THE ROYAL SOCIETY

Published online 18 December 2002

Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote–eukaryote chimaeras (meta-algae)

T. Cavalier-Smith

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK (tom.cavalier-smith@zoo.ox.ac.uk)

Chloroplasts originated just once, from cyanobacteria enslaved by a biciliate protozoan to form the plant kingdom (green plants, red and glaucophyte algae), but subsequently, were laterally transferred to other lineages to form eukaryote–eukaryote chimaeras or meta-algae. This process of secondary symbiogenesis (permanent merger of two phylogenetically distinct eukaryote cells) has left remarkable traces of its evolutionary role in the more complex topology of the membranes surrounding all non-plant (meta-algal) chloroplasts. It took place twice, soon after green and red algae diverged over 550 Myr ago to form two independent major branches of the eukaryotic tree (chromalveolates and cabozoa), comprising both meta-algae and numerous secondarily non-photosynthetic lineages. In both cases, enslavement probably began by evolving a novel targeting of endomembrane vesicles to the perialgal vacuole to implant host porter proteins for extracting photosynthate. Chromalveolates arose by such enslavement of a unicellular red alga and evolution of chlorophyll *c* to form the kingdom Chromista and protozoan infrakingdom Alveolata, which diverged from the ancestral chromalveolate chimaera. Cabozoa arose when the common ancestor of euglenoids and cercozoan chlorarachnean algae enslaved a tetraphyte green alga with chlorophyll *a* and *b*. I suggest that in cabozoa the endomembrane vesicles originally budded from the Golgi, whereas in chromalveolates they budded from the endoplasmic reticulum (ER) independently of Golgi-targeted vesicles, presenting a potentially novel target for drugs against alveolate Sporozoa such as malaria parasites and *Toxoplasma*. These hypothetical ER-derived vesicles mediated fusion of the perialgal vacuole and rough ER (RER) in the ancestral chromist, placing the former red alga within the RER lumen. Subsequently, this chimaera diverged to form cryptomonads, which retained the red algal nucleus as a nucleomorph (NM) with approximately 464 protein-coding genes (30 encoding plastid proteins) and a red or blue phycobiliprotein antenna pigment, and the chromobiotes (heterokonts and haptophytes), which lost phycobilins and evolved the brown carotenoid fucoxanthin that colours brown seaweeds, diatoms and haptophytes. Chromobiotes transferred the 30 genes to the nucleus and lost the NM genome and nuclear-pore complexes, but retained its membrane as the periplastid reticulum (PPR), putatively the phospholipid factory of the periplastid space (former algal cytoplasm), as did the ancestral alveolate independently. The chlorarachnean NM has three minute chromosomes bearing approximately 300 genes riddled with pygmy introns. I propose that the periplastid membrane (PPM, the former algal plasma membrane) of chromalveolates, and possibly chlorarachneans, grows by fusion of vesicles emanating from the NM envelope or PPR. Dinoflagellates and euglenoids independently lost the PPM and PPR (after diverging from Sporozoa and chlorarachneans, respectively) and evolved triple chloroplast envelopes comprising the original plant double envelope and an extra outermost membrane, the EM, derived from the perialgal vacuole. In all meta-algae most chloroplast proteins are coded by nuclear genes and enter the chloroplast by using bipartite targeting sequences—an upstream signal sequence for entering the ER and a downstream chloroplast transit sequence. I present a new theory for the four-fold diversification of the chloroplast OM protein translocon following its insertion into the PPM to facilitate protein translocation across it (of both periplastid and plastid proteins). I discuss evidence from genome sequencing and other sources on the contrasting modes of protein targeting, cellular integration, and evolution of these two major lineages of eukaryote 'cells within cells'. They also provide powerful evidence for natural selection's effectiveness in eliminating most functionless DNA and therefore of a universally useful non-genic function for nuclear non-coding DNA, i.e. most DNA in the biosphere, and dramatic examples of genomic reduction. I briefly argue that chloroplast replacement in dinoflagellates, which happened at least twice, may have been evolutionarily easier than secondary symbiogenesis because parts of the chromalveolate protein-targeting machinery could have helped enslave the foreign plastids.

Keywords: nucleomorph; chlorarachneans; chromalveolates; euglenoids; chloroplast evolution; evolution of genome size

One contribution of 21 to a Discussion Meeting Issue 'Chloroplasts and mitochondria: functional genomics and evolution'.

© 2002 The Royal Society
DOI 10.1098/rstb.2002.1194

1. INTRODUCTION

That chloroplasts evolved from an intracellular cyanobacterium (see Mereschkowsky 1905) is an established fact (Delwiche & Palmer 1998; Cavalier-Smith 2000*a*). However, Mereschkowsky's later idea that chloroplasts of differently pigmented groups of algae originated independently from correspondingly pigmented bacteria has been disproved (Mereschkowsky 1910). Instead, algal diversity has arisen in three different ways: (i) by differential losses of pigments in lineages diverging from the richly endowed ancestrally enslaved cyanobacterium; (ii) by secondary symbiogenetic enslavements of eukaryotic products of this primary divergence; (iii) and by evolution in the resulting eukaryote–eukaryote chimaeras of novel, purely eukaryotic pigments, notably chlorophyll *c* and the brown carotenoids fucoxanthin and peridinin. Secondary symbiogenesis, in effect, is the lateral transfer of entire eukaryotic organelles from one evolutionary lineage to another. Such lateral transfer of pre-existing organelles implants both novel genomes and novel genetic membranes into a cell (Cavalier-Smith 1995, 2000*a*), thereby dramatically increasing cellular complexity much more than is normally achieved by mutation and selection alone. However, secondary symbiogenesis—conversion of a foreign eukaryotic cell into an enslaved organelle—does not occur in one step, but involves several thousand mutations.

Secondary symbiogenesis generated only three major algal groups during the history of life: euglenoids and chlorarachneans, whose ancestors enslaved a green alga from which they inherited chlorophylls *a* and *b*, and the chromalveolates (chromists and alveolates; Cavalier-Smith (1999)), a vast assemblage of eukaryotes stemming from a single ancestor that enslaved a red alga and evolved chlorophyll c_2 in lieu of phycobilisomes. Chromalveolates include not only the algae traditionally called chromophytes, but also numerous descendants that secondarily lost photosynthesis, such as the protozoan ciliates and Sporozoa (e.g. malaria parasites) and colourless chromists such as oomycetes and other pseudofungi, opalinids, labyrinthulids and thraustochytrids.

Thus, secondary symbiogenesis played key roles not only in the origins of most major groups of eukaryotic algae, but also in the history of many non-photosynthetic groups, some of which (e.g. most Sporozoa) have retained plastids and degenerate plastid genomes despite having abandoned photosynthesis hundreds of millions of years ago. Even those that may no longer have plastids are likely to retain genic evidence of past secondary symbiogenesis, e.g. oomycetes, formerly misclassified as fungi, retain a 6-phosphogluconate dehydrogenase gene ultimately of cyanobacterial origin (Andersson & Roger 2002). The cabozoan theory (Cavalier-Smith 1999) that euglenoids and chlorarachneans were both products of a single symbiogenetic incorporation of a green algal cell is less well established than the chromalveolate theory. Nevertheless, if it is correct, then several other heterotrophic protozoan groups evolved from chimaeric meta-algae by chloroplast losses. However, I focus on organisms known to have retained novel genetic membranes as a result of secondary symbiogenesis, thereby increasing their cellular complexity.

A genetic membrane (Cavalier-Smith 1995) is defined as one that always arises by growth and division from pre-existing membranes of the *same* kind (e.g. thylakoid membranes and the inner and outer membranes of the chloroplast envelope). They are contrasted with non-genetic membranes that arise developmentally by differentiation from a different type of membrane, e.g. lysosome membranes. In effect, genetic membranes constitute the germ line of membrane heredity, whereas non-genetic membranes are more temporary somatic developmental states. Increases in the number of genetic membranes are very rare in the history of life (Cavalier-Smith 1991*a*) and have arisen in three ways: (i) subdivision of an existing genetic membrane into two topologically and biogenetically distinct parts, e.g. the origin of thylakoids (Cavalier-Smith 2002*a*) or of the ER (Cavalier-Smith 2002*b*); (ii) acquisition of novel genetic membranes by symbiogenesis as in the origin of mitochondria (Cavalier-Smith 1983, 2002*b*) or chloroplasts (Cavalier-Smith 1982*a*); and (iii) endowment of a pre-existing membrane with new properties, including a novel ability to divide to perpetuate them directly. The latter was especially significant during secondary symbiogenesis: the enslaved alga's plasma membrane was converted into the PPM (figure 1); and the phagosomal membrane that originally surrounded the algae engulfed during secondary symbiogenesis into the EM—the name adopted here for the outermost membrane surrounding the plastid in alveolates and cabozoans.

All three modes of origin of new genetic membranes involve the evolution of substantially new kinds of protein targeting and are therefore evolutionarily onerous. By contrast, loss of genetic membranes can be substantially easier and has been more frequent in eukaryotes (but not bacteria; Cavalier-Smith (2002*a*)), contributing significantly to their cellular diversity. Although genetic membranes show only a limited form of heredity compared with the unlimited heredity of genomes (Maynard Smith & Szathmáry 1995), they are just as important for understanding cell and organismal evolution, which would be impossible without the perpetual interplay of both DNA and membrane heredity (Cavalier-Smith 2000*a*). Chloroplasts import proteins by complex macromolecular assemblies (translocons) embedded in their envelope. These consist of two structurally and functionally separate parts stuck together to make a gated channel across the envelope: Toc in the OM and Tic in the inner one.

I discuss the innovations in protein targeting and membrane biogenesis that integrated the algal cells into their hosts in more detail than in the original statement of the cabozoan and chromalveolate theories (Cavalier-Smith 1999). As before, I assign central significance to the novel insertion of the chloroplast Toc into the former plasma membrane of the algal symbiont (now PPM) so as to mediate protein import into the algal symbiont: the transit mechanism co-option theory. I shall argue that soon after this insertion the PPM Toc diverged in several respects from its ancestor. To emphasize this and ensure clarity I designate the putative derived translocon of the PPM of chlorarachneans and chromalveolates Top (Tpc would be inappropriate as the PPM, contrary to common usage (van Dooren *et al.* 2001), is *not* a chloroplast envelope membrane, but topologically, evolutionarily and (I predict) chemically fundamentally distinct).

In addition to the origin of chromalveolates and cabozoa by secondary symbiogenesis there are two minor cases of lateral organelle transfer, both involving dinoflagellates. These are really chloroplast replacements (Saldarriaga *et al.* 2001): one very small group of dinoflagellates (e.g. *Gymnodinium breve*) permanently replaced its typical peridinin-containing chloroplasts by haptophyte fucoxanthin-containing chloroplasts (Tengs *et al.* 2000) and one species (*Lepidodinium*) seems to have replaced its peridinin-containing plastid by a green algal one (Watanabe *et al.* 1990; Schnepf 1993). I shall not discuss chloroplast replacements, as at present very little is known. Chloroplast replacement, occurring only in organisms already having plastids of secondary origin, was probably substantially easier than secondary symbiogenesis itself as pre-existing protein-import components could be co-opted.

2. PRIMARY SYMBIOGENESIS AND CHLOROPLAST PROTEIN IMPORT

The only three groups that originated by direct divergence from the primary chimaera of a cyanobacterium and a biciliate protozoan host are the red algae (Rhodophyta), green plants (Viridaeplantae: green algae (Chlorophyta) and land plants (Embryophyta)) and the purely unicellular blue-green Glaucophyta. These comprise the kingdom Plantae (figure 2; Cavalier-Smith 1998). Their plastids are invariably in the cytosol and bounded by an envelope of only two membranes homologous with the two membranes of the cyanobacterial envelope (Cavalier-Smith 1982*a*); the glaucophytes even retain the bacterial peptidoglycan layer between them, but their blue-green colour is probably not ancestral for chloroplasts. As the red phycoerythrin pigments of most red algae are clearly homologous with those of cyanobacteria, the enslaved cyanobacterium almost certainly had phycobilisomes with phycoerythrin, as well as blue phycocyanin and allophycocyanin, and thus would have been red like many cyanobacteria today. After enslavement, phycoerythrin was lost before the last common ancestor (cenancestor) of glaucophytes, but was retained by their sister lineage which lost peptidoglycan instead.

Recent evidence that chlorophyll *b* synthesis in green plants and prochlorophyte cyanobacteria uses related enzymes suggests that the enslaved cyanobacterium had phycobilisomes and chlorophyll *b* (Tomitani *et al.* 1999). Thus, when the sisters of glaucophytes lost peptidoglycan and diverged to form red and green algae (now established as sisters: Moreira *et al.* 2000), the green algal ancestor retained chlorophyll *b* but lost phycobilisomes, whereas the red algal lineage lost chlorophyll *b* but kept phycobilisomes. The ancestral chloroplast evolved a new chlorophyll-binding protein that diverged into versions binding chlorophyll *a* and *b* in green plants and chlorophyll *a* and *c* in chromalveolates, probably with homologues in all photosynthetic eukaryotes (see Durnford *et al.* 1999).

More fundamental were two other innovations: insertion of proteins into the cyanobacterial envelope enabling the host to tap its photosynthesate and enslave it, and evolution of chloroplast-specific protein import. Although the latter partly evolved from the cyanobacterial export machinery, it probably required changes in several host proteins (Cavalier-Smith 2000*a*). The origin of a generalized protein-targeting mechanism in the chloroplast envelope enabled any proteins that mutationally acquired N-terminal transit sequences to be imported. This marks the transition from an obligate symbiont to a true organelle (Cavalier-Smith & Lee 1985) with major repercussions on the independence of the former cyanobacterium and the evolutionary fate of its genome. Once any copies of its genes accidentally transferred to the nucleus (inevitable sooner or later) acquired transit sequences, their proteins could be imported into the plastid and the plastid-gene copy lost. The theory of plastid monophyly (Cavalier-Smith 1982*a*) predicted that protein-import mechanisms should be fundamentally the same in glaucophytes, red algae and green plants. Now that this has been clearly established (Steiner & Löffelhardt 2002) there is really no reason to call glaucophyte chloroplasts 'cyanelles'; they are simply chloroplasts that retained one more ancestral character (peptidoglycan) than others and are not actually 'the closest relatives to cyanobacteria' (Steiner & Löffelhardt 2002).

(a) *The chloroplast envelope outer membrane translocon (Toc) determines chloroplast identity*

Every distinct genetic membrane requires a specific integral membrane protein receptor that is autocatalytically self-targeting and provides the molecular basis for membrane individuality and self-perpetuation (Cavalier-Smith 1995). For the OM of plant chloroplasts Toc may provide this genetic function (Cavalier-Smith 2000*a*). In angiosperms, Toc is fundamentally trimeric, comprising two transit-peptide receptor GTPases (Toc159 and Toc34 or 33) and a hydrophilic channel, Toc75. (For simplicity, I ignore the probably peripheral Toc64, which may improve efficiency rather than provide core function (Sohrt & Soll 2000), and the variation among different angiosperm cell types (Yu & Li 2001).) Toc 75 has a transit sequence that targets it to the OM. Thus, it can only be inserted into membranes that already have it; its insertion is truly autocatalytic, but not directly, as Toc159 is the primary receptor. Toc159, like most OM proteins, has no transit peptide but its insertion is Toc-dependent, catalysed by Toc34 (Hiltbrunner *et al.* 2001). Unfortunately, it is not known how Toc34 and most other OM proteins are inserted. They have hydrophobic insertion sequences at the N-terminus as in OEP14 (Tu & Li 2001) or at the C-terminus as in OEP7 and Toc34. Although OEP7 may simply use its highly discriminating insertion sequence (Lee *et al.* 2001) to recognize the special galactolipids and phospholipid composition of the OM (Schleiff *et al.* 2001) specific proteins are probably needed to mediate insertion of others, including Toc34 and OEP14 (Tsai *et al.* 1999).

For all such non-autocatalytic self-inserting OM proteins, something must ensure that they insert into the OM, not other cell membranes. It has been assumed that the unique OM lipid composition provides this specificity (Schleiff & Klosgen 2001); galactolipids help with transit peptide binding (Bruce 2001). For the best-studied OM protein (OEP7), galactolipids and/or a sulpholipid plus the low levels of charged lipids in the outer leaflet are essential for insertion, but do not ensure specificity

(Schleiff *et al.* 2001). The presence of these specific lipids in the OM depends on synthesis within the plastid and vectorial translocation from the inner membrane to the OM. The location of their synthesis depends, in turn, on the inner membrane location of their biosynthetic enzymes. As all are nuclear encoded, they have transit peptides and enter the plastid via Tocs. Therefore, even OM lipid composition depends ultimately on its possessing Tocs, so lipid-dependent self-insertion of OM proteins, even if not directly mediated by Tocs, fundamentally depends on their location in the OM. Thus, OM self-perpetuation in all its unique complexity relies on its pre-existing lipid composition and possessing Tocs, both maintained through every cell division since plastids began around 600 Myr ago. OM compositional specificity also depends on targeting peptides in each OM protein, historically attuned to this self-perpetuating lipid composition.

The need for specific proteins for inserting Toc34 and OEP14 (Tsai *et al.* 1999) suggests that these proteins might also help ensure heritability of OM specificity. If one of them was a Toc component (needed other than for binding transit peptides, which these proteins lack) then Toc would indeed be the central determinant of OM specificity (being required for the insertion of all three of its own core proteins) as well as of entry of most stromal and thylakoid and many inner membrane proteins. If another OM protein was also crucial for specificity (not just the mechanics) of insertion into the OM, one might have to envisage a dual role for two protein complexes, not merely Toc, in chloroplast envelope OM heredity.

3. SIMILARITIES AND CONTRASTS BETWEEN SECONDARY AND PRIMARY SYMBIOGENESIS

Secondary symbiogenesis involves the same two basic principles: the insertion of transport machinery to extract photosynthate and an organelle-specific import machinery. The complexity of the latter is the main reason I have long sought to minimize the number of cases of symbiogenesis invoked to explain the history of life.

The first evidence for secondary symbiogenesis was the existence of plastids separated by four distinct membranes from the cytosol (Taylor 1974). In secondary symbiogenesis, although most novel membranes derive from the eukaryotic symbiont (not only chloroplast membranes but also its plasma membrane that became the PPM and its nuclear envelope that became that of the NM), the outermost one originated from the host's food vacuole membranes formed around it by phagocytosis when it was originally engulfed. This is in marked contrast to the origin of mitochondria and plant chloroplasts, where the double membranes are homologous with the two membranes of the negibacteria (proteobacterium and cyanobacterium, respectively), which escaped from their food vacuole and multiplied freely in the cytosol (Cavalier-Smith (1982*a*, 1983, 1993*a*), where arguments against the earlier view that they retained the vacuole membrane instead are given). This retention of the phagosomal membrane, unlike mitochondria and plant chloroplasts, means that nuclear-coded chloroplast proteins of meta-algae (i.e. those that originated by secondary symbiogenesis; Cavalier-Smith (1995)) are invariably made by ribosomes attached to the RER into which they are targeted by signal sequences upstream of the transit sequences that suffice for chloroplast entry in plants. Thus, meta-algal N-terminal targeting sequences have two (or three for thylakoid proteins) functionally distinct topogenic sequences not just one (or two) as in plants (figure 1). Even stronger evidence for secondary symbiogenesis came from the discovery of relict miniaturized nuclei (NMs) in the periplastid space between the chloroplast envelope and PPM: in cryptomonads (Greenwood 1974) and chlorarachneans (Hibberd & Norris 1984). The former originated from the red alga enslaved to form the chromalveolates while the latter is the relic of the enslaved green alga that generated the cabozoan chimaera.

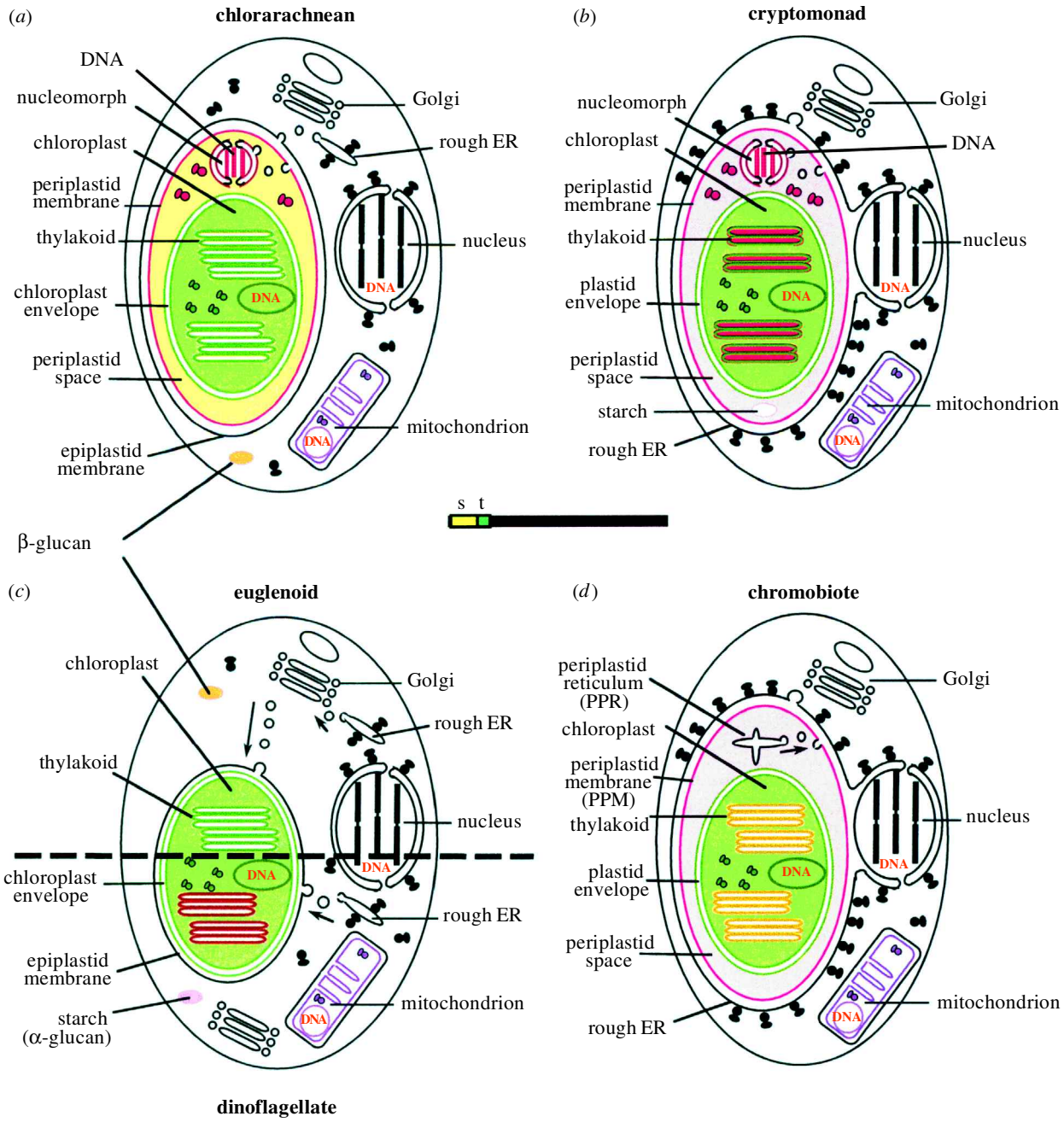
Just as protein-targeting is far more complex in meta-algae than other eukaryotes, so must be small molecule exchange across the extra membranes, all of which must somehow acquire lipids and proteins for their own growth and division; virtually nothing is known about this.

4. CHROMALVEOLATE CELLS AND GENOMES

The kingdom Chromista was established to include all algae with plastids enclosed by a PPM and RER membrane and their non-photosynthetic descendants (Cavalier-Smith 1981, 1986). Chromistan algae comprise cryptomonads, with red or blue (rarely brown) phycobilins and chlorophyll c_2, and chromobiote algae (Haptophyta and Ochrophyta, e.g. brown algae, diatoms, chrysophytes) typically brown with fucoxanthin and chlorophyll c_1, c_2 and often c_3. Photosynthesis has been lost several times by chromists, especially in heterokonts (they comprise the mainly photosynthetic Ochrophyta and heterotrophic Bigyra and Sagenista (Cavalier-Smith 1997)). Chromista exclude dinoflagellates, even though like nearly all chromistan algae they have chlorophyll *c* (specifically c_2) and so were traditionally grouped with them as chromophytes (Christensen 1989). This exclusion was because dinoflagellate chloroplast envelopes have three smooth membranes, not two as in chromists and plants, and chloroplasts are free in the cytoplasm, not inside a PPM and RER.

(a) *The protozoan infrakingdom Alveolata has a photosynthetic common ancestry with chromists*

Later, I classified dinoflagellates with Sporozoa, Ciliophora and Protalveolata as the protozoan infrakingdom Alveolata (Cavalier-Smith 1991*a*, 1993*b*). Alveolate monophyly was rapidly supported by rRNA trees (Gajadhar *et al.* 1991). The evolutionary picture apparently became more complex with the discovery that most Sporozoa actually have non-photosynthetic plastids, which fitted my contention that Alveolata were ancestrally photosynthetic. Initially it seemed that sporozoan plastids might be bounded by only three membranes (McFadden *et al.* 1996); Hopkins *et al.* (1999) still support this for *Plasmodium*, incorrectly in my opinion. After it was demonstrated that *Toxoplasma* plastids were bounded by four membranes (Köhler *et al.* 1997) and our phylogenetic analysis of dinoflagellate chloroplast genes suggested that they were related to those of *Toxoplasma* and heterokonts (Zhang *et al.* 1999, 2000), a new interpretation became more likely. This suggested that alveolates were sisters to



dinoflagellate

Figure 1. The four membrane topologies found in algae of secondary symbiogenetic origin. Successively greater degrees of loss of acquired genetic membranes are shown. In all cases, the food vacuole membrane that originally surrounded the alga when it was first internalized by phagocytosis is retained. All such algae import nuclear-coded chloroplast stromal proteins using a bipartite N-terminal targeting peptide, comprising a distal signal peptide (S) for crossing the RER membrane and a proximal transit peptide (T) for crossing the chloroplast envelope and PPM. Nuclear-encoded thylakoid proteins are tripartite, with an additional downstream signal sequence for entry into the thylakoids (Hiller *et al.* 2001). Except in dinoflagellates and euglenoids, the former plasma membrane of the enslaved alga is retained as a periplastid membrane; how this is traversed by imported proteins is unknown—it probably contains modified transit peptide receptors. In chlorarachneans and cryptomonads, its nucleus also remains as a nucleomorph, which codes many of its own proteins, many for periplastid ribosomes and periplastid enzymes, and some with unipartite N-terminal transit sequences that are imported into the chloroplast. Their PPM probably grows from vesicles budded from the nucleomorph envelope. (*a*) Chlorarachnean algae have the least modified state. The former symbiont has lost its mitochondria and Golgi but retains nucleus (NM) and plasma membrane (PPM). The outermost EM is smooth and proteins are thought to be imported by fusing of endomembrane vesicles. (*b*) Cryptomonads. Fusion of the EM with the nuclear envelope allowed it to be colonized by ribosome receptors and become functionally RER. This allows imported proteins to cross the OM co-translationally, obviating the need for vesicle fusion. (*c*) Euglenoids and dinoflagellates differ from others by the loss of both NM and PPM and the adhesion of the former food vacuole membrane (EM) to the former chloroplast OM to form a novel triple envelope. This happened independently in euglenoids, which import proteins by fusion of Golgi-derived vesicles with the OM and in dinoflagellates, where vesicle fusion is suspected, but their origin not demonstrated. (*d*) Chromobiote algae are similar to cryptomonads except for the loss of the NM following the transfer of its chloroplast protein genes to the nucleus. Found in all haptophytes (two classes: prymnesiophytes and pavlovophytes) and ochrophytes (heterokont algae—12 classes, e.g. brown algae, diatoms, chrysophytes, xanthophytes and raphidophytes).

chromists (collectively chromalveolates) and that their common ancestor had a plastid surrounded by four smooth membranes like Sporozoa but with chlorophyll c_2 and phycobilins like cryptomonads. According to this chromalveolate theory the photosynthetic chromalveolate cenancestor arose by a single symbiogenetic merger between a biciliate protozoan host and a unicellular red alga (Cavalier-Smith 1999). Instead of assuming there are five separate symbiogenetic events to generate chromalveolate plastid diversity (Delwiche & Palmer 1998), only one is needed (figure 3). This simplification was compellingly supported, unexpectedly swiftly, by the discovery that four different chromalveolate groups (assumed by others to have got chloroplasts independently (Delwiche & Palmer 1998)) had all replaced their original red algal plastid GAPDH enzyme by one of host origin (Fast *et al.* 2001).

The extreme improbability of the gene for the host cytosolic GAPDH undergoing gene duplication and acquiring bipartite targeting signals for transport across both RER and plastid envelope independently in cryptomonads, ochrophytes, dinoflagellates and Sporozoa makes it almost certain that this duplication and retargeting occurred once only in their common ancestor (Fast *et al.* 2001). Although such data are still lacking for haptophytes (the fifth plastid-bearing chromalveolate group), the ultrastructural and pigment identity of their fucoxanthin-containing chloroplasts and surrounding membranes to those of ochrophytes, plus a uniquely shared fluorescent cilium and similar mitochondrial ultrastructure, make it highly probable that haptophytes are sisters to heterokonts (Cavalier-Smith 1994) and that chromobiotes are holophyletic, even though they seldom group together on 18S rRNA trees. It is probable that nuclear 18S rRNA trees are simply slightly misleading (as they have been all too often in the past (Cavalier-Smith 2002*a*)) in separating cryptophytes from other chromalveolates, putting them instead among plants (Cavalier-Smith & Chao 2003*a*). Recent maximum-likelihood chloroplast 16S rRNA trees show chromobiotes and chromists as monophyletic with moderate bootstrap support (Tengs *et al.* 2000); richly sampled trees using five plastid genes show their monophyly extremely robustly (Yoon *et al.* 2002), though earlier too sparsely sampled distance trees often did not (Oliveira & Bhattacharya 2000).

(b) *Cryptomonads as eukaryote–eukaryote chimaeras: cells within cells*

The complete genome sequence of the cryptomonad NM (Douglas *et al.* 2001) fully confirms that it is a true, albeit highly miniaturized, nucleus as its ultrastructure with double envelope and nuclear-pore-like structures initially suggested to Greenwood (1974). NMs encode genes for three core histones and for acetylating and deacetylating them. Because they have 17 spliceosomal introns, encode numerous RNAs and proteins for removing them, and have telomeres at their ends (albeit unusual in sequence (Zauner *et al.* 2000)), the three chromosomes are clearly true eukaryotic ones. NM pores must really be nuclear pores, as the genome encodes two importin proteins and chromosomal region maintenance, a protein that interacts with nuclear pores, and is needed for transport (Douglas *et al.* 2001). It encodes numerous other factors indicating that transcription, messenger capping and polyA addition occur as in full-sized nuclei and that fibrillarin and small nucleolar ribonucleoproteins are involved in nucleolar rRNA processing. Several characteristically eukaryotic replication and DNA repair proteins are encoded. No major nuclear functions seem absent, though genes for some (notably DNA polymerase for replication) have been transferred to the main (former host) nucleus.

The periplastid space really is a relict cytoplasm containing five key functions:

(i) protein synthesis by ribosomes (the NM encodes their rRNA and most of their proteins, but almost all aminoacyl-tRNA synthetases must be imported from the cytosol);
(ii) protein assembly and export (both into the chloroplast and NM: 30 proteins have transit sequences and many others have NLSs);
(iii) protein degradation by proteasomes using ubiquitin-labelling;
(iv) starch synthesis and degradation (located in the cytosol in red algae and glaucophytes); and
(v) cell cycle control.

The presence of a gene for geranyl-geranyl-transferase lacking a transit peptide suggests that a key step in carotenoid synthesis may occur in the periplastid space. In addition to these basic functions, kinases, GTPases and a protein phosphatase that help regulate them are NM-coded.

There is compelling evidence for mitotic division, despite the inability to see any kind of spindle, microtubules or centrosome ultrastructurally. Four centrosome proteins (g-tubulin, Ranbpm, Hsp82, Hsp70), the centromeric histone CenpA, and the microtubule components a- and b-tubulin, together indicate that the NM must divide mitotically. Furthermore, the periplastid complex (NM, periplastid space and PPM) is conceptually equivalent to a complete eukaryotic cell in having machinery for coordinating NM and cell division. The two key cell cycle controls are the G1/S-phase transition (Mcm2, a key player, is NM-coded) and the G2/M phase checkpoint (requiring cyclin B and cyclin-dependent kinase: both also involved in centrosome activation and NM-coded).

Coding by the NM of 30 chloroplast proteins explains why it was never lost. These include proteins for protein import and targeting to thylakoids, chaperones, DNA gyrase, FtsZ for chloroplast division and rubredoxin, but no photosynthetic or antenna proteins (Douglas *et al.* 2001). There are also a few identified only by having transit sequences, which are distinctly different from those of nuclear-coded chloroplast proteins.

The NM genome sequence has been least revealing about the PPM, the homologue of the red algal plasma membrane. Three NM-coded transporters that lack transit sequences (one for sulphate, one for potassium and an ABC transporter) might be located in the PPM, which must have a capacity for actively or passively importing potassium, sulphate and other essential ions. It is also likely that the NM-coded phosphatidyl-4-OH kinase is in the PPM. Coupled with the presence of GTP-binding proteins this implies that the PPM/space retains elements

of the red algal signalling system via the second messenger inositol triphosphate.

The periplastid complex thus represents a eukaryotic cell pared down to a bare minimum. The cryptomonad cell is effectively a eukaryote cell within a cell. However, many genes for periplastid functions must have moved to the host nucleus, so the NM genome sequence reveals only the tip of the genic iceberg.

(c) *Protein targeting into and across the periplastid space*

The cryptomonad NM genome sequence has important implications for protein targeting via the periplastid space. First, it reveals that a novel pathway must exist across the PPM to import nuclear-coded periplastid and NM proteins. Second, it shows that the import machineries for NM- and nucleus-encoded chloroplast proteins differ substantially. The first conclusion follows because the NM genome does not encode all proteins needed for its own perpetuation (e.g. DNA polymerases) or for other periplastid functions (e.g. most aminoacyl-tRNA synthetases and starch-making enzymes are absent). Therefore, these proteins must be imported through RER and PPM membranes, but must not go into the chloroplast. A simple way of achieving this would be if periplastid proteins share the same PPM receptors as nuclear-coded chloroplast proteins but have different transit peptidase recognition sequences for a transit peptidase located only in the periplastid space (figure 4). This would remove the transit peptide, thus preventing entry into the plastid; periplastid chaperones and ATP-driven motors would pull these proteins into the periplasm, where they would remain unless they had NLS for entry into the NM or other signals for membrane insertion. Plastid stromal and thylakoid proteins would retain the transit sequence that could engage in a translocon (Tocm) in the plastid OM for import. I suggest that plastid OM proteins that are inserted in higher plants without having a transit peptide (Schleiff & Klosgen 2001) will share the same version of the transit peptide as the periplastid proteins and insert directly into the OM after its removal.

Second, the transit peptides for NM-coded chloroplast proteins are significantly different from those of the bipartite leaders of nuclear-coded proteins. Although both are rich in hydroxylated amino acids, NM peptides do not exhibit the red algal GPXM|XX motif (Douglas *et al.* 2001) suggesting that they use a different, novel transit peptidase. Likewise, they do not show the highly conserved upstream FXP that characterizes nuclear-coded chromist transit sequences (van Dooren *et al.* 2001); however, there is evidence for a weakly conserved motif FXN adjacent to or 1–4 amino acids from the N-terminal methionine, which might be related. Figure 4 presents a model for targeting cryptomonad chloroplast proteins assuming that two different OM translocons now exist having diverged from that of the red alga during its enslavement. The difference in transit sequences used by the two translocons may stem from a divergence between the OM Toc1 and the putative PPM Top1, especially as the conserved FXP motif is downstream of the usual specificity site for the signal peptidase. The fact that NM transit sequences target proteins to pea chloroplasts *in vitro*, but nuclear preproteins with deleted signal sequence do not (Wastl & Maier 2000), implies that NM and nuclear transit sequences are functionally different. Therefore, there must be a distinct protein-import machinery in the PPM that recognizes different targeting signals from those on the 30 NM-coded chloroplast proteins, which presumably still use the original Toc machinery in the chloroplast OM.

The NM genome encodes a homologue of Tic 22, which mediates association of Toc and Tic, and of the major Tic110 (formerly Iap100; Kessler & Blobel (1996)), which attaches the NM-coded stromal chaperonin Cpn60 (Douglas *et al.* 2001). No Toc genes were identified. A core component (SecE) of the thylakoid translocon for unfolded proteins and of the thylakoid TAT machinery for importing folded proteins (Tha4) are NM-encoded. Thus, cryptomonads have both normal thylakoid targeting pathways (Mori & Cline 2001).

(d) *Growth of the periplastid membrane and nucleomorph envelope*

We know nothing about PPM composition or the origin of its lipids. Although the NM envelope is a relic of the symbiont ER, there is no evidence that it retains the ancestral ability to make phospholipids or other membrane lipids. I suggest it does, and that the smooth vesicles (periplastid vesicles) in the periplastid space transport them to and fuse with the PPM enabling it to grow. Cryptomonad periplastid vesicles can be observed ultrastructurally fusing with, or budding from, the PPM. The indirect evidence from the NM sequence for nuclear-coded periplastid proteins, requiring their own import machinery, calls into question the idea that periplastid vesicles fuse with the chloroplast OM (Gibbs 1979, 1981*a*; Cavalier-Smith 1999). If vesicle transport were bidirectional and non-selective between the PPM and OM, it would equilibrate their protein and lipid composition and be incompatible with the existence of distinct receptors in the PPM and OM (Cavalier-Smith 1999). The existence of nuclear-coded periplasmic and NM proteins, plus the differences in NM- and nuclear-coded transit peptides, allows us reliably to infer for the first time that there must be differences in the composition of these two membranes. I suggest that periplastid vesicles fuse only with the PPM as in the ancestral membrane, and never evolved the capacity to fuse with the chloroplast OM first postulated by Gibbs (1979).

The NM membrane was originally part of the algal ER, with its own acyl transferase for making phospholipids and SRP machinery for inserting proteins, which together made it a distinct genetic membrane from the plasma membrane; the SRP RNA is probably lost (Douglas *et al.* 2001). The plasma membrane would have grown by the fusion of vesicles from the Golgi, which no longer remains. I suggest that the NM envelope retains an acyl transferase (imported from the cytosol or coded by a divergent NM orf) and ability to make phospholipids from fatty acids imported from the plastid. If it also retains the ability to bud off copII-coated vesicles, PPM growth could occur simply by their fusion directly with it rather than with the Golgi as ancestrally. This change could have come about simply by some Golgi vesicles bearing the appropriate vesicle fusion receptors fusing with the PPM during Golgi degeneration. This exocytosis model involves less change from the ancestral state than the alternative assumption

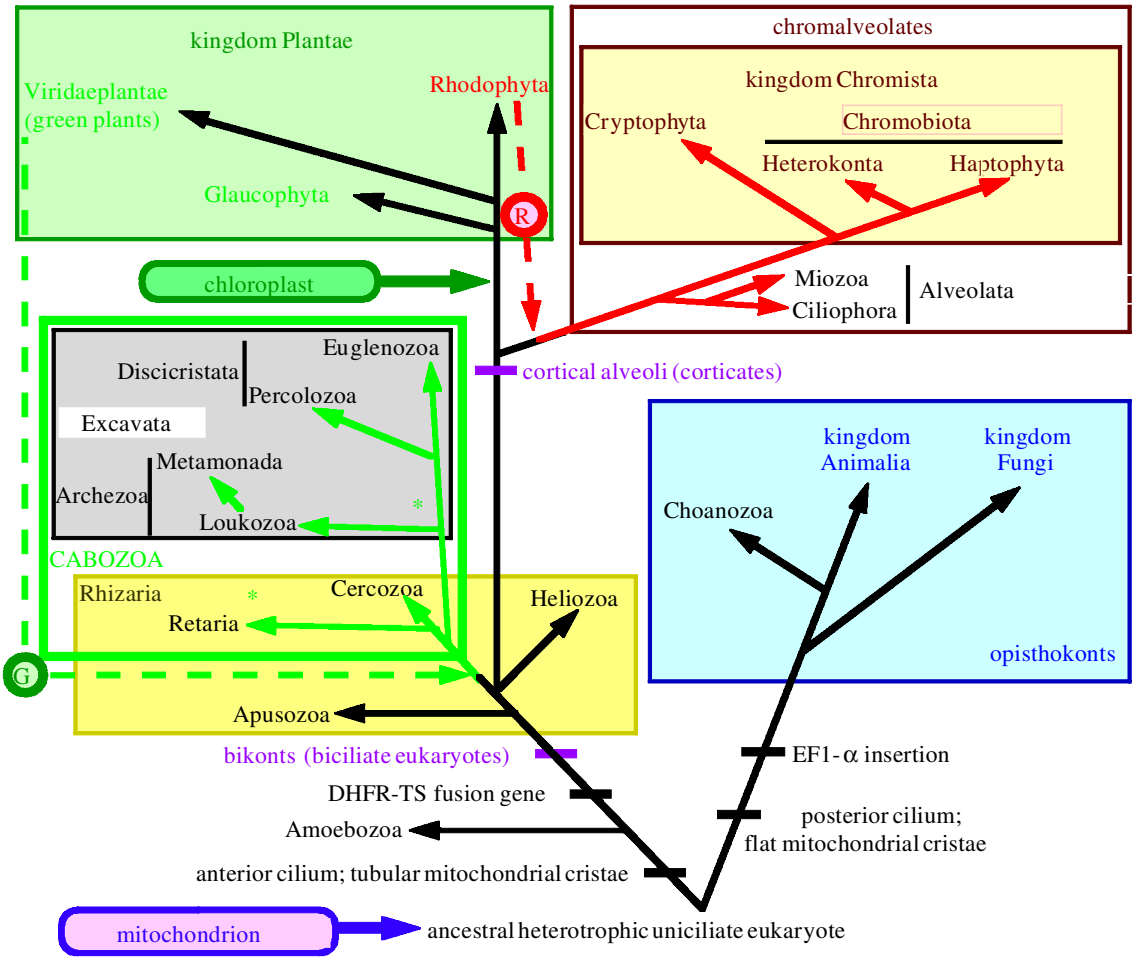


Figure 2. The eukaryotic phylogenetic tree based on a synthesis of ultrastructural, cell biological and molecular evidence, showing the four major symbiogenetic events. The ancestral eukaryote is held to have been a phagotrophic uniciliate aerobic zooflagellate that arose from a neomuran bacterial ancestor by the simultaneous origin of the cytoskeleton, endomembrane system, nucleus and cilium, coupled with the overlapping symbiogenetic origin of mitochondria from an intracellular a-proteobacterium (Cavalier-Smith 2000*b*, 2002*b*). The root of the tree (Stechmann & Cavalier-Smith 2002) is between the posteriorly uniciliate opisthokonts and the bikonts, which ancestrally evolved two cilia and ciliary transformation with a younger anterior cilium (Cavalier-Smith 2002*b*), but it remains uncertain whether Amoebozoa are really sisters to bikonts or opisthokonts or diverged even earlier. It is reasonably certain that opisthokonts and Amoebozoa are ancestrally heterotrophic and that chloroplasts arose symbiogenetically in an early bikont. It is firmly established that chromalveolates are monophyletic, ancestrally photophagotrophic, and evolved by the single enslavement of a red alga by a bikont host to form a eukaryote–eukaryote chimaera. There are equally strong protein-targeting arguments for the single secondary origin of the cabozoan chloroplast (G) in a common ancestor of Cercozoa and Euglenozoa (Cavalier-Smith 1999), but the idea remains controversial as compelling sequence evidence is unavailable to rule out the less parsimonious alternative possibility that euglenoid and chlorarachnean plastids were separately implanted as shown by the asterisks. Retaria comprise Foraminifera and Radiolaria and are probably sisters of Cercozoa. Heliozoa might actually be part of the cabozoan (or less probably the chromalveolate) clade, and thus also secondary heterotrophs. Whether Apusozoa are actually the most divergent bikonts and ancestrally heterotrophic, as shown, or are really also cabozoa is also uncertain.

that phospholipids are made only in the host ER and imported into the PPM by novel intraER soluble carriers allowing it to grow and generate periplastid vesicles by a relic of endocytosis.

This view of periplastid phospholipid synthesis means that the PPM will not have chloroplast glycolipids and sulpholipids or OM proteins needed for their insertion, so chloroplast OM proteins (probably all nuclear-coded) will not insert into it after their transit peptide is removed, but only enter the OM. Thus, conserved lipid composition of the chloroplast OM and PPM and a fundamental differentiation in protein content during secondary symbiogenesis were probably essential for successfully retargeting chloroplast envelope proteins whose genes were transferred to the nucleus. Superimposed on such conservation were the innovations in PPM translocons (Tops) and any additional insertion machinery needed to insert other nuclear-coded PPM proteins from the ER lumen. The PPM is a novel kind of genetic membrane of chimaeric origin, with some biogenetic features simply inherited from the red algal slave and new ones imposed from outside by the slave owner.

(e) *The chromobiote periplastid reticulum: a relict phospholipid factory?*

Chromobiotes lost the NM, periplastid ribosomes and starch, but retained the PPM. As the 30 NM-encoded chloroplast proteins are essential, their genes must have

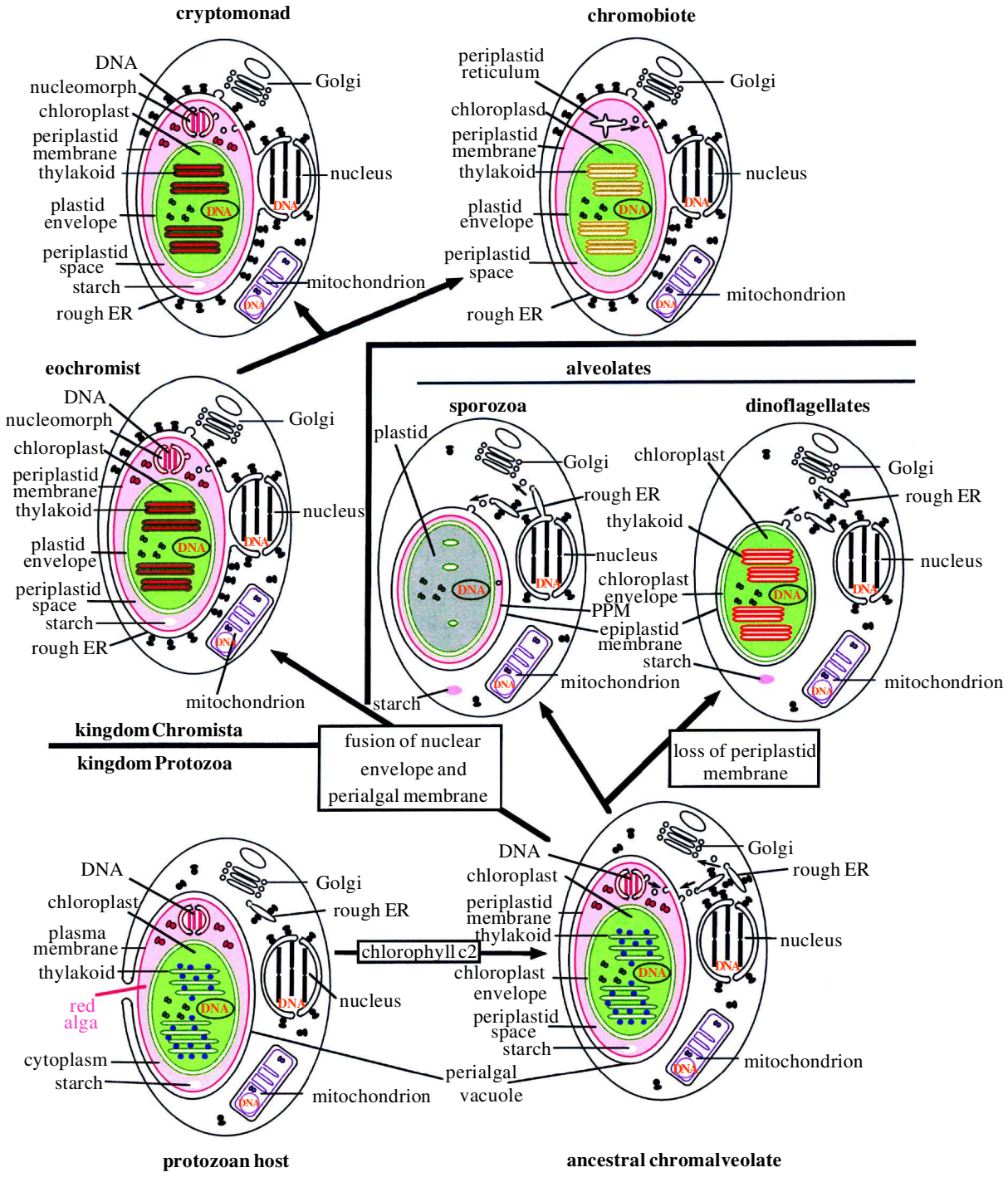


Figure 3. Secondary symbiogenetic origin of chromalveolate cells and the kingdom Chromista. After the red alga was phagocytosed into a food vacuole (phagosome), the host enslaved it by inserting translocators into the surrounding membranes to extract useful molecules. This entailed the evolution of novel protein-import machinery involving four major innovations: (i) addition of N-terminal signal sequences for translocation across the ER membrane to each pre-protein; (ii) a new type of membrane vesicle budding from the ER and the targeting of their fusion with the perialgal vacuole membrane by means of novel SNARE receptors; (iii) a new machinery for import across the periplastid membrane (see figure 4) that recognizes subterminal transit sequences. Symbiont genes transferred to the host nucleus that acquired similar bipartite plastid-targeting sequences were lost from the symbiont nucleus. The red algal mitochondrion and Golgi (not shown) were lost early on, but its nucleus remains in cryptomonads as the NM, because gene transfer to the host nucleus was never completed. The NM genome was lost independently in chromobiotes and alveolates, which managed to transfer all essential red algal nuclear genes to the host, but its membrane remains as the periplastid reticulum. In the ancestor of chromists the phagosomal membrane fused with the nuclear envelope and was colonized by ribosome receptors; this did not occur in alveolates.

moved into the nucleus and retargeted across the ER and PPM in chromobiotes, which probably retained the PPM because they became dependent on it for protein import before losing the nucleus (see below). As the PPM must still grow and divide in synchrony with the plastid and periplastid ER, I suggest that the PPR, found in all chro-

mobiotes but not cryptomonads, makes phospholipids for PPM growth by fusion of vesicles budded from it (figures 1 and 3). Thus, chromobiotes must retain the periplastid targeting machinery for importing acyl synthetase and vesicle budding and fusion proteins, including the periplastid transit peptidase and chaperones, i.e. they should have two Top receptors and Tocm, but will have lost Toc1 from the plastid OM (see figure 4). They also must have a post-translational protein-insertion system to insert acyl transferase and other proteins into the PPR. Probably this evolved in the ancestral chromist as our inability to find a NM SRP-RNA gene suggests that protein insertion into the NM membrane is post-translational (Douglas *et al.* 2001).

This interpretation now clarifies the results of classical experiments using protein synthesis inhibitors on the ochrophyte alga *Ochromonas* (Gibbs 1979). Spectinomycin or chloramphenicol, which inhibit plastid protein synthesis, hypertrophied the PPR, whereas cycloheximide (inhibiting host cytosolic protein synthesis) shrank it. Cycloheximide, by preventing import of PPR precursor proteins, would impede its growth, so it would shrink by continued vesicle transport to the PPM. Conversely, preventing chloroplast protein synthesis would restrict growth of the chloroplast (and indirectly the surrounding PPM) but allow the PPR to continue to accept cytoplasmic proteins imported by Top2 and make more and more phospholipid and so hypertrophy.

Thus, the PPR is probably a unique phospholipid-synthesizing organelle of chromobiotes, homologous with the cryptomonad NM membrane. Although we commonly speak of NM loss in the ancestral chromobiote, it would be more accurate to say loss of its genome and pore complexes. The membrane survives hundreds of millions of years after the genome it once housed was lost; yet another example of the enduring character of membrane heredity, to be added to the multiple origins of hydrogenosomes that retained the double proteobacterial envelope long after the mitochondrial genome was lost (Cavalier-Smith 1987).

(f) *Miozoan plastids and the divergence between chromists and alveolates in targeting biology*

Dinoflagellates are grouped with Sporozoa and protalveolates in the phylum Miozoa, which must have been photosynthetic ancestrally (Cavalier-Smith 1999). The outermost membrane (the EM) surrounding dinoflagellate chloroplasts and sporozoan plastids is apparently not connected to the RER. Thus, chloroplast proteins must be carried by budding from the endomembrane system and inserted into the EM by vesicle fusion. Such vesicle fusion required evolution of a novel SNARE targeting system (McFadden 1999), which dinoflagellates will share, but which is likely to differ in detail from that of cabozoans. Previously, it was assumed that in Miozoa transport vesicles budded from the Golgi as in *Euglena* (Cavalier-Smith 1999; McFadden 1999), but discovery of euglenoid stop-transfer sequences (Sulli *et al.* 1999), absent from miozoan pre-proteins, leads me to suggest that Miozoa follow a more direct route by budding from the ER and bypassing the Golgi (figures 1 and 3). At present, there is no evidence from labelling studies that *Toxoplasma* or *Plasmodium* pre-proteins exist in the Golgi. Furthermore, as neither Brefeldin A, nor adding ER-retrieval sequences, inhibit protein targeting to the apicoplast (Joiner & Roos 2002), this cannot be via the Golgi. Joiner & Roos (2002) argue that failure of ER-retrieval sequences to block targeting implies that the EM is continuous with the ER (i.e. the plastid and PPM are inside the ER as in chromists, including aberrant ones with largely smooth periplastid ER (Ishida *et al.* 2000)). However, they may be misled by the term 'ER retention sequences'; those in question are actually retrieval sequences that function not in the ER but by retrieving soluble ER proteins from the Golgi. If, as I propose here, apicoplast and dinoflagellate plastid proteins are targeted by a novel class of vesicle that buds directly from the ER, totally bypassing the Golgi, ER-retrieval sequences should have no effect on their successful targeting, as there would be no machinery in the apicoplast (unlike the Golgi) to recognize them. In favour of direct ER budding is immunogold evidence that N-ethylmaleamide sensitive factor, which helps associate the attachment proteins (SNAPs) and receptors (SNAREs) that mediate vesicle docking and fusion (Clague & Herrmann 2000), is located at the apicoplast envelope (Hayashi *et al.* 2001). Serial sectioning *Plasmodium* (Hopkins *et al.* 1999) also suggests that the plastid and PPM are not inside the ER lumen.

The PPM was lost in the ancestral dinoflagellate as in euglenoids but retained in Sporozoa. *Plasmodium* and *Toxoplasma* transit peptides differ from those of all other eukaryotes in having very few hydroxylated amino acids; they are particularly rich in leucine, isoleucine and asparagine. This greater hydrophobicity suggests that their translocons, or lipid environment, are unusually divergent. But as plant transit sequences work in *Toxoplasma* they cannot be fundamentally different (Crawford & Roos cited by van Dooren *et al.* 2001). However, as the hydroxylated amino acids are postulated to hydrogen bond with the galactose moiety of galactolipids, Sporozoa possibly lost both galactolipids and sulpholipids when they lost photosynthesis, because maintaining the proper lipid environment for the photosynthetic machinery would cease to be a selective constraint. Few dinoflagellate transit sequences are known (McFadden 1999); although some have hydroxylated amino acids, they are sparse in peridinin-binding protein transit sequences, which are rich in alanine instead.

In contrast to Miozoa, chromist transit sequences are much more similar to those of plants, especially red algae (Deane *et al.* 2000). Their cleavage site sequence is GPXM|XX; as the conserved SPK|AN at the corresponding site in Chlorarachnea is more similar to that of green plants, this is further evidence for a red algal ancestry for chromistan plastids. As Miozoa lack either motif, they have diverged more from the ancestral state. The beginning of the chromist transit sequence has the conserved motif FXP, a shared derived character (absent from red algae) supporting a single origin for the chromistan targeting machinery, their transit sequences are also generally shorter than in red algae. Absence of this motif from Sporozoa (Zuegge *et al.* 2001) means that miozoan targeting diverged more from the ancestral chromalveolate state, possibly as a result of changed lipid composition through loss of photosynthesis. Some dinoflagellate transit sequences have FXP or FXXP motifs suggesting that they arose in the chromalveolate common ancestor.

If Miozoa target plastid proteins by budding vesicles from the ER, not the Golgi (Cavalier-Smith 1999; McFadden 1999), this was probably ancestral for all chromalveolates. We can now understand how the EM fused with the RER in the ancestral chromist. Such ER-derived vesicles would have an inherent capacity to fuse specifically with the EM; if one did not separate properly from the ER but went on nonetheless to fuse with its target membrane this would inevitably merge the membranes into a topological continuum, placing the periplastid complex within the ER lumen (figure 3). Thereafter, budding ER vesicles containing plastid proteins would be energetically wasteful but not lethal: during the transitional stage t-SNAREs would still be present on the epiplastid/RER membrane, leading to futile budding and fusion cycles with the same membrane, eventually suppressed as selection for efficiency favoured defective mutants in key genes. Never undergoing this fusion their alveolate sisters retained the vesicle budding and fusion mechanism; possibly the greater divergence of their transit sequences reflects an additional role in ensuring specificity of ER vesicle budding that led to modifications following the primary chromalveolate divergence.

Retention of both vesicle fusion with the EM and phagotrophy by dinoflagellates probably explains why they alone among chromalveolates have undergone chloroplast replacement. The vesicles were there ready to be used in enslaving the incoming foreign eukaryotic alga.

Plastid genomes of Miozoa are greatly simplified compared with other groups. The sporozoan plastid has very few genes, apparently retained to allow expression of approximately three end-product genes involved in fatty acid synthesis but its genome organization is conventional (Wilson *et al.* 1996). Whereas Sporozoa lost all photosynthetic genes, in dinoflagellates virtually the converse pertains: almost the only identified chloroplast genes, apart from plastid 16S and 23S rRNA genes, are for photosynthetic proteins (Zhang *et al.* 1999; Barbrook & Howe 2000; Barbrook *et al.* 2001; Hiller 2001). These genes are uniquely located on plasmid-sized minicircular chromosomes, usually each bearing only one gene (rarely two). Each minicircle has a non-coding region that undergoes very rapid evolution, concerted among the different minicircles of a species, probably by frequent DNA conversion (Zhang *et al.* 2002). In addition to functional minicircles, dinoflagellates have evolved minute selfish circles with only fragments of formerly functional genes (Zhang *et al.* 2001) or none at all (Hiller 2001).

However, in the dinoflagellate *Prorocentrum* one gene (*psbA*) is on much larger molecules than is customary for minicircles, and no 23S rRNA was detected (Zhang 1999). This raises the possibility that in *Prorocentrum* all chloroplast genes moved to the nucleus and all chloroplast proteins are imported. However, cytologically detectable DNA in *Prorocentrum* plastids argues against this, unless it was completely selfish. As minicircular chromosomes can easily grow larger by dimerization and deletion of replicon origins (Zhang *et al.* 2000), large size of plastid-gene-containing DNA might simply result from secondary joining or expansion of minicircles rather than transfer to the nucleus. I postulate that minicircles arose as a genomic evolutionary quirk of no fundamental significance after most genes present in other plastids moved to the nucleus in the ancestral dinoflagellate; quirky and ultra-fast evolution is often associated with extreme genomic reduction probably because of reduced-strength stabilizing selection (Zhang *et al.* 2000). Dinoflagellate minicircles may have such a high copy number that they do not need a specific segregation mechanism. Even the tiny plastid of the sporozoan *Toxoplasma* has approximately 25 genomes (Matsuzaki *et al.* 2001) and the single apicoplast is segregated by attachment to the centrosome (Striepen *et al.* 2000); its accuracy is essential for viability (He *et al.* 2001).

5. NOVEL SNARES: INITIATORS OF ALGAL ENSLAVEMENT DURING SECONDARY SYMBIOGENESIS

In both chromalveolates and cabozoa, the first key step in evolving chloroplast-specific targeting would have been the evolution of a mechanism for fusing endomembrane vesicles carrying chloroplast pre-proteins with the former food vacuole membrane. The second key innovation (at least for chlorarachneans and chromalveolates) was a mechanism for crossing the PPM. We do not know how this is achieved. Simplest would be the insertion of transit peptide receptors and translocators (the Toc machinery) into the PPM so that a transit sequence could direct the pre-protein across it (Cavalier-Smith 1999; van Dooren *et al.* 2001).

Fusion specificity of endomembrane vesicles depends on receptors on target membranes (t-SNAREs) recognized by complementary ones on the vesicles (v-SNAREs), helped by soluble cofactors (SNAPs). If they also mediate vesicle fusion with the EM, origins of novel SNAREs and SNAPs would have been key primary events in secondary symbiogenesis (McFadden 1999). Gene duplication presumably generated novel t-SNAREs on the former perialgal membrane and v-SNAREs on the endomembrane transport vesicle. Only then could nuclear-encoded proteins with a signal sequence be targeted to the space around the PPM. In principle, the novel transport vesicles might initially bud from the Golgi or from the ER. Apparently, chromalveolates invented novel budding from the ER, I suggest by duplicating copII vesicles, whereas cabozoa hit on budding from the Golgi instead, probably using copI vesicles.

A novel SNARE targeting mechanism to the perialgal vacuole and insertion of Toc machinery into the PPM would have allowed nuclear-coded pre-proteins to enter the chloroplast, but must have occurred independently in the ancestors of chromalveolates and cabozoa. However, a good fraction of each protein would leak into the default secretory pathway and be lost from the cell unless chloroplast proteins were specifically and efficiently included in the new type of transport vesicle, e.g. by transit-peptide binding. Until ways were found to exclude secretory proteins from plastid-destined vesicles, they would also enter the plastid, thus subsequent improvement was needed after the origin of the novel SNAREs.

The selective advantage of this novel vesicle targeting was obviously not to allow protein import into the organelle, which must have evolved later. It was to tap photosynthate by inserting transport proteins in the perialgal vacuole membrane, and subsidiarily to add lipids made

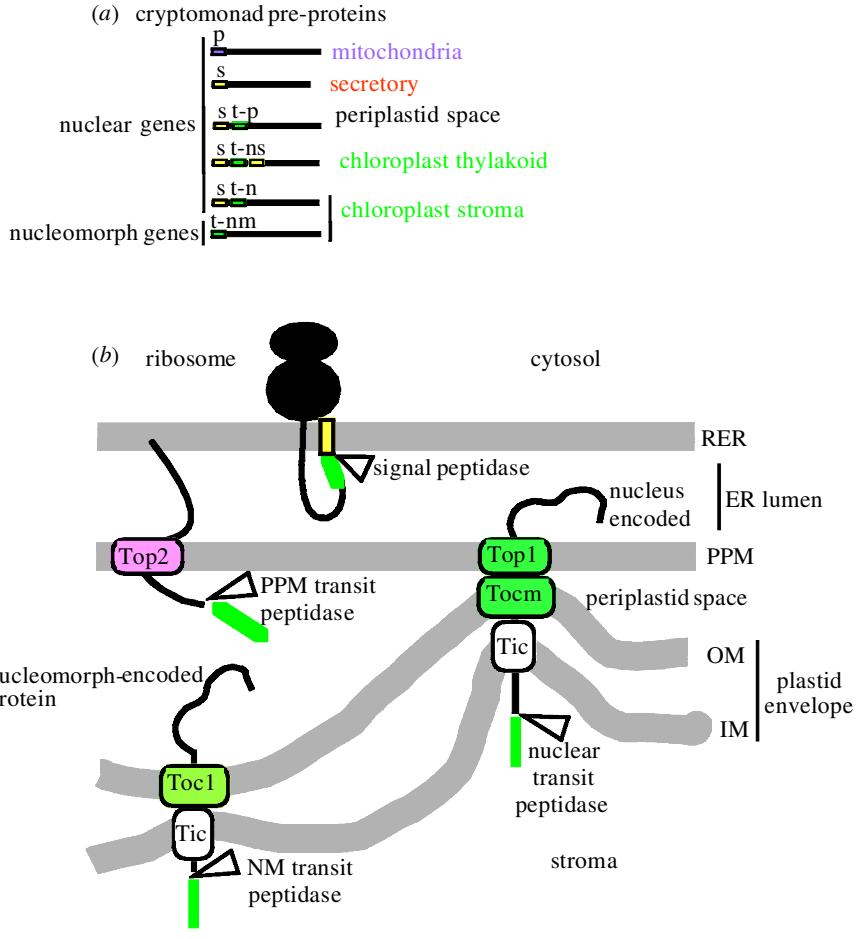


Figure 4. Protein targeting in cryptomonads. See text for explanation.

in the ER to let the perialgal membrane grow. In nature today free-living unicellular algae often leak photosynthate into their environment, as do those cultivated intracellularly by animals (e.g. corals) or protists. As enslavement evolved, carriers for increasing numbers of leaked molecules would be inserted into the perialgal membrane via the ER-derived vesicles. An even more efficient exploitation could occur by inserting host proteins post-translationally into the algal plasma membrane to extract other organic molecules into the ER lumen. This would initiate the conversion of the algal plasma membrane into the chimaeric PPM and could have occurred without evolving specialized insertion machinery, if the difference in lipid composition of the outer leaflet of the plasma membrane and the perialgal membrane's lumenal leaflet were sufficiently different to provide specificity or PPM proteins encoded by the algal nucleus did so.

However, translocating proteins across the PPM into the algal cytoplasm required novel machinery, making this the central, most difficult step in secondary symbiogenesis. Its initial selective advantage must initially have been to add a host protein to the cytosol that distorted the symbiont metabolism to increase the concentration of a key metabolite already efficiently exported through the PPM and perialgal membrane. Thus, the machinery must have evolved to import host proteins. Later it incidentally allowed re-import of chloroplast or other symbiont proteins encoded by genes accidentally transferred to the nucleus.

6. Tocs AND Tops: KEYS TO THE SYMBIONT-TO-ORGANELLE CONVERSION IN SECONDARY SYMBIOGENESIS

To evolve the transport machinery of figure 4, a novel translocon had to be inserted *de novo* into the PPM, most simply by relocating pre-existing Tocs to the PPM (Cavalier-Smith 1999). If the cell already had Toc genes in the nucleus, with signal sequences fortuitously added many Tocs would be present in the ER lumen; eventually one Toc34 might spontaneously self-insert. As Toc34 is self-inserting given suitable lipid composition, accidental vesicle blebbing and fusion might have helped indirectly by inserting an enabling patch of galactolipid membrane. Toc75 could then easily be inserted using its pre-existing transit sequence and Toc159 would follow it.

Suitable chaperones for pulling pre-proteins across the membrane (Zhang & Glaser 2002) would probably have existed in the algal cytosol, e.g. Hsp70. If the original imported protein had to enter the chloroplast, as there would have been no transit peptidase in the algal cytosol, the transit peptide would remain to engage the OM Toc for import into the stroma. Both host and symbiont would have numerous peptidases recruitable as the novel periplastid transit peptidase. There would have been a chicken-and-egg problem if the peptidase were originally of host origin, because it could not enter the periplastid space until it was there to remove its own transit peptide and prevent an onward march into the plastid. Therefore, it

was most probably of symbiont origin—as the symbiont had mitochondrial peptidases that became redundant with the loss of the mitochondria during enslavement one could be recruited by losing its mitochondrial targeting signal or another cytosolic peptidase used instead. Recruiting a non-plastid peptidase might have been sufficient to ensure a different specificity, allowing distinct but related chloroplast and periplastid transit sequences to evolve.

Once the endosymbiotic alga became incapable of independent existence, by loss or mutation of a vital gene not needed inside the host, the symbiont/host became a single unit for selection via the reproductive success of the whole consortium. Whether innovatory mutations beneficial to it arose in former symbiont or host genes would be irrelevant to their success.

There would have been an intermediate stage during establishment of the PPM translocon when host nuclear genes encoded it and OM Toc proteins were still encoded by symbiont (future NM) genes. Thus, they would be free to evolve independently through drift and differential selective forces until the OM version was lost or, I suggest, transferred to the host nucleus. There would almost certainly have been differential selection on them, if only because the OM Toc would continue to interact with Tic through Tic22, whereas the PPM Toc (henceforth Top) would have to interact instead with periplastid chaperones for which it would need new binding properties. Another selective difference is that the Top would have to recognize transit sequences constrained by having to maintain compatibility with an upstream signal peptide and its ER lumenal signal peptidase. Whereas Toc would have to recognize these and the ancestral type of transit sequence with N-terminal methionine (and perhaps subterminal FXN) not the N-terminal FXP of those cleaved from a signal peptidase. Thus Top and Toc would diverge and Toc genes be independently transferred to the nucleus to retain both translocon types.

Efficient transfer of plastid proteins would be favoured if Top evolved specific binding to Toc. As this would be likely to be inimical to efficient recognition of NM-coded transit peptides, Toc itself became duplicated (multiple versions of Toc are known in *Arabidopsis* (Jarvis & Soll 2001) and may function in different cell types). One version (Toc1) would act alone and be specialized for NM proteins, whereas the other (Tocm; m symbolizes its medial position between Top1 and Tic and its modification compared with ancestral Tocs) would be modified for binding the periplastid end of Top. Top function itself could be better optimized by duplication to form Top1 (the putative original Top) for plastid proteins and Top2 for periplastid space proteins.

The ancestral chromobiote lost the Toc1 machinery but retained the Top1/Tocm and Top2 translocons. The alveolates also lost the nuclear genome and one of the Tocs. Sporozoa must have retained the Top1/Tocm translocon but lost Toc1; they apparently have a vestigial PPR, possibly derived from the algal nuclear envelope (independently from Chromobiota); if it has specific proteins they must retain Top2. Dinoflagellates additionally lost the PPM and PPR and thus the Top2 translocon. I suspect they retained the Tocm translocon not Toc1, which was probably lost when the ancestral miozoan lost the algal nuclear genome.

How was the dinoflagellate ancestor able to lose the PPM when no chromobiotes or Sporozoa did? Tocm would eventually become so modified that it could no longer act as a transit peptide receptor in the absence of Top. Possibly the dinoflagellate happened to lose the PPM sufficiently early that Tocm could still function without it, albeit initially inefficiently and needing further improvement. Alternatively, at the time of loss a new chimaeric receptor that could function alone was made from Tocm and Top1 proteins. Either scenario would involve concerted modifications to transit peptides, explaining why they differ considerably from those of chromists, despite a common ancestry. A divergence in properties of Tocm from Top1, eventually made it impossible for Tocm to function as a receptor in the absence of Top1, rendering protein import completely dependent on the PPM so that thereafter it could never be lost. This difficulty of recreating normal Toc or Top was an epigenetic constraint, ensuring that chromobiotes and Sporozoa could never lose this complex four-membrane topology unless they lost plastids altogether, as some did. Thus, historically determined epigenetic constraints associated with membrane growth and protein targeting are the fundamental reasons for the double envelope of most plastids and the PPM of chromalveolates other than dinoflagellates and the fact that the periplastid complex is located within the ER.

(a) *Terminology*

So long as the dinoflagellate ancestor could still import proteins using a modified Tocm no significant modification would be needed for the former perialgal membrane to become, in effect, a third outermost-chloroplast envelope membrane. However, to call this membrane the chloroplast OM in either dinoflagellates or euglenoids, as is usual, is evolutionarily and functionally misleading because it is not homologous with the OM of other plastids. It is better to call it the EM (figures 1 and 3) or simply the outermost membrane. I predict that the middle membrane of dinoflagellate and euglenoid plastids will have galactolipids and sulpholipids but the EM will not and be predominantly phospholipids (plus some sterols). The outermost membrane of Sporozoa, chlorarachneans and euglenoids should also be called the EM, as they are cytologically equivalent.

7. THE CABOZOAN THEORY

As chlorarachnean and euglenoid plastids both originated from green algal chloroplasts, did they have a common origin as postulated by the cabozoan theory (Cavalier-Smith 1999) or originate independently (Delwiche & Palmer 1998)? On sequence trees, chlorarachnean algae nest well within the phylum Cercozoa (Cavalier-Smith & Chao 1997; Ishida *et al.* 1999; Cavalier-Smith 2000*b*), now classified in the protozoan infrakingdom Rhizaria (see figure 2), whereas euglenoids are an ancient group in the phylum Euglenozoa within infrakingdom Excavata (Cavalier-Smith 2002*b*, 2003). In favour of a common origin of their chloroplasts is a tremendous economy in the evolution of their mechanisms for importing nuclear-coded chloroplast proteins (Cavalier-Smith 1999). As in excess of 2000 proteins are imported, at least 2000 fewer mutations would be needed

to attach import signals to each of them. More important is the need to evolve import machinery, which is probably evolutionarily more onerous that adding targeting signals. Eventually, study of their import machinery should show if it has a single origin, but studies of protein targeting in euglenoids are in their infancy (Sulli *et al.* 1999) and in chlorarachneans they have not begun. Lacking a complete plastid genome sequence for any chlorarachnean also impedes interpretation. The few chloroplast or nuclear genes available from both groups are subject to such long-branch problems that we cannot trust their non-grouping together as evidence against a common origin. For example, light-harvesting-complex genes of *Euglena* are so divergent from those of other eukaryotes that they do not even group with green algal genes let alone those of Chlorarachnea (Deane *et al.* 2000).

A common origin of the euglenoid and chlorarachnean chloroplasts would require at least three chloroplast losses within Rhizaria and at least six within excavates, but as several losses can now be inferred within excavates (Andersson & Roger 2002; Cavalier-Smith 2002*b*) this is not a sound reason to reject the idea. As some cytoskeletal evidence suggested that Rhizaria and excavates might have diverged before the origin of the plant kingdom, I recently argued that a sister relationship between Rhizaria and excavates is unlikely and a common photosynthetic origin implausible (Cavalier-Smith 2002*b*). However, the early branching order of ancestrally biciliate eukaryotes (bikonts) is still uncertain (Cavalier-Smith & Chao 2003*a*,*b*) and that argument weaker than I thought. Rhizaria may be paraphyletic and excavates not actually closer to plants and chromalveolates. I now think excavates are probably sisters to the Cercozoa/Retaria clade, here designated 'core rhizarians' or (less likely) to all Rhizaria. The common ancestor of excavates and core Rhizaria could have acquired green algal plastids in one symbiogenetic event, thereby originating an ancestrally photosynthetic 'cabozoan' clade (figure 2): the cabozoan hypothesis (Cavalier-Smith 1999, 2000*a*). As excavates and core rhizarians do form a clade on a recent maximum-likelihood rRNA tree (Cavalier-Smith 2002*b*), it was premature to reject the cabozoan hypothesis. If it is correct, then secondary symbiogenesis created two of the four major clades on the eukaryote tree (figure 2) and many more heterotrophic protists had a temporarily photosynthetic ancestry than commonly thought (Cavalier-Smith 2002*c*).

8. CHLORARACHNEAN CELLS AND GENOMES

(a) *Chlorarachneans are cercozoan algae, the least well-known meta-algal group*

Chlorarachneans are tropical or subtropical unicellular marine algae with green chloroplasts of green algal origin and a unique cell structure very different from green algae (figure 1), some purely photosynthetic (e.g. *Lotharella globosa*), others (e.g. *Chlorarachnion*) also phagotrophic. They are the only algal members of the protozoan phylum Cercozoa (Cavalier-Smith 1998), whose monophyly is supported by protein and RNA trees (Keeling *et al.* 1998; Cavalier-Smith 2000*b*; Vickerman *et al.* 2002). The four named genera differ in ultrastructure of the pyrenoid, a dense proteinaceous aggregate, presumably of RUBISCO as in green algae, filling a pear-shaped evagination of the chloroplast envelope. Photosynthate from the chloroplast is stored in the cytoplasm as β-1,3-glucan within a membrane-bounded organelle appressed to the pyrenoid (cabozoan theory implies homology with the euglenoid β-1,3-glucan paramylum granule membrane). Starch-making machinery was presumably lost from the enslaved green alga during symbiogenesis. The chlorarachnean cenancestor probably had a life cycle with three different cell types: (i) a naked uniciliate dispersal stage; (ii) a benthic feeding and photosynthesizing stage as a filose amoeba; and (iii) a coccoid stage with a protective enveloping wall. Some (*Bigelowiella natans*; Moestrup & Sengco 2001) lost amoeboid and coccoid phases, being planktonic phytoflagellates, whereas *Gymnochlora* retained only the amoeboid phases. Some have lost only the amoeboid phase, (e.g. *L. globosa*). Similar phenotypic diversity is found in non-photosynthetic Cercozoa, ancestrally naked biciliate flagellates, but often with amoeboid phases and ability to encyst.

As benthic chlorarachneans are slow growing, only the flagellate *Bigelowiella* has been used for detailed studies of NM DNA. As in cryptomonads their NMs have three small linear chromosomes, with ends terminated by telomeres (Gilson & McFadden 1995) and a single subterminal rRNA gene cluster, but genome size is less on average (380–455 kb; Gilson & McFadden 1996*a*). Its smallest chromosome (III) is fully sequenced; others are in progress (Gilson & McFadden 2002). The genomes are very compact and gene-rich (approximately one gene per 1141 bp: average space between genes 97 bp on chromosome III) and encode predominantly housekeeping proteins plus a few chloroplast proteins.

(b) *Pygmy introns*

Chlorarachnean NMs have the smallest known nuclear genomes, but have been radically streamlined by losing most genes and virtually all intergenic DNA; yet they retain a high intron density (3.3 introns per kb protein coding sequence). This density, similar to *Arabidopsis*, may be typical of green plants and thus probably similar to their green algal ancestors. Thus, chlorarachnean NMs depleted their non-coding intragenic DNA not by eliminating introns, but by drastically shortening them: they now have only 19 ± 1 nucleotides, the shortest known spliceosomal introns (Gilson & McFadden 1996*b*). The presence of hundreds of pygmy introns in this 'bonsaied' genome can hardly be functional, as the larger cryptomonad NM has only 17 introns; it is an inescapable relic of its evolutionary history as a many-fold larger genome where introns were less burdensome. The NM encodes many spliceosomal RNAs and proteins, which are an additional burden. Possibly the splicing mechanism is simplified compared with other eukaryotes; uniformity in intron size and many ill-spliced transcripts suggests they may rely on a ruler mechanism for excision. If this depends less on intron structure it would have facilitated a marked reduction in intron size—by being less than half the size of spliceosomal introns in any other protist, this would halve the metabolic burden of useless introns. Presumably, shortening introns is mutationally easier than precise excision; such extreme shortening also reduces the selective advantage of total excision.

(c) *Why chlorarachnean nucleomorph genomes are retained*

Logically, NMs should not be retained unless they contribute at least one vital function (Maier *et al.* 2000). As they are only found in organisms with plastids (being absent from *Goniomonas*, the closest plastid-free relative of cryptomonads), it was postulated that NMs should contain at least one gene for an essential protein imported into the chloroplast (McFadden *et al.* 1997). Five such genes have been identified (three ClpPs, an ABC transporter and SecY; Gilson & McFadden 2002): it is unlikely that the total number of NM genes directly useful to the cell (end-product genes: Cavalier-Smith & Beaton (1999)) will exceed 20. No genes encoding non-plastid end-product functions are known. As the nucleus of the green alga would probably have encoded 2000 chloroplast proteins or more, 99% must have been transferred from the NM to the cercozoan host nucleus early on: an example is the light-harvesting complex proteins that bind chlorophyll *a* and *c* (Deane *et al.* 2000). Unlike cryptomonads, there is no evidence that chlorarachnean NMs encode their own tubulins (Keeling *et al.* 1998).

Chlorarachnean NM genomes have therefore been retained for several hundred million years just because they still encode *ca*. 1% of the proteins necessary to make a chloroplast, and the requisite mutations for transferring them to the nucleus have never occurred in the appropriate order. Compared with euglenoids, which managed to transfer this last residue and lose the algal nucleus, they are locked into a less efficient stable state with a large evolutionary burden of several hundreds of other genes, ribosomes and nuclear membranes and division machinery essential only to enable perpetuation and expression of a paltry few end-product genes. This is a superb example of phylogenetic constraint and historical contingency (Gould 1989), refuting the dogma that natural selection can always optimize function or is equivalent to a good designer.

(d) *The source of the chlorarachnean plastids*

The sequence of their light-harvesting complex protein (Deane *et al.* 2000) is distinctly related to that of the green alga *Tetraselmis* and more distantly to *Chlamydomonas* (both class Chlorophyceae; Cavalier-Smith 2000*b*). This is consistent with other sequence evidence for an affinity of the green algal symbiont with the Ulvophyceae (Ishida *et al.* 1997, 1999), as Ulvophyceae and Chlorophyceae are both in the green algal infraphylum Tetraphytae (Cavalier-Smith 1998). Enolase indels have been interpreted as supporting an alternative relationship to streptophytes (charophytes and land plants) (Keeling & Palmer 2001). However, enolase seems prone to lateral transfer, e.g. that of *Euglena* appears to come from a spirochaete (Hannaert *et al.* 2000). Keeling & Palmer (2001) suggested that the two *Arabidopsis* nuclear genes for chloroplast enolase came by lateral transfers from other eukaryotes. However, a gene duplication in a green algal ancestor followed by differential losses of the two types seems more probable, in which case the streptophyte version with the double (one plus five amino acids) insertion might have been present in a tetraphyte lineage at the time of the secondary symbiogenesis. As alveolates share this very enolase double insertion, I suggest that this gene duplication, followed by insertions in one copy, occurred in the common ancestor of red and green algae (almost certainly sisters) and both versions were in the red alga incorporated into the ancestral alveolate and only one version was retained in alveolates and the other in chromists. Instead of three independent lateral transfers (Keeling & Palmer 2001), multiple losses are more likely.

9. EUGLENOID CHLOROPLASTS AND PROTEIN TARGETING

Euglenoids are nutritionally very diverse, only about half being photosynthetic like *Euglena* itself. Many are phagotrophs (petalomonads eat bacteria; peranemids ingest other eukaryotes) or saprotrophs, e.g. rhabdomonads with no trace of plastids and *Astasia*, effectively a *Euglena* that lost photosynthesis but retained a colourless plastid (leucoplast) with reduced genome. Molecular phylogeny reveals that the osmotrophic *Khawkinea*, where plastids are unknown, lost photosynthesis independently of *Astasia* (Linton *et al.* 1999; Mullner *et al.* 2001; Preisfeld *et al.* 2001). As no photosynthetic euglenoids are phagotrophic, their common ancestor must have lost phagotrophy and must have had a photophagotrophic ancestor. Euglenoids are most closely related to the purely heterotrophic kinetoplastid and diplonemid zooflagellates grouped with them as the phylum Euglenozoa by several shared ultrastructural characters (Cavalier-Smith 1981, 1993*b*; Simpson 1997) and a unique trans-splicing of all nuclear pre-messengers to add mini-exons and make mature mRNA (Cavalier-Smith 1993*b*,*c*; Frantz *et al.* 2000).

The sister group to Euglenozoa is probably the heterotrophic amoeboflagellate Percolozoa, these phyla constituting the Discicristata. A common photosynthetic ancestry for Percolozoa and Euglenozoa is strongly supported by shared 6-phosphogluconate dehydrogenase genes, ultimately of cyanobacterial ancestry (Krepinsky *et al.* 2001) but evolutionarily closer to those of other eukaryotic algae (see Andersson & Roger 2002). There are reasonable, but not yet compelling, arguments for the view that the common ancestor of the protozoan infrakingdom Excavata, which includes discicristates with the anaerobic Metamonada and the aerobic Loukozoa, was ancestrally photosynthetic, and that metamonads and Loukozoa also lost chloroplasts.

As euglenoid plastids are bounded by an envelope of three membranes and no NM is present, they have entirely lost the green algal nucleus and one membrane compared with chlorarachneans. The outermost of the triple membranes (the EM) is the recipient of vesicles budding from the Golgi apparatus carrying chloroplast precursor proteins (Sulli *et al.* 1999), strongly showing that it is homologous with the former food vacuole membrane, as postulated by the cabozoan theory (Cavalier-Smith 1999), not the plasma membrane as Gibbs (1978, 1981*b*) postulated. Thus, the former plasma membrane of the green alga was lost (Cavalier-Smith 1999).

The chloroplast genome sequence of *Euglena gracilis* (Hallick *et al.* 1993) showed that euglenoids have lost many chloroplast genes compared with green plants (Martin *et al.* 1998). The secondarily non-photosynthetic *Astasia* has lost many more: the only gene needed for

photosynthesis that remains is the large subunit of RUBISCO (Gockel & Hachtel 2000). Concatenated gene sequences confirm that the *Euglena* chloroplast is of green algal origin as it branches within the green algae (Turmel *et al.* 1999; Lemieux *et al.* 2000). As its closest relatives are the chlorophycean algae *Chlamydomonas* and *Chlorella* (infraphylum Tetraphytae), rather than land plants, the algal partner of the symbiosis was clearly not a charophyte, but probably a tetraphyte, as probably also true for chlorarachneans. Thus, present knowledge of chloroplast genomes, like nuclear gene phylogeny, is broadly consistent with the cabozoan theory that the green algal chloroplast was implanted only once into the common ancestor of Euglenozoa and Cercozoa. A slight counter indication is the chloroplast EF-Tu tree, which places Chlorarachnea as sister to Ulvophyceae and *Euglena* as sister to Chlorophyceae, a different tetraphyte class (Ishida *et al.* 1997). However, as single-gene trees can sometimes be misleading it is premature to reject the cabozoan theory.

The most striking feature of the chloroplast genome of *E. gracilis* (Hallick *et al.* 1993) is the immensely larger number of group II introns (Thompson *et al.* 1997*a*) than in green plants (91) and the presence also of 69 highly modified versions (group III introns; Sheveleva *et al.* 2002) and 15 twintrons: introns within introns (Doetsch *et al.* 1998). Comparative survey suggests that the ancestral euglenoid had many fewer group II introns and twintrons (Thompson *et al.* 1995, 1997*b*), but that group III introns evolved relatively early in euglenoid evolution (Doetsch *et al.* 2001). This fits the idea that such introns are selfish genetic elements that were ancestrally mobile (Cavalier-Smith 1991*b*) and spread by self-insertion, not only into ordinary protein-coding genes, but also into pre-existing introns to form twintrons and into the maturase proteins present in some. Although Doetsch *et al.* (1998) seriously consider the possibility that group III introns were present first in euglenoids, and that their unusually short group II introns (some the shortest known) evolved from them, there is little doubt that group III introns arose from group II introns (Cavalier-Smith 1993*c*) as the latter would have been present in the engulfed green alga, whereas group III introns are unknown outside euglenoids. Group II introns are present in chloroplast genes in *Chlamydomonas* (Rivier *et al.* 2001) and higher plants, so probably entered chloroplasts early in green plant evolution, probably from mitochondria where they were almost certainly present in the ancestral green alga (Turmel *et al.* 2002). As they are unknown in red algal or chromalveolate plastid genomes transfer was probably after green and red algae diverged: the assumption that group II introns were absent in the ancestral euglenoid plastid (Thompson *et al.* 1995; Doetsch *et al.* 1998) is unparsimonious.

The finding that nuclear-coded chloroplast proteins co-translationally inserted into the RER do not pass fully into the lumen, but remain stuck in the membrane, with the bulk of their molecule exposed on its cytosolic face during trans-Golgi transit until after the transport vesicles fuse with the EM (Sulli *et al.* 1999) is of key importance for the origin of euglenoid plastids. This unique membrane-bound state is caused by a hydrophobic stop-transfer sequence downstream from the transit peptide; so N-terminal targeting sequences of stroma-targeted proteins are tripartite in euglenoids not bipartite (signal sequence and transit sequence only) as in chlorarachneans and chromalveolates (Sulli *et al.* 1999). Thus, intermediate import stages are entirely different from the stages in chromalveolates, where chloroplast protein precursors pass completely into the ER lumen. Later stages across the two inner membranes (the chloroplast envelope of the engulfed green alga) are essentially the same as in green plants (Inagaki *et al.* 2000).

10. ORIGIN OF CABOZOAN PROTEIN TARGETING

The cabozoan theory assumed that protein targeting to the chlorarachnean plastid was via the Golgi as was known in euglenoids (Cavalier-Smith 1999). However, the presence of a stop-transfer sequence in the euglenozoan preproteins (Sulli *et al.* 1999), but apparently not in chlorarachneans (Deane *et al.* 2000), implies a significant difference in the targeting mechanism of the two groups. Targeting has not been directly studied in chlorarachneans. However, the presence of an N-terminal signal sequence and absence of a stop-transfer sequence on their pre-proteins (Deane *et al.* 2000) implies that in chlorarachneans like chromalveolates, the pre-protein fully enters the lumen. This substantial difference from euglenoids is compatible with the cabozoan theory.

The key advantage of the cabozoan theory is that it avoids independent origins of protein targeting in euglenoids and chlorarachnean. In both groups, as in chromalveolates, the first key innovation in evolving the chloroplast-specific targeting mechanism would have been a specific SNARE mechanism for fusing endomembrane vesicles carrying chloroplast pre-proteins to the former food vacuole membrane. Thereafter nuclear-encoded proteins with a signal sequence could be targeted to the space around the PPM. The second key step (at least for chlorarachneans) was the evolution of a mechanism for crossing the PPM. I suggest that both these evolutionarily complex innovations occurred in the common ancestor of euglenoids and chlorarachneans, much as in chromalveolates, and that they were followed by the addition of signal sequences to hundreds of (eventually over 2000) former green algal nuclear genes for chloroplast proteins.

The endomembrane vesicles might have budded off indiscriminately from both Golgi and ER, with specificity increased independently in the ancestors of euglenoids and chlorarachneans after they diverged. The fact that euglenoids target plastid proteins via the Golgi means that in marked contrast to chromalveolates they failed to evolve efficient budding from the ER. The simplest explanation would be that chlorarachneans also use the Golgi route and that their transit sequences have a dual role and are recognized by cargo receptors in the Golgi for inclusion in vesicles destined for the EM, as postulated for the analogous case of budding from the ER in alveolates. I suggest that there is a novel coated-vesicle budding machinery in the euglenoid and, probably, chlorarachnean Golgi, and that it evolved from the copI machinery used for retrograde Golgi transport; its proteins should show such homology when discovered. A key question is the role of the euglenoid stop-transfer sequence and why such a sequence is absent in chlorarachneans. Might it be a secondary consequence of the loss of the PPM and its postulated Top receptors serving to hold the pre-protein in the correct orientation for

entry into the former plastid OM? I suggest that even in euglenoids the transit peptide provides the specificity for cargo identification in the Golgi, though protruding C-terminal domains of the pre-proteins and the stop-transfer sequence might serve as subsidiary signals for EM-destined vesicle assembly. The closeness of the Golgi to the base of the pyrenoid in *Chlorarachnion* and electron microscope images of smooth vesicles fit targeting via Golgi. Vesicles from the transitional ER are on the side facing the Golgi, not the plastid envelope as would be expected if they were the source of the vesicles (Hibberd & Norris 1984). We cannot currently exclude the less parsimonious possibility that chlorarachneans use ER budding instead, in which case their chloroplast vesicle-budding machinery might have evolved instead from copII vesicles that bud from ER for anterograde transport to the Golgi.

Thus, the cabozoan theory expects euglenoids and chlorarachneans to have homologous SNARE machinery for targeting chloroplast transport vesicles to the EM. They probably also have the same coated-vesicle budding machinery, but might have evolved different coated-vesicle budding machinery from the Golgi and ER, respectively.

At some stage, the euglenoid lineage alone must have lost the NM, which could only occur after all essential NM-coded chloroplast proteins successfully moved to the nucleus and retargeted to the plastid using the novel machinery. The plasma membrane of the green alga could have been lost quite early, but on the cabozoan theory only after the divergence from chlorarachneans and probably after the initial origin of vesicle targeting by SNAREs. As the latter would be unlikely to be perfected without a means for crossing the PPM, it is likely that the Toc protein inserted into the PPM before membrane loss. If at the time of loss there was a functional Toc in both the PPM and former OM, then from the point of view of membrane targeting either membrane could have been the one that was lost. However, as the OM would have an established mechanism of growth and interaction with the inner membrane, as well as porins and other transporters important for small molecule exchange, whereas the plasma membrane would not, it is inherently more likely that it would be retained and the plasma membrane lost. The insertion of the majority of the OM proteins probably depends on its unique lipid composition with digalactosyl and sulpholipids (Schleiff *et al.* 2001). As the plasma membrane would not have this, it could not substitute for the OM merely by acquiring Tocs. The probability that it could acquire the unique membrane composition and asymmetry if it never had it before is minuscule, and the chance that this could occur at the same time as the loss of the OM is even less. Thus, the cyanobacterial OM has probably never been lost in the history of plastids for reasons of its complex biogenesis (Cavalier-Smith 1982*a*). If the plasma membrane were lost very early, there would not be any need to have evolved an import mechanism across it, but as such a mechanism was essential for the chlorarachneans there is no gain in parsimony through assuming such an early loss.

11. CONVERGENCE IN 'BONSAIED' CHROMOSOME ORGANIZATION

Since the discovery that cryptomonad and chlorarachnean NMs both have precisely three minute chromosomes terminated in rRNA genes (Eschbach *et al.* 1991; Maier *et al.* 1991; Gilson & McFadden 1996*a*,*b*) this coincidence has been intriguing. The possibility that they share a common origin (Cavalier-Smith *et al.* 1994) is excluded by unequivocal phylogenetic evidence for separate origins within unicellular red algae and tetraphyte green algae, respectively (Cavalier-Smith *et al.* 1996; van der Peer *et al.* 1996). Therefore, they have converged on a similar structure. The opposite orientation of the rRNA genes in the two groups supports convergence.

As nuclear chromosome number is generally so variable, why should NMs all have exactly three? As chromosomes below *ca.* 100 kb may be too readily lost by mitosis (Murray & Szostak 1985), the *Bigelowiella* sizes (98, 140, 145 kb) may be almost as low as practically possible, explaining why they have not fragmented into even smaller ones. But why could they not have aggregated into just one or two? If this happened, the chromosomes would be longer. This would be no problem for normal eukaryotic chromosomes, which can fold their basic 30 nm chromatin threads into highly compact structures easily segregated at mitosis. As there is no evidence for compact chromosomes during NM division (McKerracher & Gibbs 1982; Morrall & Greenwood 1982; Meyer 1987), we suggested that NM chromosomes lost the ability for such higher-level folding and participate in mitosis simply as 30 nm threads with metacentric chromosomes. Even without higher-order folding each chromosome arm would fit into the small dimensions of dividing NMs (Douglas *et al.* 2001), but if their length doubled or tripled by chromosome aggregation spindles would have to be much longer, which would be inefficient. This interesting constraint on minute eukaryotic genomes arises because the length of chromatin threads scales directly with genome size, whereas the diameter of the nucleus scales with its cube root.

The highly reduced genome of the microsporidia, energy parasites on their obligate hosts, not energy donors like the enslaved algae, also shows remarkable convergences with NM genomes (Katinka *et al.* 2001). They also have rRNA genes at each end of every chromosome as well as repeated subterminal sequences. However, in other small eukaryotic genomes such as *Giardia*, *Saccharomyces*, *Schizosaccharomyces* and *Plasmodium*, not every chromosome has rRNA genes at both telomeres. What seems universal is the presence of terminal and subterminal sequences repeated on several chromosomes, and a high propensity for recombination in these regions. One can surmise that the general occurrence of the Rabl chromosome configuration after mitosis, and the clustering of telomeres in the bouquet phase of meiotic prophase in those nuclei that do undergo meiosis (there is no evidence that NMs do), plus the high propensity for subterminal recombination that generates all these repeats, makes it easy to maintain them identically by gene conversion.

Miniaturization of the NM genome even influences polypeptide length, which is shorter for many proteins, both in NMs (Archibald *et al.* 2001; Douglas *et al.* 2001) and microsporidia (Katinka *et al.* 2001). Another convergent similarity between microsporidian and NM genomes is the greatly elevated evolutionary rate of many, but not all, genes (Archibald *et al.* 2001; Douglas *et al.* 2001; Katinka *et al.* 2001).

12. NUCLEOMORPH MINIATURIZATION: IMPLICATIONS FOR THE UNIVERSAL FUNCTION OF SECONDARY DNA

Scaling is also important for understanding the evolution of nuclear genome size. The contrast between the NM's extreme miniaturization with, essentially, no non-coding DNA, and the nucleus with masses of non-coding DNA, strongly indicates that selection rather than mutation pressure dominates evolution of genome size and must differentially affect these two nuclei that have been co-evolving in the same cell for about 500 Myr (Cavalier-Smith & Beaton 1999). The successful elimination of virtually all non-coding DNA refutes the selfish and junk DNA theories of an inexorable upward mutation pressure on genomes that selection is powerless to resist. The positive correlation between nuclear genome size and cell size is basically the same in cryptomonads as in all other eukaryotes (Beaton & Cavalier-Smith 1999) and is simply explained by the idea that chromatin has a bulk skeletal function in nuclear assembly (for which there is solid experimental evidence; Cavalier-Smith 1982*b*). The size of the nucleus is determined by its folding pattern and genome size, logically inescapable given the attachment of the nuclear envelope to the chromatin surface. To this one need only add the established cell biological principle of balanced growth to argue that the size of the nucleus (housing the RNA synthesis machinery) relative to that of the cytoplasm (housing the protein synthesis machinery) must be of functional importance, as argued by early students of the karyoplasmic ratio (Strasburger 1893; Trombetta 1942). Skeletal DNA theory argues that this skeletal function, coupled with the principle of balanced growth, provide sufficient explanation for the universal correlation between genome size, nuclear size and cell size (Cavalier-Smith 1978, 1991*c*).

One mystery about eukaryotic genome-size evolution is why geneticists and molecular biologists so often overlook these basic cell biological principles. For example, Petrov (2001) suggested that ignoring the selective/functional reasons for different genome sizes and concentrating instead on the mutations that increase or decrease them would more rapidly advance understanding of the 'problem'. But this is not so, as it avoids the key issue. It is trivially true that genomes increase or decrease in size by mutation (duplications, insertions and deletions). Mutations are the primary cause of all evolution, but we also need to know whether selective forces favour or impede their spread. Merely enumerating accepted deletions or duplications in different groups does not tell us whether mutation pressure or selection is the decisive force and so does not tell us *why* some species have much larger genomes than others. Equating observed evolutionary acceptance of deletions with the mutation rate at which deletions occur is as conceptually confused as muddling observed nucleotide substitutions with actual mutation rates. It does not follow at all from Petrov's observations that downward mutation pressure rather than random mutation plus selection for smaller genomes has reduced *Drosophila* genome sizes. To assume that it has begs the whole question, curiously in the opposite direction from earlier selfish DNA proponents who assumed that mutation pressure was always upwards. Both cannot be right!

The correlation of genome size and cell size cannot be explained by simultaneously invoking both upward and downward mutation pressure as the dominant forces—unless one also assumes that genome size directly determines cell size (Commoner 1964; Bennett 1973; Gregory 2001), which I do not favour (Cavalier-Smith 1985*a*). Although this possibility cannot be eliminated, if it were true this would be a function for non-coding DNA: one could then argue that genome size depended directly on selection for cell size—for cell size like animal body size is not selectively neutral. Thus, a pure mutation pressure theory of the evolution of genome size (Petrov 2002), with no role for cell-size-related selection is untenable. We have to accept that overall genome size is functionally significant and that non-coding DNA has a bulk (almost certainly sequence-independent) function in the genetic control of at least one cellular parameter of profound significance. The key question is whether genome size directly affects only nuclear volume, as argued by the skeletal theory, or also physiologically controls cell size, as others postulated (Commoner 1964; Bennett 1973; Gregory 2001). As DNA is invariably attached to the nuclear envelope in interphase, nuclear volume is, necessarily, physically related to DNA amount and its folding pattern. But this is not true of cell size, for which we know there is genic control (Nurse 1985); possibly genic control is sufficient?

In contrast to the nucleus, NM genome size does not scale with cell size (Beaton & Cavalier-Smith 1999). This is to be expected on the theory of balanced growth because the NM codes for so few end-product genes (the 30 proteins imported into the plastid and a very few others). Most NM genes code for housekeeping functions needed just for perpetuating the NM genome and expressing the end-product genes; neither their transcription levels nor those of many end-product genes (e.g. *ftsZ*, *tic22*) would be expected to scale with overall cell size. Essentially, because of the dramatic reduction in NM end-product genes compared with those of the nucleus, the NM volume allowed by its coding DNA is more than enough to provide space for transcription and RNA-processing machinery without any non-coding DNA (i.e. exclusively skeletal or S-DNA; Cavalier-Smith 1978) to increase its bulk in even in the largest cryptomonad cells.

Gilson & McFadden (2002) question our interpretation on the grounds that one would not expect NM size to scale with overall cell size but with that of the periplastid space. Oddly, they seem not to realize that this is exactly what we are saying! The whole point of the idea of balanced growth is that transcription rates and translation rates must balance. The key assumption behind the skeletal theory is that in a steady state both rates are related to the amount of the two kinds of machinery. Furthermore, since each machine occupies space, the overall volume devoted to both must be kept, essentially, in proportion. Thus, throughout evolution cells of vastly differing volumes will adjust their volume of RNA-synthetically active nuclear space to protein-synthetically active space. It is the active cytoplasmic space that matters on the balanced growth theory, not overall cell volume (Cavalier-Smith 1985*a*). This is beautifully exemplified by

cryptomonads, where the volume of the former red algal cell (the PPM and its contents) is immensely greater than the bulk of the periplasmic space because the chloroplast is so huge. Probably in photosynthetic cryptomonads the former red algal cell occupies as big a volume as the former host cell, yet it has a far smaller genome. There is nothing special about cell size and genome size *per se*; they should not be treated as disembodied variables but as physical entities that occupy real space. Cell volume is used in the skeletal DNA theory as a convenient surrogate for active protein synthetic cytoplasm, just as nuclear volume or genome size is for biosynthetically active nuclear volume. However, a precise physical theory of the volume scaling laws of cell organelles must take note of the detailed physical complexity and heterogeneity of both cytoplasm and nucleus. DNA sequencing may be fashionable but cannot solve the problem, whereas quantitative studies of scaling of cell biological and growth rate parameters and of the genetic control of cell and nuclear size, and interdisciplinary thinking are unfashionable and widely ignored but will increase understanding much more.

13. GENE-TRANSFER RATCHET, SELECTION FOR ECONOMY, AND STABILIZING SELECTION

Gene transfer from mitochondria, chloroplasts and NMs to the host nucleus was massive during the early establishment of these organelles, but since then has dramatically slowed. Probably at least 2000 NM genes moved to the nucleus independently in chromalveolates and cabozoans. Thus, secondary symbiogenesis involved roughly the same number as the origins of mitochondria and plastids together. Were these at least 8000 gene transfers driven primarily by mutation pressure or by a combination of mutation and selection? In principle, selection need not have been involved, for if transfer to the nucleus were dramatically easier than the reverse this would constitute a gene transfer ratchet. Biased interorganellar transfer rates plus random loss might be sufficient. But is transfer to the nucleus easier? The facts that plant mitochondria have taken up chloroplast genes (Wintz *et al.* 1988), that the apicoplast genome probably incorporated some mitochondrial genes (Obornik *et al.* 2002) and that octocoral mitochondria apparently acquired a nuclear MutS gene (Pont-Kingdon *et al.* 1998) show that organelles can acquire DNA by gene transfer. However, the porosity of the nuclear envelope to small DNA fragments, and the multiplicity of mitochondria and chloroplasts allowing some to burst and liberate DNA without lethality, both make DNA transfer into the nucleus easier. It is less easy, however, to argue that NM DNA would move across the PPM and ER membranes dramatically more readily than the reverse and having at most two NMs per cell does not dramatically load the dice.

In addition to relative ease of transport one must consider relative survival chances before accidental loss or inactivation by mutation. Effective transfer coupled with continued function in the organelle requires that the gene acquires a targeting sequence (mitochondrial presequence, transit peptide or signal sequence for NM genes). A nuclear gene for a non-organellar protein being transferred into a plastid or mitochondrion or NM could not function in its original compartment unless its protein were exported out of the organelle. As there are no general protein export mechanisms for any of these organelles there will, essentially, be no successful transfers in that direction retaining the original site of function. Thus, it is the existence of protein-import machinery into mitochondria, plastids and the periplastid space that fundamentally biases evolution towards organelle to nucleus gene transfers. The gene transfer ratchet lies in these protein-import mechanisms rather than in the ease of physical transfer of DNA (Doolittle 1998), which may be a relatively trivial factor. Thus, the ratchet is not a simple mutational one, but the consequence of mutation pressure plus a particular cellular organization, which acts as an epigenetic constraint to channel mutation pressure.

But that is not the whole story. The shrinkage of these three types of organelle genome is not purely a matter of mutation pressure and epigenetic constraint. Selection is also involved. Organellar replicative competition has been frequently invoked as a selective force that will keep mitochondrial genomes small and eliminate non-coding DNA (Cosmides & Tooby 1981). Even for mitochondria, this is probably insufficient explanation. Secondary expansion of non-coding DNA in higher plant mitochondria can be attributed to a weakening of stabilizing selection against useless DNA (Cavalier-Smith 1985*b*). To attribute NM genome reduction to replicative competition (Gilson & McFadden 2002) is implausible as there is usually only one NM per cell, and even when there are two, for example before division and in one small cryptomonad clade (Deane *et al.* 2002), their genes are probably not genuinely in replicative competition as mitochondrial and chloroplast genomes may be. Cryptomonad NMs encode Mcm2 which mediates the G1/S transition and a cyclin-dependent kinase, indicating that their replication and cell-cycle control is fundamentally eukaryotic and therefore that all replicons probably initiate replication synchronously. They are so short that they can easily complete replication within the S-phase. Because of these different cell cycle controls, replicative competition is unlikely to be important in keeping eukaryotic genome size low (Cavalier-Smith 1985*c*).

I previously argued that the most important selective force favouring genome reduction in mitochondria and plastids was selection for economy (Cavalier-Smith 1985*b*, 1987). This is such a generally important selective force that it is puzzling why it is so seldom considered, whereas numerous implausible fancy mechanisms are discussed at length. Selosse *et al.* (2001) are virtually the only other authors to recognize the basic point that these other mechanisms all apply only to genes and therefore do not explain why non-coding DNA is also lost from the organelle (replicative competition could, but, as it probably does not apply to NMs, it cannot be the general mechanism and may never have been the decisive one). Muller's ratchet, sometimes suggested as a factor in organelle genome reduction (Andersson & Kurland 1998; Martin & Herrmann 1998), is irrelevant to unicellular algae because of their immense populations.

Putting organelle genes in the nucleus economizes simultaneously on materials, energy and space, all scarce resources. There are two reasons for this. First, the roughly two orders of magnitude lower copy number of

nuclear genes means they cost much less in resources to replicate and occupy much less space. One hundred copies of 4000 mitochondrial/plastid genes would occupy 4.3×10^6 nm^3 or 0.004 mm^3 without any proteins or transcription machinery or water, space that is better devoted to organelle enzymes. Second, according to skeletal DNA theory a cell of a certain size must have nuclei of a given volume and therefore a set genome size, irrespective of the number of genes its DNA contains; therefore, transferring genes to the nucleus saves all their replicative cost, as they do not increase the genome size, but merely replace S-DNA by G-DNA. Genes can serve as a skeleton just as well as pure S-DNA, so it is obviously better to kill two birds with one stone and function simultaneously as a gene and a skeleton. Thus, gene transfer to the nucleus economizes not only on genes but also on purely skeletal DNA. Coupled with the inevitable gene-transfer ratchet arising from the absence of a protein export machinery from the organelles, it is hardly surprising that most genes that could be were transferred to the nucleus relatively quickly and that the organelle genomes were dramatically pared down early in their evolution.

When most genes were transferred, the selective advantage of transfer of the refractory residue was relatively small, but there is no reason to think that failure of transfer in cryptomonad or chlorarachnean NMs has a selective advantage: it was probably just an accidental failure of all the requisite mutations to occur in those two lineages. Chromobiotes, alveolates and euglenoids managed it independently. Mitochondria managed 100% transfer several times over when they became hydrogenosomes and lost their proteobacterial inner membrane proteins—and chloroplasts probably would have done too if they had been able to re-import all their photosynthetic proteins (Cavalier-Smith 2000*a*). It was just an accident that chlorarachneans and cryptomonads did not—but a lucky accident for us, as it has revealed so much about the role of secondary symbiogenesis in the history of life and the functional nature of most DNA in the biosphere: nuclear non-coding DNA.

14. SOME GENERAL CONCLUSIONS FROM SECONDARY SYMBIOGENESIS

(i) Retention of localization of ancestral functions has played a major part, but the location of genes is much less stable than the location of proteins and lipids. Chloroplast organization has been remarkably constant, despite the movement of thousands of genes first to the nucleus and then from the NM to the host nucleus. However, some novel functions are acquired by retargeting to different compartments; very probably the cytoplasmic starch-making machinery of Miozoa came from the red algal symbiont as a result of gene transfer to the nucleus without re-targeting to the periplastid space as occurred in cryptomonads.

(ii) None the less some functionally trivial replacement or addition of equivalent proteins has occurred. This works both ways: from host to symbiont, e.g. the retargeting of a variant of the cytosolic GAPDH to replace the red algal plastid version (Fast *et al.* 2001) and from symbiont to host, as the acquisition of an actin by the cryptomonad host from the red algal symbiont (Stibitz *et al.* 2000). When there is functional redundancy between compartments, there is no fundamental requirement that a gene be retargeted to its original compartment after gene transfer to the nucleus (Cavalier-Smith 1990); such replacements may be more common than sometimes thought, especially for water-soluble proteins.

(iii) Membrane heredity is central to secondary symbiogenesis. Whereas symbiogenesis provides a cell with several novel genetic membranes in one fell swoop, integrating them into the existing cell to make a chimaeric organism requires the origin of novel protein-targeting machinery, for which proteins are typically recruited from both host and symbiont.

(iv) Loss of redundant function, e.g. of mitochondria and Golgi of the algal endosymbionts and often fatty acid biosynthesis of the host (Cavalier-Smith 2000*a*), has been equally important. Photosynthesis has been repeatedly lost in chromalveolates and cabozoa, as in plants; unlike plants, both secondary symbiogenetic groups seem to have lost plastids completely several times. However, pedinellid chromists *Ciliophrys* and *Pteridomonas*, previously thought to lack plastids primitively (Smith & Patterson 1986) or secondarily (Cavalier-Smith *et al.* 1995), have actually retained leucoplasts, PPM and RUBISCO (Sekiguchi *et al.* 2002). Possibly, therefore, all Ochrophyta depend on plastids for fatty acid synthesis, like plants. If oomycetes genuinely lack plastids, their 6-phosphogluconate dehydrogenase (Andersson & Roger 2002) would originally exemplify plastid proteins of secondary symbiogenetic origin now in the cytosol; however, might oomycetes have genome-free plastid relics analogous to the mitosomes of *Entamoeba* that are all that remain of mitochondria (Tovar *et al.* 1999)? My interpretation of the PPR is an analogous example of the retention of nuclear membranes after the genome was lost.

It used to be argued that because cryptomonads and chlorarachneans still have NMs they must have arisen more recently then other meta-algae that lost them. But this assumed an inexorable and uniform rate of gene transfer. Very little in cell evolution follows such an assumption of uniform rates. Much more typical is dramatically rapid early change followed by hundreds of millions of years of relatively trivial fiddling around with details (Cavalier-Smith 2002*a*). Gene transfer in symbiogenesis, both primary and secondary, clearly follows this pattern. The 1–2% of genes remaining in NMs are not recent intermediates but products of a half-billion-year-old frozen accident, locked in metastable state—just like chloroplast and mitochondrial genomes.

I thank NERC for a Professorial Fellowship and research grant.

REFERENCES

Andersson, S. G. & Kurland, C. G. 1998 Ancient and recent horizontal transfer events: the origins of mitochondria. *APMIS Suppl.* **84**, 5–14.

Andersson, J. A. & Roger, A. J. 2002 A cyanobacterial gene in non-photosynthetic protests—an early chloroplast acquisition on eukaryotes? *Curr. Biol.* **12**, 115–119.

Archibald, J., Cavalier-Smith, T., Zauner, M., Maier, U.-G. & Douglas, S. 2001 Molecular chaperones encoded by a reduced nucleus—the cryptomonad nucleomorph. *J. Mol. Evol.* **52**, 490–501.

Barbrook, A. C. & Howe, C. J. 2000 Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*. *Mol. Gen. Genet.* **263**, 152–158.

Barbrook, A. C., Symington, H., Nisbet, R. E., Larkum, A. & Howe, C. J. 2001 Organisation and expression of the plastid genome of the dinoflagellate *Amphidinium operculatum*. *Mol. Genet. Genom.* **266**, 632–638.

Beaton, M. J. & Cavalier-Smith, T. 1999 Nuclear non-coding DNA is functional: evidence from differential scaling of cryptomonad genomes. *Proc. R. Soc. Lond.* B **266**, 2053–2060. (DOI 10.1098/rspb.1999.0886.)

Bennett, M. D. 1973 Nuclear characters in plants. *Brookhaven Symp. Biol.* **25**, 344–366.

Bruce, B. D. 2001 The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochim. Biophys. Acta* **1541**, 2–21.

Cavalier-Smith, T. 1978 Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J. Cell Sci.* **34**, 247–278.

Cavalier-Smith, T. 1981 Eukaryote kingdoms: seven or nine? *BioSystems* **14**, 461–481.

Cavalier-Smith, T. 1982*a* The origins of plastids. *Biol. J. Linn. Soc.* **17**, 289–306.

Cavalier-Smith, T. 1982*b* Skeletal DNA and the evolution of genome size. *A. Rev. Biophys. Bioeng.* **11**, 273–302.

Cavalier-Smith, T. 1983 Endosymbiotic origin of the mitochondrial envelope. In *Endocytobiology II* (ed. W. Schwemmler & H. E. A. Schenk), pp. 265–279. Berlin: de Gruyter.

Cavalier-Smith, T. (ed.) 1985*a* Cell volume and the evolution of eukaryotic genome size. In *The evolution of genome size*, pp. 211–251. Chichester: Wiley.

Cavalier-Smith, T. (ed.) 1985*b* Introduction: the evolutionary significance of genome size. In *The evolution of genome size*, pp. 1–36. Chichester: Wiley.

Cavalier-Smith, T. (ed.) 1985*c* DNA replication and the evolution of genome size. In *The evolution of genome size*, pp. 1–36. Chichester: Wiley.

Cavalier-Smith, T. 1986 The kingdom Chromista: origin and systematics. In *Progress in phycological research*, vol. 4 (ed. F. E. Round & D. J. Chapman), pp. 309–347. Bristol: Biopress Ltd.

Cavalier-Smith, T. 1987 The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. *Ann. NY Acad. Sci.* **503**, 55–71.

Cavalier-Smith, T. 1990 Symbiotic origin of peroxisomes. In *Endocytobiology IV* (ed. P. Nardon, V. Gianinazzi-Pearson, A. M. Grenier, L. Margulis & D. C. Smith), pp. 515–521. Paris: Institut National de la Recherche Agronomique.

Cavalier-Smith, T. 1991*a* The evolution of prokaryotic and eukaryotic cells. In *Fundamentals of medical cell biology*, vol. I (ed. G. E. Bittar), pp. 217–272. Greenwich, CN: JAI Press.

Cavalier-Smith, T. 1991*b* Intron phylogeny: a new hypothesis. *Trends Genet.* **7**, 145–148.

Cavalier-Smith, T. 1991*c* Coevolution of vertebrate genome, cell and nuclear sizes. In *Symposium on the evolution of terrestrial vertebrates*, Selected Symp. Monogr. U.Z.I., 4 (ed. G. Ghiara), pp. 51–88. Modena, Italy: Muchi.

Cavalier-Smith, T. 1993*a* The origin, losses and gains of chloroplasts. In *Origin of plastids: symbiogenesis, prochlorophytes and the origins of chloroplasts* (ed. R. A. Lewin), pp. 291–348. New York: Chapman & Hall.

Cavalier-Smith, T. 1993*b* Kingdom Protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994.

Cavalier-Smith, T. 1993*c* Evolution of the eukaryotic genome. In *The eukaryotic genome* (ed. P. Broda, S. G. Oliver & P. Sims), pp. 333–385. Cambridge University Press.

Cavalier-Smith, T. 1994 Origin and relationships of Haptophyta. In *The Haptophyte algae* (ed. J. C. Green & B. S. C. Leadbeater), pp. 413–435. Oxford: Clarendon Press.

Cavalier-Smith, T. 1995 Membrane heredity, symbiogenesis, and the multiple origins of algae. In *Biodiversity and evolution* (ed. R. Arai, M. Kato & Y. Doi), pp. 75–114. Tokyo: The National Science Museum Foundation.

Cavalier-Smith, T. 1997 Sagenista and Bigyra, two phyla of heterotrophic heterokont chromists. *Archiv. Protistenk.* **148**, 253–267.

Cavalier-Smith, T. 1998 A revised six-kingdom system of life. *Biol. Rev.* **73**, 203–266.

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. 2000*a* Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5**, 174–182.

Cavalier-Smith, T. 2000*b* Flagellate megaevolution: the basis for eukaryote diversification. In *The flagellates* (ed. J. R. Green & B. S. C. Leadbeater), pp. 361–390. London: Taylor & Francis.

Cavalier-Smith, T. 2002*a* The neomuran origin of archaebacteria, the negibacterial root of the universal tree, and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* **52**, 7–76.

Cavalier-Smith, T. 2002*b* The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297–354.

Cavalier-Smith, T. 2002*c* Chloroplast evolution: secondary symbiogenesis and multiple losses. *Curr. Biol.* **22**, R62–R64.

Cavalier-Smith, T. 2003 The excavate protozoan phyla Metamonada Grassé emend. (Anaeromonadea, Parabasalia and Eopharyngia) and Loukozoa emend. (Jakobea, *Malawimonas* and Diphylleida): their evolutionary affinities and new higher taxa. *Int. J. Syst. Evol. Microbiol.* (Submitted.)

Cavalier-Smith, T. & Beaton, M. J. 1999 The skeletal function of non-genic nuclear DNA: new evidence from ancient cell chimaeras. *Genetica* **106**, 3–13.

Cavalier-Smith, T. & Chao, E. E. 1997 Sarcomonad ribosomal RNA sequences, rhizopod phylogeny, and the origin of euglyphid amoebae. *Archiv f. Protistenkunde.* **147**, 227–236.

Cavalier-Smith, T. & Chao, E. E. 2003*a* Phylogeny of Choanozoa, Apusozoa and other Protozoa and early eukaryote megaevolution. *J. Mol. Evol.* (In the press.)

Cavalier-Smith, T. & Chao, E. E. 2003*b* Molecular phylogeny of centrohelid heliozoa, a novel lineage of bikont eukaryotes that arose by ciliary loss. *J. Mol. Evol.* (In the press.)

Cavalier-Smith, T. & Lee, J. J. 1985 Protozoa as hosts for endosymbioses and the conversion of symbionts into organelles. *J. Protozool.* **32**, 376–379.

Cavalier-Smith, T., Allsopp, M. T. E. P. & Chao, E. E. 1994 Chimeric conundra: are nucleomorphs and chromists monophyletic or polyphyletic? *Proc. Natl Acad. Sci. USA* **91**, 11 368–11 372.

Cavalier-Smith, T., Chao, E. E. & Allsopp, M. T. E. P. 1995 Ribosomal RNA evidence for chloroplast loss within Heterokonta: pedinellid relationships and a revised classification of ochristan algae. *Arch. Protistenk.* **145**, 209–220.

Cavalier-Smith, T., Couch, J. A., Thorsteinsen, K. E., Gilson, P., Deane, J., Hill, D. A. & McFadden, G. I. 1996 Crypto-

monad nuclear and nucleomorph 18S rRNA phylogeny. *Eur. J. Phycol.* **31**, 315–328.
Christensen, T. 1989 The Chromophyta, past and present. In *The chromophyte algae: problems and perspectives* (ed. J. C. Green, B. S. C. Leadbeater & W. C. Diver), pp. 1–12. Oxford University Press.
Clague, M. J. & Herrmann, A. 2000 Membrane transport: deciphering fusion. *Curr. Biol.* **10**, R750–R752.
Commoner, B. 1964 Roles of deoxyribonucleic acid in inheritance. *Nature* **202**, 960–968.
Cosmides, L. M. & Tooby, J. 1981 Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* **89**, 83–129.
Deane, J. A., Fraunholz, M., Su, V., Maier, U.-G., Martin, W., Durnford, D. G. & McFadden, G. I. 2000 Evidence for nucleomorph to host nucleus gene transfer: light-harvesting complex proteins from cryptomonads and chlorarachniophytes. *Protist* **151**, 239–252.
Deane, J. A., Strachen, I. M., Saunders, G. W., Hill, D. R. A. & McFadden, G. I. 2002 Cryptomonad evolution: nuclear 18S and DNA phylogeny versus cell morphology and pigmentation. *J. Phycol.* **38**, 1–10.
Delwiche, C. F. & Palmer, J. D. 1998 The origin of plastids and their spread via secondary endosymbiosis. *Plant Syst. Evol. Suppl.* **11**, 51–86.
Doetsch, N. A., Thompson, M. D. & Hallick, R. B. 1998 A maturase-encoding group III twintron is conserved in deeply rooted euglenoid species: are group III introns the chicken or the egg? *Mol. Biol. Evol.* **15**, 76–86.
Doetsch, N. A., Thompson, M. D., Favreau, M. R. & Hallick, R. B. 2001 Comparison of psbK operon organization and group III intron content in chloroplast genomes of 12 Euglenoid species. *Mol. Gen. Genet.* **264**, 682–690.
Doolittle, W. F. 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. J., Penny, S., Deng, L.-T., Wu, X., Reith, M., Cavalier-Smith, T. & Maier, U.-G. 2001 The highly reduced genome of an enslaved algal nucleus. *Nature* **410**, 1081–1086.
Durnford, D. G., Deane, J. A., Tan, S., McFadden, G. I., Gantt, E. & Green, B. R. 1999 A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J. Mol. Evol.* **48**, 59–68.
Eschbach, S., Hofmann, C. J., Maier, U.-G., Sitte, P. & Hansmann, P. 1991 A eukaryotic genome of 660 kb: electrophoretic karyotype of nucleomorph and cell nucleus of the cryptomonad alga, *Pyrenomonas salina*. *Nucleic Acids Res.* **19**, 1779–1781.
Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. 2001 Nuclear encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellates plastids. *Mol. Biol. Evol.* **18**, 418–426.
Frantz, C., Ebel, C., Paulus, F. & Imbault, P. 2000 Characterization of trans-splicing in Euglenoids. *Curr. Genet.* **37**, 349–355.
Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V. & Sogin, M. L. 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, S. P. 1978 The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Can. J. Bot.* **56**, 2883–2889.
Gibbs, S. P. 1979 The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. *J. Cell Sci.* **35**, 253–266.
Gibbs, S. P. 1981*a* The chloroplast endoplasmic reticulum. *Int. Rev. Cytol.* **72**, 49–99.
Gibbs, S. P. 1981*b* The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Ann. NY Acad. Sci.* **361**, 193–208.
Gilson, P. & McFadden, G. I. 1995 The chlorarachniophyte: a cell with two different nuclei and two different telomeres. *Chromosoma* **103**, 635–641.
Gilson, P. R. & McFadden, G. I. 1996*a* Molecular and morphological characterisation of six chlorarachniophyte strains. *Phycol. Res.* **47**, 7–19.
Gilson, P. R. & McFadden, G. I. 1996*b* The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. *Proc. Natl Acad. Sci. USA* **93**, 7737–7742.
Gilson, P. R. & McFadden, G. I. 2002 Jam packed genomes—a preliminary, comparative analysis of nucleomorphs. *Genetica* **115**, 13–28.
Gockel, G. & Hachtel, W. 2000 Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate *Astasia longa*. *Protist* **151**, 347–351.
Gould, S. J. 1989 *Wonderful life*. New York: Norton.
Gregory, T. 2001 Coincidence coevolution or causation? DNA content, cell size and the C-value enigma *Biol. Rev.* **76**, 65–101.
Greenwood, A. D. 1974 The Cryptophyta in relation to phylogeny and photosynthesis. In *8th Int. Congr. of Electron Microscopy* (ed. J. Sanders & D. Goodchild), pp. 566–567. Canberra: Australian Academy of Sciences.
Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A. & Stutz, E. 1993 Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* **21**, 3537–3544.
Hannaert, V., Brinkmann, H., Nowitzki, U., Lee, J. A., Albert, M. A., Sensen, C. W., Gaasterland, T., Muller, M., Michels, P. & Martin, W. 2000 Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol. Biol. Evol.* **17**, 989–1000.
Hayashi, M., Taniguchi, S., Ishizuka, Y., Kim, H. S., Wataya, J., Yamamoto, A. & Moriyama, Y. 2001 A homologue of N-ethylmaleimide-sensitive factor in the malaria parasite *Plasmodium falciparum* is exported and localized in vesicular structure in the cytoplasm in the brefeldin A-sensitive pathway. *J. Biol. Chem.* **276**, 15 249–15 255.
He, C. Y., Shaw, M. K., Pletcher, C. H., Striepen, B., Tilney, L. G. & Roos, D. S. 2001 A plastid segregation defect in the protozoan parasite *Toxoplasma gondii*. *EMBO J.* **20**, 330–339.
Hibberd, D. J. & Norris, R. E. 1984 Cytology and ultrastructure of *Chlorarachnion reptans* (Chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). *J. Phycol.* **20**, 310–330.
Hiller, R. G. 2001 'Empty' minicircles and *petB*/*atpA* and *psbD*/*psbE* (cytb559 alpha) genes in tandem in *Amphidinium carterae* plastid DNA. *FEBS Lett.* **505**, 449–452.
Hiller, R. G., Crossley, L. G., Wrench, P. M., Santucci, N. & Hofmann, E. 2001 The 15-kDa forms of the apo-peridinin-chlorophyll a protein (PCP) in dinoflagellates show high identity with the apo-32 kDa PCP forms, and have similar N-terminal leaders and gene arrangements. *Mol. Genet. Genom.* **266**, 254–259.
Hiltbrunner, A., Bauer, J., Vidi, P. A., Infanger, S., Weibel, P., Hohwy, M. & Kessler, F. 2001 Targeting of an abundant cytosolic form of the protein import receptor at Toc159 to the outer chloroplast membrane. *J. Cell Biol.* **154**, 309–316.
Hopkins, J., Fowler, R., Krishna, S., Wilson, I., Mitchell, G. & Bannister, L. 1999 The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis. *Protist* **150**, 283–295.

Inagaki, J., Fujita, Y., Hase, T. & Yamamoto, Y. 2000 Protein translocation within chloroplast is similar in *Euglena* and higher plants. *Biochem. Biophys. Res. Commun.* **277**, 436–442.

Ishida, K., Green, B. R. & Cavalier-Smith, T. 1999 Diversification of a chimaeric algal group, the chlorarachniophytes: phylogeny of nuclear and nucleomorph small subunit rRNA genes. *Mol. Biol. Evol.* **16**, 321–331.

Ishida, K., Cao, Y., Hasegawa, M., Okada, N. & Hara, Y. 1997 The origin of chlorarachniophyte plastids, as inferred from phylogenetic comparisons of amino acid sequences of Ef-Tu. *J. Mol. Evol.* **45**, 682–687.

Ishida, K., Cavalier-Smith, T. & Green, B. R. 2000 Endomembrane structure and the chloroplast protein targeting pathway in *Heterosigma akashiwo* (Raphidophyceae, Chromista). *J. Phycol.* **36**, 1135–1144.

Jarvis, P. & Soll, J. 2001 Toc, Tic, and chloroplast protein import. *Biochim. Biophys. Acta* **1541**, 64–79.

Joiner, K. A. & Roos, D. S. 2002 Secretory traffic in the eukaryotic parasite *Toxoplasma gondii*: less is more. *J. Cell Biol.* **157**, 557–563.

Katinka, M. D. (and 16 others) 2001 Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453.

Keeling, P. J. & Palmer, J. D. 2001 Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. *Proc. Natl Acad. Sci. USA* **98**, 10 745–10 750.

Keeling, P. J., Deane, J. A. & McFadden, G. I. 1998 The phylogenetic position of alpha- and beta-tubulins from the *Chlorarachnion* host and *Cercomonas* (Cercozoa). *J. Euk. Microbiol.* **45**, 561–570.

Kessler, F. & Blobel, G. 1996 Interaction of the protein import and folding machineries of the chloroplast. *Proc. Natl Acad. Sci. USA* **93**, 7684–7689.

Köhler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J., Palmer, J. D. & Roos, D. S. 1997 A plastid of probable green algal origin in Apicomplexan parasites. *Science* **275**, 1485–1489.

Krepinsky, K., Plaumann, M., Martin, W. & Schnarrenberger, C. 2001 Purification and cloning of chloroplast 6-phosphogluconate dehydrogenase from spinach. Cyanobacterial genes for chloroplast and cytosolic isoenzymes encoded in eukaryotic chromosomes. *Eur. J. Biochem.* **268**, 2678–2686.

Lee, Y. J., Kim, D. H., Kim, Y. W. & Hwang, I. 2001 Identification of a signal that distinguishes between the chloroplast outer envelope membrane and the endomembrane system *in vivo*. *Plant Cell.* **13**, 2175–2190.

Lemieux, C., Otis, C. & Turmel, M. 2000 Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* **403**, 649–652.

Linton, E., Hittner, D., Levandowski, C. F., Auld, T. & Triemer, R. 1999 A molecular study of euglenoid phylogeny using small subunit rDNA. *J. Euk. Microbiol.* **46**, 217–223.

McFadden, G. I. 1999 Plastids and protein targeting. *J. Euk. Microbiol.* **46**, 339–346.

McFadden, G. I., Reith, M. E., Munholland, J. & Lang-Unnasch, N. 1996 Plastid in human parasites. *Nature* **381**, 482.

McFadden, G. I., Gilson, P. R., Douglas, S. E., Cavalier-Smith, T., Hofmann, C. J. B. & Maier, U.-G. 1997 Bonsai genomics: sequencing the smallest eukaryotic genomes. *Trends Genet.* **13**, 46–49.

McKerracher, L. & Gibbs, S. P. 1982 Cell and nucleomorph division in the alga *Cryptomonas*. *Can. J. Bot.* **60**, 2440–2452.

Maier, U.-G., Hofmann, C. J., Eschbach, S., Wolters, J. & Igloi, G. L. 1991 Demonstration of nucleomorph-encoded eukaryotic small subunit ribosomal RNA in cryptomonads. *Mol. Gen. Genet.* **230**, 155–160.

Maier, U.-G., Douglas, S. & Cavalier-Smith, T. 2000 The nucleomorph genomes of cryptophytes and chlorarachniophytes. *Protist* **151**, 103–109.

Martin, W. & Herrmann, R. G. 1998 Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* **118**, 9–17.

Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M. & Kowallik, K. V. 1998 Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165.

Matsuzaki, M., Kikuchi, T., Kita, K., Kojima, S. & Kuroiwa, T. 2001 Large amounts of apicoplast nucleoid DNA and its segregation in *Toxoplasma gondii*. *Protoplasma* **218**, 180–191.

Maynard Smith, J. & Szathmáry, E. 1995 *The major transitions in evolution*. Oxford University Press.

Mereschkowsky, C. 1905 Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Zentralbl.* **25**, 593–604.

Mereschkowsky, C. 1910 Theorie der Zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. *Biol. Zentralbl.* **30**, 278–303, 321–347, 353–367.

Meyer, S. 1987 The taxonomic implications of the ultrastructure and cell division of a stigma-containing *Chroomonas* sp. (Cryptophyceae) from Rocky Bay, Natal, South Africa. *S. Afr. J. Bot.* **53**, 129–139.

Moestrup, Ø. & Sengco, M. 2001 Ultrastuctural studies on *Bigelowiella natans*, gen. et sp. nov., a chlorarachniophyte flagellate. *J. Phycol.* **37**, 624–646.

Moreira, D., Le Guyader, H. & Philippe, H. 2000 The origin of red algae and the evolution of chloroplasts. *Nature* **405**, 69–72.

Mori, H. & Cline, K. 2001 Post-translational protein translocation into thylakoids by the Sec and DeltapH-dependent pathways. *Biochim. Biophys. Acta* **1541**, 80–90.

Morrall, S. & Greenwood, A. D. 1982 Ultrastructure of nucleomorph division in species of Cryptophyceae and its evolutionary implications. *J. Cell Sci.* **54**, 311–328.

Mullner, A. N., Angeler, D. G., Samuel, R., Linton, E. W. & Triemer, R. E. 2001 Phylogenetic analysis of phagotrophic, photomorphic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence. *Int. J. Syst. Evol. Microbiol.* **51**, 783–791.

Murray, A. & Szostak, J. 1985 Chromosome segregation in mitosis and meiosis. *A. Rev. Cell Biol.* **1**, 289–315.

Nurse, P. 1985 The genetic control of cell volume. In *The evolution of genome size* (ed. T. Cavalier-Smith), pp. 185–196. Chichester: Wiley.

Obornik, M., Van de Peer, Y., Hypsa, V., Frickey, T., Slapeta, J. R., Meyer, A. & Lukes, J. 2002 Phylogenetic analyses suggest lateral gene transfer from the mitochondrion to the apicoplast. *Gene* **285**, 109–118.

Oliveira, M. C. & Bhattacharya, D. 2000 Phylogeny of the Bangiophyceae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. *Am. J. Bot.* **87**, 482–492.

Petrov, D. A. 2001 Evolution of genome size: new approaches to an old problem. *Trends Genet.* **17**, 23–28.

Petrov, D. A. 2002 Mutational equilibrium model of genome size evolution. *Theor. Popul. Biol.* **61**, 531–544.

Pont-Kingdon, G. A., Okada, N. A., Macfarlane, J. L., Beagley, C. T., Wolstenholme, D. R., Cavalier-Smith, T. & Clark-Walker, G. D. 1998 Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. *J. Mol. Evol.* **46**, 419–431.

Preisfeld, A., Busse, I., Klingberg, M., Talke, S. & Ruppel, H. G. 2001 Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *Int. J. Syst. Evol. Microbiol.* **51**, 751–758.

Rivier, C., Goldschmidt-Clermont, M. & Rochaix, J. D. 2001 Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. *EMBO J.* **20**, 1765–1773.

Saldarriaga, J., Taylor, F. J. R., Keeling, P. J. & Cavalier-Smith, T. 2001 Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *J. Mol. Evol.* **53**, 204–213.

Schleiff, E. & Klosgen, R. B. 2001 Without a little help from 'my' friends: direct insertion of proteins into chloroplast membranes? *Biochim. Biophys. Acta* **1541**, 22–33.

Schleiff, E., Tien, R., Salomon, M. & Soll, J. 2001 Lipid composition of outer leaflet of chloroplast outer envelope determines topology of OEP7. *Mol. Biol. Cell.* **12**, 4090–4102.

Schnepf, E. 1993 From prey via endosymbiont to plastid: comparative studies in dinoflagellates. In *Origins of plastids: symbiogenesis, prochlorophytes and the origins of chloroplasts* (ed. R. A. Lewin), pp. 53–76. London: Chapman & Hall.

Sekiguchi, H., Moriya, M., Nakayama, T. & Inouye, I. 2002 Vestigial chloroplasts in heterotrophic stramenopiles *Pteridomonas danica* and *Ciliophrys infusionum* (Dictyochophyceae). *Protist* **153**, 157–167.

Selosse, M., Albert, B. & Godelle, B. 2001 Reducing the genome size of organelles favours transfer to the nucleus. *Trends Ecol. Evol.* **16**, 135–141.

Sheveleva, E. V., Giordani, N. V. & Hallick, R. B. 2002 Identification and comparative analysis of the chloroplast alpha-subunit gene of DNA-dependent RNA polymerase from seven *Euglena* species. *Nucleic Acids Res.* **30**, 1247–1254.

Simpson, A. G. B. 1997 The identity and composition of the Euglenozoa. *Arch. Protistenk.* **148**, 318–328.

Smith, R. & Patterson, D. J. 1986 Analysis of heliozoan interrelationships: an example of the potentials and limitations of ultrastructural approaches to the study of protistan phylogeny. *Proc. R. Soc. Lond.* B **227**, 325–366.

Sohrt, K. & Soll, J. 2000 Toc64, a new component of the protein translocon of chloroplasts. *J. Cell Biol.* **148**, 213–221.

Stechmann, A. & Cavalier-Smith, T. 2002 Rooting the eukaryote tree by using a derived gene fusion. *Science* **297**, 89–91.

Steiner, J. M. & Löffelhardt, W. 2002 Protein import into cyanelles. *Trends Plant Sci.* **7**, 72–77.

Stibitz, T. B., Keeling, P. J. & Bhattacharya, D. 2000 Symbiotic origin of a novel actin gene in the cryptophyte *Pyrenomonas helgolandii*. *Mol. Biol. Evol.* **17**, 1731–1738.

Striepen, B., Crawford, M. J., Shaw, M. K., Tilney, L. G., Seeber, F. & Roos, D. S. 2000 The plastid of *Toxoplasma gondii* is divided by association with the centrosomes. *J. Cell Biol.* **151**, 1423–1434.

Strasburger, E. 1893 Über die Wirkungssphäre der Kerne und die Zellgrösse. *Histol. Beitr.* **5**, 97–124.

Sulli, C., Fang, Z.-W., Muchhal, U. & Schwartzbach, S. D. 1999 Topology of *Euglena* chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *J. Biol. Chem.* **274**, 457–463.

Taylor, F. J. R. 1974 Implications and extensions of the serial endosymbiosis theory of the origin of eukaryotes. *Taxon* **23**, 229–258.

Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C. F. & Jakobsen, K. S. 2000 Phylogenetic analyses indicate that the 19′ hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol. Biol. Evol.* **17**, 718–729.

Thompson, M. D., Copertino, D. W., Thompson, E., Favreau, M. R. & Hallick, R. B. 1995 Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. *Nucleic Acids Res.* **23**, 4745–4752.

Thompson, M. D., Zhang, L., Hong, L. & Hallick, R. B. 1997*a* Extensive structural conservation exists among several homologs of two *Euglena* chloroplast group II introns. *Mol. Gen. Genet.* **257**, 45–54.

Thompson, M. D., Zhang, L., Hong, L. & Hallick, R. B. 1997*b* Two new group-II twintrons in the *Euglena gracilis* chloroplast are absent in basally branching *Euglena* species. *Curr. Genet.* **31**, 89–95.

Tomitani, A., Okada, K., Miyashita, H., Matthijs, H. C., Ohno, T. & Tanaka, A. 1999 Chlorophyll *b* and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature* **400**, 159–162.

Tovar, J., Fischer, A. & Clark, C. G. 1999 The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol. Microbiol.* **32**, 1013–1021.

Trombetta, V. V. 1942 The cytonuclear ratio. *Bot. Rev.* **8**, 317–336.

Tsai, L. Y., Tu, S. L. & Li, H. M. 1999 Insertion of atToc34 into the chloroplastic outer membrane is assisted by at least two proteinaceous components in the import system. *J. Biol. Chem.* **274**, 18 735–18 740.

Tu, S. L. & Li, H. M. 2001 Insertion of OEP14 into the outer envelope membrane is mediated by proteinaceous components of chloroplasts. *Plant Cell.* **12**, 1951–1960.

Turmel, M., Otis, C. & Lemieux, C. 1999 The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. *Proc. Natl. Acad. Sci. USA* **96**, 10 248–10 253.

Turmel, M., Otis, C. & Lemieux, C. 2002 The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Mol. Biol. Evol.* **19**, 24–38.

van de Peer, Y., Rensing, S. A., Maier, U. G. & De Wachter, R. 1996 Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc. Natl Acad. Sci. USA* **93**, 7732–7736.

van Dooren, G. G., Schwartzbach, S. D., Osafune, T. & McFadden, G. I. 2001 Translocation of proteins across the multiple membranes of complex plastids. *Biochim. Biophys. Acta* **1541**, 34–53.

Vickerman, K., Le Ray, D., Hoef-Emden, K. & De Jonckheere, J. 2002 The soil flagellate *Proleptomonas faecicola*: cell organisation and phylogeny suggest that the only described free-living trypanosomatid is not a kinetoplastid but has cercomonad affinities. *Protist* **153**, 9–24.

Wastl, J. & Maier, U. G. 2000 Transport of proteins into cryptomonads complex plastids. *J. Biol. Chem.* **275**, 23 194–23 198.

Watanabe, M. M., Suda, S., Inouye, I., Sawaguchi, T. & Chihara, M. 1990 *Lepidodinium viride* gen. sp. nov. (Gymnodiniales, Dinophyta) a green dinoflagellate with a chlorophyll *a*- and *b*-containing endosymbiont. *J. Phycol.* **26**, 741–751.

Wilson, R. J. (and 10 others) 1996 Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **261**, 155–172.

Wintz, H., Grienenberger, J. M., Weil, J. H. & Lonsdale, D. M. 1988 Location and nucleotide sequence of two tRNA genes and a tRNA pseudo-gene in the maize mitochondrial genome: evidence for the transcription of a chloroplast gene in mitochondria. *Curr. Genet.* **13**, 247–254.

Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. 2002 The single, ancient origin of chromist plastids. *Proc. Natl Acad. Sci. USA* **99**, 15 507–15 512.

Yu, T. S. & Li, H. 2001 Chloroplast protein translocon components atToc159 and atToc33 are not essential for chloroplast biogenesis in guard cells and root cells. *Plant Physiol.* **127**, 90–96.

Zauner, S., Fraunholz, M., Wastl, J., Penny, S., Beaton, M. J., Cavalier-Smith, T., Maier, U.-G. & Douglas, S. 2000 Chloroplast protein-coding and centrosomal genes, odd telomeres, and a tRNA intron in an unusually compact eukaryotic genome—the cryptomonad nucleomorph. *Proc. Natl Acad. Sci. USA* **97**, 200–205.

Zhang, X. P. & Glaser, E. 2002 Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. *Trends Plant Sci.* **7**, 14–21.

Zhang, Z. 1999 Single gene circles in dinoflagellate chloroplast genomes: characterization and phylogeny. PhD thesis, University of British Columbia.

Zhang, Z., Green, B. R. & Cavalier-Smith, T. 1999 Single gene circles in dinoflagellate chloroplast genomes. *Nature* **400**, 155–159.

Zhang, Z., Green, B. R. & 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.

Zhang, Z., Cavalier-Smith, T. & Green, B. R. 2001 A family of selfish minicircular chromosomes with jumbled chloroplast gene fragments from a dinoflagellate. *Mol. Biol. Evol.* **18**, 1558–1565.

Zhang, Z., Cavalier-Smith, T. & Green, B. R. 2002 Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. *Mol. Biol. Evol.* **19**, 489–500.

Zuegge, J., Ralph, S., Schmuker, M., McFadden, G. I. & Schneider, G. 2001 Deciphering apicoplast targeting signals—feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* **280**, 19–26.

Discussion

M. van der Giezen (*School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey, UK*). I gather that red algae have mitochondria, but you did not show anything in the secondary endosymbiosis to indicate that mitochondria were left by the endosymbiont.

T. Cavalier-Smith. The mitochondria of the secondary endosymbiont must have gone half a billion years ago. In both these cases (chlorarachniophytes and cryptomonads) there is no trace of mitochondria and no trace of a Golgi body in the periplastid space, so it is a very ancient symbiosis that probably happened 500 million years ago. The cytoplasm of the endosymbiont would have been stripped down to almost nothing then and has hardly changed since. So they just were not retained.

W. Martin (*Biological-Medical Research Centre, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany*). Certainly the principle of membrane heredity is a very important one, but looking at eukaryotic cells we have unique endomembrane systems, the endoplasmic reticulum and the nuclear membrane, so my question to you is in what order would you derive the endoplasmic reticulum and the nuclear membrane, and from what?

T. Cavalier-Smith. I would derive the endo-membrane system first, that is the endoplasmic reticulum, from the surface of the bacterial ancestor of the eukaryote, by invagination and budding off, to make a separate membrane system; but to retain it as a separate system you have to evolve for the first time in the history of life coated vesicles that bud off from these internalized membranes, to return to the cell surface the lipids and proteins so that the plasma membrane can grow. One of the most fundamental things in making a eukaryote is the evolution of coated vesicle budding and targeting; the nuclear envelope is really just a specialization of part of the rough endoplasmic reticulum. To make a nucleus you need to have two things in addition to the endoplasmic reticulum: one is the nuclear pore complex, which is the key thing that many people leave out of consideration; secondly, you need to evolve mitosis. I think both had to evolve more or less simultaneously, but my key point is that the budding from the cell surface arose by the evolution of phagotrophy, which is opposed to the view you tend to adopt.

A. Wilkins (*BioEssays, Wallington, Surrey, UK*). You have mentioned one possible selective force for retention of non-coding DNA in the nucleus, but it is quite apparent that quite a lot of non-coding DNA is involved in regulation of genes. Should that not be incorporated into your scheme as an important function?

T. Cavalier-Smith. I think it is a minor part and that such regulatory needs do not explain why you get the correlation, over a 40 000-fold range, between the total amount of DNA in the nucleus and cell size. You can regulate things perfectly well with only about one-tenth of the DNA we have or one-hundredth of what a lily plant has. Obviously, within that non-coding DNA, which is explained primarily for the reasons I gave, there would be two other sorts of DNA there; one, as you said, would have some regulatory sequences; secondly, there will be some genuinely selfish sequences, like Alu sequences or different sorts of transposons, but although they are 'selfish' because they have a specific sequence which has made them multiply, they are doing so *within an overall global constraint* that is determined by these cell-volume-related selective forces. So it is not that selfish DNA or regulation are unimportant, but in my view the overall thing that accounts for the correlation between cell size and the amount of nuclear DNA is not regulation of genes by non-coding DNA.

W. F. Doolittle (*Dalhousie University, Halifax, Nova Scotia, Canada*). It seems to me, however, that the graph of nuclear volume and DNA content versus nucleomorph could still be explained by selfish DNA by arguing the nucleomorphs are under a different kind of selective regime than the cells as a whole, because nucleomorphs are competing with each other for space within that genome. They are, therefore, being selected just as prokaryotes are, because they have to out-compete each other, and there is not such strong selective pressure on the cell as a whole; so the nucleus can just relax and let more selfish DNA accumulate.

T. Cavalier-Smith. But they are not out-competing each other. It is not like mitochondrial or chloroplast genomes where you have in fact got multiple genomes within the cells, with inter-organellar competition. In most cryptomonads you have just one nucleomorph genome per cell and it is regulated in the same way as the host nucleus, which is by the cyclin-dependent protein kinases, and so they are not competing. The DNA replication in nucleomorphs is controlled in a manner suggesting that there must be the equivalent of an organellar S-phase, so I think that the cell biology behind your assumption is wrong.

W. F. Doolittle. OK. You win that one.

J. F. Allen (*Plant Biochemistry, Lund University, Lund, Sweden*). Non-functional DNA—is it possible that this is simply a storage organ for nucleic acid precursors? Has this been suggested? Does it make any sense?

T. Cavalier-Smith. I expect it has, but cells could not get it out because it is in between the genes and within the genes. If you suddenly need nucleic acids what do you do? Chop your chromosomes up and get them out? There are a few organisms that do chop their chromosomes up, but it does not seem to relate to the storage of the precursors of nucleic acids. Sorry, I do not quite follow the argument.

C. J. Leaver (*Department of Plant Sciences, University of Oxford, Oxford, UK*). I had always thought that the large variation in length of targeting sequences could be another means of getting amino acids into the chloroplast or mitochondria whether they wanted them or not, but still.

F. R. Whatley (*Department of Plant Sciences, University of Oxford, Oxford, UK*). I would like to ask a pre-genomics question. Where did the basic argument come from that says there should be only one origin of plastids, rather than one for the greens and one for the reds, taking into account that the Chromista all came from the reds?

T. Cavalier-Smith. The original argument, when I first argued this in 1982, was twofold; it was first of all the protein targeting argument, that to originate an organelle required the origin of a novel protein targeting machinery, which I submit is a complex thing, and also the addition of transit sequences to all the thousands of genes that we have been talking about. The second argument, which was one of parsimony and economy from a phylogenetic viewpoint, is that we should not invoke several separate symbioses unless there is a good phylogenetic reason for doing so, and at that time I looked at the nature of the host and deduced that you could actually have a host of particular characteristics that could have evolved into both the reds and the greens, and also the glaucophytes. This third group contains *Cyanophora*, which at the time no one thought was related to the other two.

Since then, of course, there has been so much molecular evidence that supports this view, including common nature of the protein targeting machinery in the three groups. People can do experiments where they take proteins from one of the three groups and target them to chloroplasts of another group. Secondly, there are multiple gene trees, for example David Moeira's *Nature* paper, where he showed all three plant groups are more closely related to each other than to other groups. A third line of argument has to do with the operon structure and gene organization of the chloroplast genomes, which have a few features that look like derived features that are unique to chloroplasts and are not found in cyanobacteria. Of course, someone might discover the cyanobacterium, or rather the three cyanobacteria, of all different colours once postulated as separate ancestors, but really, the strength of the argument for a single origin is such that it is the most reasonable view by far.

C. J. Howe (*Department of Biochemistry, University of Cambridge, Cambridge, UK*). I have a comment on the answer to the previous question. It is clear that elements at least of the protein import machinery were probably present in the ancestor, so whether one had a polyphyletic or monophyletic origin from some photosynthetic bacterium one might expect to find similar import machinery.

T. Cavalier-Smith. Some elements are derived from cyanobacterial proteins, and I had always supposed that would be the case, but for some elements it is not true. It is the novel elements and the combination of novel elements with pre-existing elements that makes it more parsimonious to suppose it happened only once, and since all the other evidence now supports it. You do not dispute that do you?

C. J. Howe. That was my second point. It is increasingly clear that there are significant problems with some of the sequence analyses; they do not take account of 'covariable' evolution and that kind of thing, and to my mind what is going on to be much more convincing is the operon structure data rather than a simplistic interpretation of the sequence-based trees.

T. Cavalier-Smith. No one should make simplistic interpretations of these but I think that taking all the different arguments together there is not any strong indication against this view.

GLOSSARY

EM: epiplastid membrane
ER: endoplasmic reticulum
GAPDH: glyceraldehyde phosphate dehydrogenase
NLS: nuclear localization signal
NM: nucleomorph
OM: outer membrane
PPM: periplastid membrane
PPR: periplastid reticulum
RER: rough endoplasmic reticulum
RUBISCO: ribulose-1,5-bisphosphate carboxylase–oxygenase
SNAP: soluble *N*-ethylmaleimide-sensitive-factor attachment protein
SNARE: SNAP receptors
SRP: signal recognition particle