was cobbled together from genes of cyanobacterial, mitochondrial, and foreign origin. Furthermore, evidence from plant models convincingly demonstrates that organellar DNA is frequently integrated into the nuclear genome. These studies also provide reasonable scenarios for endosymbiont gene activation. Given this body of knowledge and strong selection for the activation of organellar genes in the nucleus of the 'Plantae' ancestor we hypothesize that primary plastid endosymbiosis was largely a process of refinement and retooling of the host-plastid relationship that spanned several hundred million years of evolution. Much like Russian dolls, this complex and highly derived machinery was then transferred in toto (with subsequent modification) to chromalveolates and other protists through secondary (and tertiary) endosymbiosis.

Looking to the future, we suggest that, whereas plastid genomes offer relatively few surprises because they are largely impervious to HGT (e.g., 79a), the significantly more fluid nuclear genomes need to become the target for endosymbiosis research. This approach will provide many important insights into eukaryotic evolution. For example, algal nuclear genome data will allow us to identify the ancestral cyanobacterial gene set shared by all 'Plantae' in contrast to lineage-specific (e.g., red versus green algae) gene losses and recruitment events. The incorporation of a rich sample of EGT candidates from different algae in phylogenetic analyses will also provide greater resolution with regard to uncovering key events in their gene and gene family evolution. Resolving the eukaryotic tree of life is also crucial to endosymbiosis research and must be pursued at the genome level with the utmost urgency. A static or wildly fluctuating framework based on single or limited multigene gene data is not sufficient; rather the effort should be to understand gene and genome evolution on a grand scale. With the availability of complete or nearly complete genome sequences from a wide diversity of free-living protist taxa, we will be able to reconstruct with higher confidence the photosynthetic tree of life and gain insights into gene recruitment in these and other taxa through duplication or HGT and gene losses. In addition, effort needs to be expended on sequencing the genomes of a broader diversity of cyanobacteria to explore the metabolic diversity and the extent of HGT in these taxa (and therefore potentially of the ancestral endosymbiont) and to keep alive the search for the closest sister to the canonical plastid. Finally, the analysis of protist biodiversity in nature is key to advancing the field of plastid endosymbiosis. The identification of other models of novel plastid capture such as *Paulinella* would significantly accelerate our understanding of how predatory cells are converted into photoautotrophs or mixotrophs. These taxa may also help us better understand the role of the host in driving plastid establishment and test the ideas we have proposed in this review. The encouraging news is that, given an increasing appreciation of microbial biodiversity and the sinking costs of sequencing due to the advent of revolutionary new sequencing technology, the dream of unlimited data and opportunities shared by endosymbiosis researchers will in fact shortly become a reality.

SUMMARY POINTS

1. The eukaryotic tree of life is divided into supergroups.
2. A single primary endosymbiosis gave rise to the plastid in 'Plantae'.
3. The distribution and phylogeny of plastid solute transporters supports 'Plantae' monophyly and the origin of most of these transporters from the host genome.

4. *Paulinella chromatophora* contains a plastid of recent origin.
5. Plastid protein import can occur independently of the Tic-Toc system.
6. Most chromalveolates contain a plastid of red algal origin.
7. Primary EGT is essentially limited to genes of plastid function in algae.
8. Plastid gene transfer to the nucleus occurs frequently in plants.

DISCLOSURE

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

A.R.P. and D.B. are indebted to the U.S. National Science Foundation for generous, continuing support of our endosymbiosis and algal research in grants awarded to D.B. (MCB 0236631, EF 0431117, EF 0625440) and to NASA (NNG04GM17G) for supporting the *Cyanophora* primary EGT work. A.P.M.W. acknowledges support from the National Science Foundation (EF 0332882).

LITERATURE CITED

1. Abdallah F, Salamini F, Leister D. 2000. A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant Sci.* 5:141–42
2. Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* 29:380–95
3. Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* 52:399–451
4. Andersson MX, Goksor M, Sandelius AS. 2007. Optical manipulation reveals strong attracting forces at membrane contact sites between endoplasmic reticulum and chloroplasts. *J. Biol. Chem.* 282(2):1170–74
5. Ayliffe MA, Timmis JN. 1992. Plastid DNA sequence homologies in the tobacco nuclear genome. *Mol. Gen. Genet.* 236:105–12
6. Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM, Delwiche CF. 2004. Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* 155:65–78
7. Baginsky S, Siddique A, Gruissem W. 2004. Proteome analysis of tobacco bright yellow-2 (BY-2) cell culture plastids as a model for undifferentiated heterotrophic plastids. *J. Proteome Res.* 3:1128–37
8. Barbrook AC, Howe CJ. 2000. Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*. *Mol. Gen. Genet.* 263:152–58
9. Barbrook AC, Santucci N, Plenderleith LJ, Hiller RG, Howe CJ. 2006. Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes. *BMC Genomics* 7:297
10. Bhattacharya D, Helmchen T, Melkonian M. 1995. Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent rhizopod lineage containing the Euglyphina and the Chlorarachniophyta. *J. Eukaryot. Microbiol.* 42:65–69

11. Bhattacharya D, Katz LA. 2005. Frontiers in genomics: insights into protist evolutionary biology, University of Iowa, May 19–21, 2004. *J. Eukaryot. Microbiol.* 52:170–72
12. Bhattacharya D, Medlin LK. 1995. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. *J. Phycol.* 31:489–98
13. Bhattacharya D, Yoon HS, Hackett JD. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays* 26:50–60
14. Blair J, Shah P, Hedges SB. 2005. Evolutionary sequence analysis of complete eukaryote genomes. *BMC Bioinformat.* 6:53
15. Blanchard JL, Lynch M. 2000. Organellar genes: Why do they end up in the nucleus? *Trends Genet.* 16:315–20
16. 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.* 46:347–66
17. Cavalier-Smith T. 2000. Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* 5:174–82
18. 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. London Ser. B* 358:109–34
19. Delwiche CF. 1999. Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.* 154:S164–S77
20. Dolezal P, Likic V, Tachezy J, Lithgow T. 2006. Evolution of the molecular machines for protein import into mitochondria. *Science* 313:314–18
21. Douce R, Joyard J. 1981. Does the plastid envelope derive from the endoplasmic reticulum? *Trends Biochem. Sci.* 6:237–39
22. Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, et al. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature* 410:1091–96
23. Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. USA* 101:15386–91
24. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300:1005–16
25. Fast NM, Kissinger JC, Roos DS, Keeling PJ. 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.* 18:418–26
26. Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, et al. 2003. Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science* 299:705–8
27. Friso G, Giacomelli L, Ytterberg AJ, Peltier JB, Rudella A, et al. 2004. In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* 16:478–99
28. Gray MW. 1992. The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* 141:233–357
29. Gray YH. 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet.* 16:461–68
30. Greenwood AD. 1974. The Cryptophyta in relation to phylogeny and photosynthesis. In *Electron Microscopy 1974*, ed. JV Sanders, DJ Goodchild, pp. 566–67. Canberra: Aust. Acad. Sci.
31. Hackett JD, Anderson DM, Erdner DL, Bhattacharya D. 2004. Dinoflagellates: a remarkable evolutionary experiment. *Am. J. Bot.* 91:1523–34

32. Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, et al. 2004. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr. Biol.* 14:213–18
33. Harper JT, Keeling PJ. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol. Biol. Evol.* 20:1730–35
34. Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. 2004. Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* 16:241–56
35. Hedges S, Blair J, Venturi M, Shoe J. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* 4:2
36. Howe CJ, Barbrook AC, Koumandou VL, Nisbet RE, Symington HA, Wightman TF. 2003. Evolution of the chloroplast genome. *Philos. Trans. R. Soc. London Ser. B* 358:99–106; discussion 107
37. Huang CY, Ayliffe MA, Timmis JN. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422:72–76
38. Huang CY, Ayliffe MA, Timmis JN. 2004. Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *Proc. Natl. Acad. Sci. USA* 101:9710–15
39. Huang CY, Grunheit N, Ahmadinejad N, Timmis JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. *Plant Physiol.* 138:1723–33
40. Ishida K, Green BR. 2002. Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc. Natl. Acad. Sci. USA* 99:9294–99
41. Jarvis P, Robinson C. 2004. Mechanisms of protein import and routing in chloroplasts. *Curr. Biol.* 14:R1064–77
42. Jarvis P, Soll J. 2002. Toc, tic, and chloroplast protein import. *Biochim. Biophys. Acta* 1590:177–89
43. Johnson PW, Hargraves PE, Sieburth JM. 1988. Ultrastructure and ecology of *Calycomonas ovalis* Wulff, 1919, (Chrysophyceae) and its redescription as a testate rhizopod, *Paulinella ovalis* n. comb. (Filosea: Euglyphina). *J. Protozool.* 35:618–26
44. Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, et al. 2005. The tree of eukaryotes. *Trends Ecol. Evol.* 20:670–76
45. Kies L. 1974. Elektronenmikroskopische Untersuchungen an *Paulinella chromatophora* Lauterborn, einer Thekamobe mit blau-grunen Endosymbionten (Cyanellen). *Protoplasma* 80:69–89
46. Kies L, Kremer BP. 1979. Function of cyanelles in the thecamoeba *Paulinella chromatophora*. *Naturwissenschaften* 66:578
47. Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, et al. 2004. The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* 14:354–62
48. Knight J. 2002. All genomes great and small. *Nature* 417:374–76
49. Koumandou VL, Howe CJ. 2007. The copy number of chloroplast gene minicircles changes dramatically with growth phase in the dinoflagellate *Amphidinium operculatum*. *Protist* 158:89–103
50. Laatsch T, Zauner S, Stoebe-Maier B, Kowallik KV, Maier UG. 2004. Plastid-derived single gene minicircles of the dinoflagellate *Ceratium horridum* are localized in the nucleus. *Mol. Biol. Evol.* 21:1318–22

51. Lauterborn R. 1895. Protozoenstudien II. *Paulinella chromatophora* nov. gen., nov. spec., ein beschalter Rhizopode des Susswassers mit blaugrunen chromatophorenartigen Einschlussen. *Z. Wiss. Zool.* 59:537–44
52. Leister D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet.* 21:655–63
53. Lerouge P, Cabanes-Macheteau M, Rayon C, Fischette-Laine AC, Gomord V, Faye L. 1998. N-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol. Biol.* 38:31–48
54. Li S, Nosenko T, Hackett JD, Bhattacharya D. 2006. Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. *Mol. Biol. Evol.* 23:663–74
55. Lister DL, Bateman JM, Purton S, Howe CJ. 2003. DNA transfer from chloroplast to nucleus is much rarer in *Chlamydomonas* than in tobacco. *Gene* 316:33–38

55a. Margulis L, Fester R, eds. 1991. *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*. Cambridge: MIT Press. 470 pp.

56. Marin B, Nowack EC, Melkonian M. 2005. A plastid in the making: evidence for a second primary endosymbiosis. *Protist* 156:425–32
57. Martin W. 2003. Gene transfer from organelles to the nucleus: frequent and in big chunks. *Proc. Natl. Acad. Sci. USA* 100:8612–14
58. Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiol.* 118:9–17
59. Martin W, Rotte C, Hoffmeister M, Theissen U, Gelius-Dietrich G, et al. 2003. Early cell evolution, eukaryotes, anoxia, sulfide, oxygen, fungi first (?), and a tree of genomes revisited. *Iubmb. Life* 55:193–204
60. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, 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. USA* 99:12246–51
61. Matsuo M, Ito Y, Yamauchi R, Obokata J. 2005. The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. *Plant Cell* 17:665–75
62. Matsuzaki M, Misumi O, Shin IT, Maruyama S, Takahara M, et al. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428:653–57
63. McFadden GI. 2001. Primary and secondary endosymbiosis and the origin of plastids. *J. Phycol.* 37:951–59
64. McFadden GI, van Dooren GG. 2004. Evolution: Red algal genome affirms a common origin of all plastids. *Curr. Biol.* 14:R514–16
65. Melkonian M, Mollenhauer D. 2005. Robert Lauterborn (1869–1952) and his *Paulinella chromatophora*. *Protist* 156:253–62
66. Millar AH, Whelan J, Small I. 2006. Recent surprises in protein targeting to mitochondria and plastids. *Curr. Opin. Plant Biol.* 9:610–15
67. Moreira D, Le Guyader H, Phillippe H. 2000. The origin of red algae and the evolution of chloroplasts. *Nature* 405:69–72
68. Murcha MW, Elhafez D, Lister R, Tonti-Filippini J, Baumgartner M, et al. 2007. Characterization of the preprotein and amino acid transporter gene family in *Arabidopsis*. *Plant Physiol.* 143:199–212
69. Nosenko T, Lidie KL, Van Dolah FM, Lindquist E, Cheng JF, Bhattacharya D. 2006. Chimeric plastid proteome in the Florida "red tide" dinoflagellate *Karenia brevis*. *Mol. Biol. Evol.* 23:2026–38

69a. Okamoto N, Inouye I. 2005. The katablepharids are a distant sister group of the Cryptophyta: a proposal for Katablepharidophyta Divisio Nova/Kathablepharida Phylum Novum based on SSU rDNA and beta-tubulin phylogeny. *Protist* 156:163–79
70. Palmer JD. 2003. The symbiotic birth of plastids: How many times and whodunit? *J. Phycol.* 39:4–11
71. Parfrey LW, Barbero E, Lasser E, Dunthorn M, Bhattacharya D, et al. 2006. Evaluating support for the current classification of eukaryotic diversity. *PLoS Genet.* 2:e220
72. Peltier JB, Friso G, Kalume DE, Roepstorff P, Nilsson F, et al. 2000. Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell* 12:319–41
73. Pichersky E, Logsdon JM Jr, McGrath JM, Stasys RA. 1991. Fragments of plastid DNA in the nuclear genome of tomato: prevalence, chromosomal location, and possible mechanism of integration. *Mol. Gen. Genet.* 225:453–58
74. Radhamony RN, Theg SM. 2006. Evidence for an ER to Golgi to chloroplast protein transport pathway. *Trends Cell Biol.* 16:385–87
75. Reumann S, Davila-Aponte J, Keegstra K. 1999. The evolutionary origin of the protein-translocating channel of chloroplastic envelope membranes: identification of a cyanobacterial homolog. *Proc. Natl. Acad. Sci. USA* 96:784–89
76. Reumann S, Inoue K, Keegstra K. 2005. Evolution of the general protein import pathway of plastids. *Mol. Membrane Biol.* 22:73–86
77. Reumann S, Keegstra K. 1999. The endosymbiotic origin of the protein import machinery of chloroplastic envelope membranes. *Trends Plant Sci.* 4:302–7
78. Reyes-Prieto A, Bhattacharya D. 2007. Phylogeny of Calvin cycle enzymes supports Plantae monophyly. *Mol. Phylogenet. Evol.* In Press
79. Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D. 2006. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr. Biol.* 16:2320–25
79a. Rice DW, Palmer JD. 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
80. Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, et al. 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell* 5:1517–31
81. Richly E, 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
82. Richly E, Leister D. 2004. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Mol. Biol. Evol.* 21:1972–80
83. Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, et al. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* 15:1325–30
84. Rodriguez-Ezpeleta N, Philippe H. 2006. Plastid origin: replaying the tape. *Curr. Biol.* 16:R53–56
85. Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 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
86. Sato N, Ishikawa M, Fujiwara M, Sonoike K. 2005. Mass identification of chloroplast proteins of endosymbiont origin by phylogenetic profiling based on organism-optimized homologous protein groups. *Genome Inform.* 16:56–68

86a. Shalchian-Tabrizi K, Eikrem W, Klaveness D, Vaulot D, Minge MA, et al. 2006. *Telonemia*, a new protist phylum with affinity to chromist lineages. *Proc. R. Soc. London Ser. B* 273:1833–42
87. Shahmuradov IA, Akbarova YY, Solovyev VV, Aliyev JA. 2003. Abundance of plastid DNA insertions in nuclear genomes of rice and *Arabidopsis*. *Plant Mol. Biol.* 52:923–34
88. Siddique MA, Grossmann J, Gruissem W, Baginsky S. 2006. Proteome analysis of bell pepper (*Capsicum annuum* L.) chromoplasts. *Plant Cell Physiol.* 47:1663–73
89. Simpson AG, Inagaki Y, Roger AJ. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Mol. Biol. Evol.* 23:615–25
90. Simpson AG, Roger AJ. 2004. The real 'kingdoms' of eukaryotes. *Curr. Biol.* 14:R693–96
91. Soll J, Schleiff E. 2004. Protein import into chloroplasts. *Nat. Rev. Mol. Cell Biol.* 5:198–208
92. Stegemann S, Bock R. 2006. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell* 18:2869–78
93. Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc. Natl. Acad. Sci. USA* 100:8828–33
93a. Stiller JW, Reel DC, Johnson JC. 2003. The case for a single plastid origin revisited: convergent evolution in organellar genome content. *J. Phycol.* 39:95–105
94. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, et al. 2002. Human l1 retrotransposition is associated with genetic instability in vivo. *Cell* 110:327–38
95. Theissen U, Martin W. 2006. The difference between organelles and endosymbionts. *Curr. Biol.* 16:R1016–17
96. Timmis JN, Scott NS. 1983. Sequence homology between spinach nuclear and chloroplast genomes. *Nature* 305:65–67
97. Villarejo A, Buren S, Larsson S, Dejardin A, Monne M, et al. 2005. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat. Cell Biol.* 7:1224–31
98. von Zychlinski A, Kleffmann T, Krishnamurthy N, Sjolander K, Baginsky S, Gruissem W. 2005. Proteome analysis of the rice etioplast: metabolic and regulatory networks and novel protein functions. *Mol. Cell Proteomics* 4:1072–84
99. Weber AP, Linka M, Bhattacharya D. 2006. Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Eukaryot. Cell* 5:609–12
100. Weber AP, Schwacke R, Flugge UI. 2005. Solute transporters of the plastid envelope membrane. *Annu. Rev. Plant Biol.* 56:133–64
101. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21:809–18
102. Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. *Proc. Natl. Acad. Sci. USA* 99:15507–12
103. Yoon HS, Reyes-Prieto A, Melkonian M, Bhattacharya D. 2006. Minimal plastid genome evolution in the *Paulinella* endosymbiont. *Curr. Biol.* 16:R670–72
104. Yu X, Gabriel A. 1999. Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell* 4:873–81
105. Zhang Z, Cavalier-Smith T, Green BR. 2002. Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. *Mol. Biol. Evol.* 19:489–500
106. Zhang Z, Green BR, Cavalier-Smith T. 1999. Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400:155–59