Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 11368–11372, November 1994
Evolution

Chimeric conundra: Are nucleomorphs and chromists monophyletic or polyphyletic?

T. CAVALIER-SMITH, M. T. E. P. ALLSOPP*, AND E. E. CHAO

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

Communicated by Russell F. Doolittle, August 1, 1994

ABSTRACT **All algae with chloroplasts located not freely in the cytosol, but inside two extra membranes, probably arose chimerically by the permanent fusion of two different eukaryote cells: a protozoan host and a eukaryotic algal symbiont. Two such groups, cryptomonads (phylum Cryptista) and Chlorarachniophyta, still retain a DNA-containing relic of the nucleus of the algal endosymbiont, known as the nucleomorph, as well as the host nucleus. These two phyla were traditionally assumed to have obtained their chloroplasts separately by two independent symbioses. We have sequenced the nuclear and the nucleomorph 18S rRNA genes of the nonphotosynthetic cryptomonad *Chilomonas paramecium*. Our phylogenetic analysis suggests that cryptomonad and chlorarachniophyte nucleomorphs may be related to each other and raises the possibility that both phyla may have diverged from a common ancestral chimeric cell that originated by a single endosymbiosis involving an algal endosymbiont related to the ancestor of red algae. But, because of the instability of the molecular trees when different taxa are added, there is insufficient evidence to overturn the traditional view that *Chlorarachnion* nucleomorphs evolved separately from a relative of green algae. The four phyla that contain chromophyte algae (those with chlorophyll c—i.e., Cryptista, Heterokonta, Haptophyta, Dinozoa) are distantly related to each other and to Chlorarachniophyta on our trees. However, all of the photosynthetic taxa within each of these four phyla radiate from each other very substantially after the radiation of the four phyla themselves. This favors the view that the common ancestor of these four phyla was not photosynthetic and that chloroplasts were implanted separately into each much more recently. This probable polyphyly of the chromophyte algae, if confirmed, would make it desirable to treat Cryptista, Heterokonta, and Haptophyta as separate kingdoms, rather than to group them together in the single kingdom Chromista.**

The predominantly photosynthetic kingdom Chromista (1) differs fundamentally from plants (kingdom Plantae) in having chloroplasts not in the cytosol but within a subcellular compartment, the periplastid space, bounded by two distinct membranes (2–5). The innermost one, the periplastid membrane (5), is interpreted as a relic of the plasma membrane of a former eukaryotic algal symbiont, or symbionts, that merged with a protozoan host or hosts to create the complex chromist cell (4, 5). The periplastid space represents the former cytosol of that endosymbiont, and in two chromist phyla Cryptista (5) and Chlorarachniophyta (6) still contains 80S ribosomes and a relic of the algal nucleus (the nucleomorph) with three tiny linear chromosomes bearing multiple rRNA genes (7–9). Cryptista are divided into two classes (3, 5): one comprising cryptomonads, which have plastids, nucleomorphs, and two different kinds of 80S ribosomes; the other containing only *Goniomonas*, which lacks plastids and nucleomorphs and (like most eukaryotes) has only one kind of 80S ribosome. Nuclear and nucleomorph 18S rRNA genes were previously sequenced for two photosynthetic cryptomonads, *Cryptomonas* ϕ (10) and *Pyrenomonas salina* (11). We have now cloned and sequenced both nuclear and nucleomorph 18S RNA genes from the nonphotosynthetic cryptomonad *Chilomonas*. The other chromist phyla Heterokonta (2) and Haptophyta (2), together called Chromobiota (3, 12) lack nucleomorphs and periplastid ribosomes and so have only one type of 18S rRNA gene.

Chlorarachniophyta (*Chlorarachnion* and relatives) have often been thought to be unrelated to the other three chromist phyla recently classified in the subkingdom Euchromista (3) because they differ in two fundamental respects: (*i*) their chloroplasts contain chlorophyll b, not c as in Euchromista, and (*ii*) the membrane outside the periplastid membrane lacks 80S ribosomes on its cytosolic surface (6, 13). The presence in Euchromista of ribosomes on the cytosolic face of the outermost membrane probably originated when the food vacuole membrane enclosing the algal endosymbiont fused with the outer membrane of the host's nuclear envelope (2, 4); clearly such fusion never occurred in *Chlorarachnion*, but this does not necessarily mean that *Chlorarachnion* and euchromists obtained their chloroplasts in separate symbioses (13)—they might, instead, have diverged from a single chimeric ancestor (2, 14) prior to the fusion of the food vacuole and outer nuclear envelope membranes. The membrane topology around chromist chloroplasts means that nuclearly coded chloroplast proteins are imported across four membranes, by mechanisms yet unknown. The necessarily much greater complexity of such topogenic processes in photosynthetic chromists (2, 5) than in plants was a major reason for postulating that they evolved once only (2) and for classifying all three euchromist phyla (1, 2), and more recently all four phyla (3), in a single kingdom, the Chromista—i.e., a third botanical kingdom in addition to Plantae and Fungi. Here we test the validity of this kingdom by molecular phylogenetic analysis of nuclear 18S rRNA sequences from all four chromist phyla and of nucleomorph 18S rRNA sequences from both Cryptista and Chlorarachniophyta.

MATERIALS AND METHODS

Chilomonas paramecium DNA was purified from strain CCAP 977/2a. The two 18S rRNA genes were amplified by the polymerase chain reaction using standard primers (15). Two separate amplification products were seen on agarose gels, an upper band and a lower band; each was excised separately from the gel, purified, cloned into mp18 and mp19 M13 phages, and sequenced on both strands using conserved internal primers (16). The sequence from the upper band (GenBank accession no. L28811) was 2047 nt long and groups on the tree with the longer sequence from *Cryptomonas* ϕ,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Onderstepoort Veterinary Institute, Private Bag X5, Onderstepoort 0110, South Africa.

which has been proven to be from the nucleomorph (17), while that from the lower band (GenBank accession no. L28812) is only 1758 nt long and groups with the *Cryptomonas* ϕ nuclear sequence (17). The sequences were aligned manually with about 220 other eukaryotic 18S rRNA sequences using the Genetic Data Environment software; representatives of all major groups [sequences from GenBank except for our own unpublished haptophyte (*Prymnesium* and *Pavlova*) and McFadden's *Goniomonas* (18) and *Chlorarachnion* (19) ones] that branch on the eukaryotic 18S rRNA tree (3) above *Dictyostelium* were selected for phylogenetic analysis.

RESULTS AND DISCUSSION

Parsimony (Fig. 1) and distance (Fig. 2) trees for 81 sequences both show that the three cryptomonad nucleomorph sequences form a robust branch with the three *Chlorarachnion* nucleomorph sequences. This branch is most closely related to red algae (shown strongly by parsimony) or to green algae (shown weakly by neighbor joining). The monophyly of nucleomorphs and their relationship to red algae is also supported by maximum likelihood analysis (Fig. 3). Despite the reasonably high bootstrap values for the nucleomorph branch, these trees do not establish nucleomorph monophyly unambiguously: we find only one signature sequence distinctive for all nucleomorphs [all have a T at a position (900 in the *Chilomonas* sequence) where virtually all of 220 other eukaryotes (sole exceptions, *Babesia caballi* and *Hexamita* with a T; *Giardia* and two microsporidia with a G) and most bacteria have an A]. When substantially more distant eukaryotes than *Dictyostelium* are added to the tree (e.g., Archezoa, Percolozoa, Euglenozoa) the nucleomorphs move below all of the other taxa shown here except for *Dictyostelium* (3) and sometimes form two separate branches. However, this lower position is probably a "long branches attract" artifact (23) caused by adding too-distantly related outgroups: indeed, *Chlorarachnion* (with longer branches) goes lower than cryptomonads.

According to the monophyletic theory of the origin of Chromista (2, 5, 14, 24), the constituent phyla diverged in chloroplast and other characters during or immediately following the conversion of the algal endosymbiont into a permanent organelle. However, these four phyla do not all form a single branch in the distance trees (Fig. 2), where only three of them (Heterokonta and the *Chlorarachnion* nuclei in

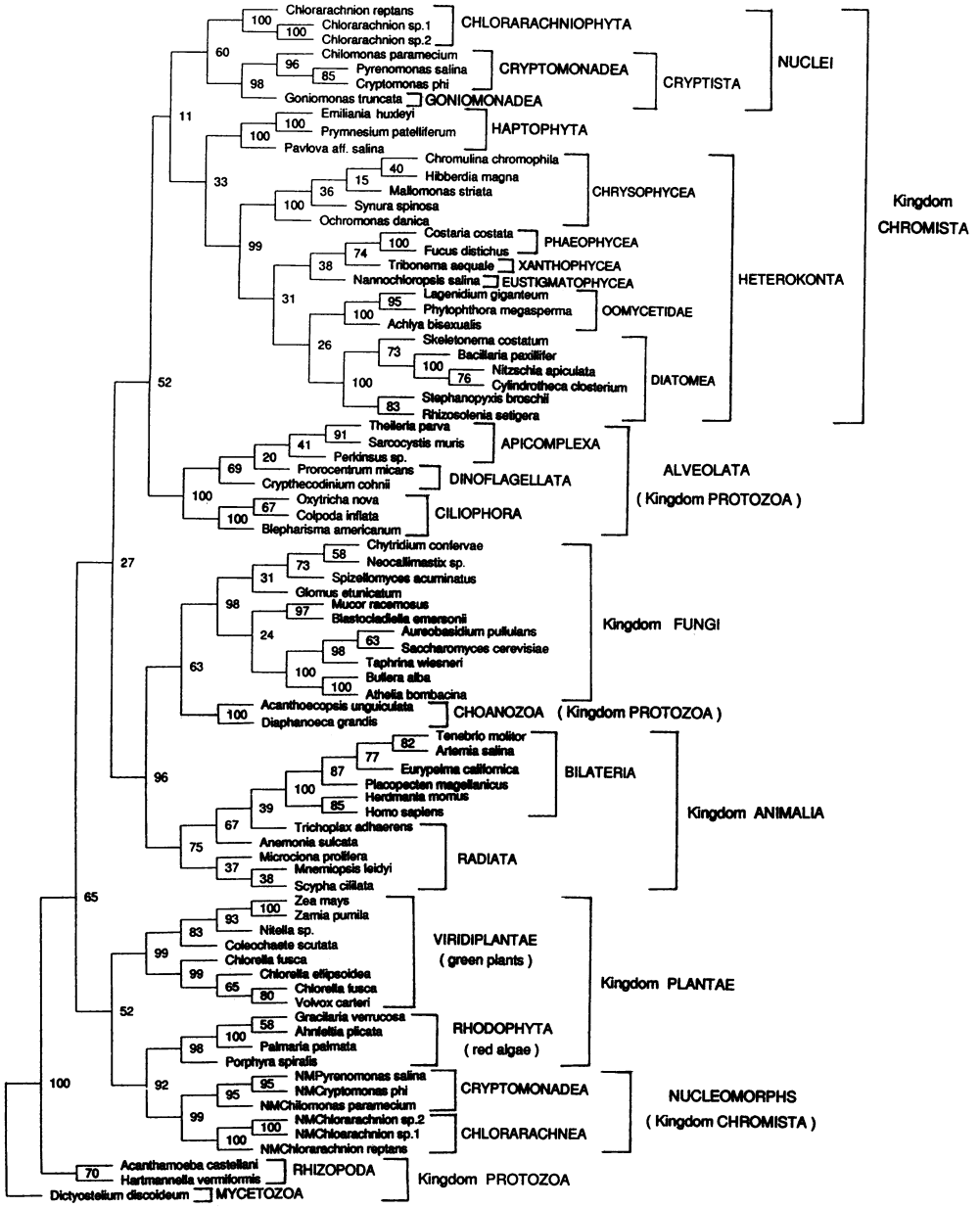

FIG. 1. Parsimony tree showing the relationship of *Chilomonas* nuclear and nucleomorph 18S rRNA sequences to 79 other eukaryotic sequences. The alignment (3500 positions: available on request from tom@tcs.botany.ubc.ca) was masked so as to exclude insertions idiosyncratic to just one major taxon (except for one nucleotide at each end), so only 2172 positions were included in the analysis. The tree shown was one of two equally parsimonious trees (requiring 13,747 steps) produced by the DNAPARS [with jumble (3 times) and global rearrangement options] program of Felsenstein's PHYLIP version 3.5; bootstrap values for 100 replicates are shown at the nodes. *Dictyostelium* was used to root the tree.

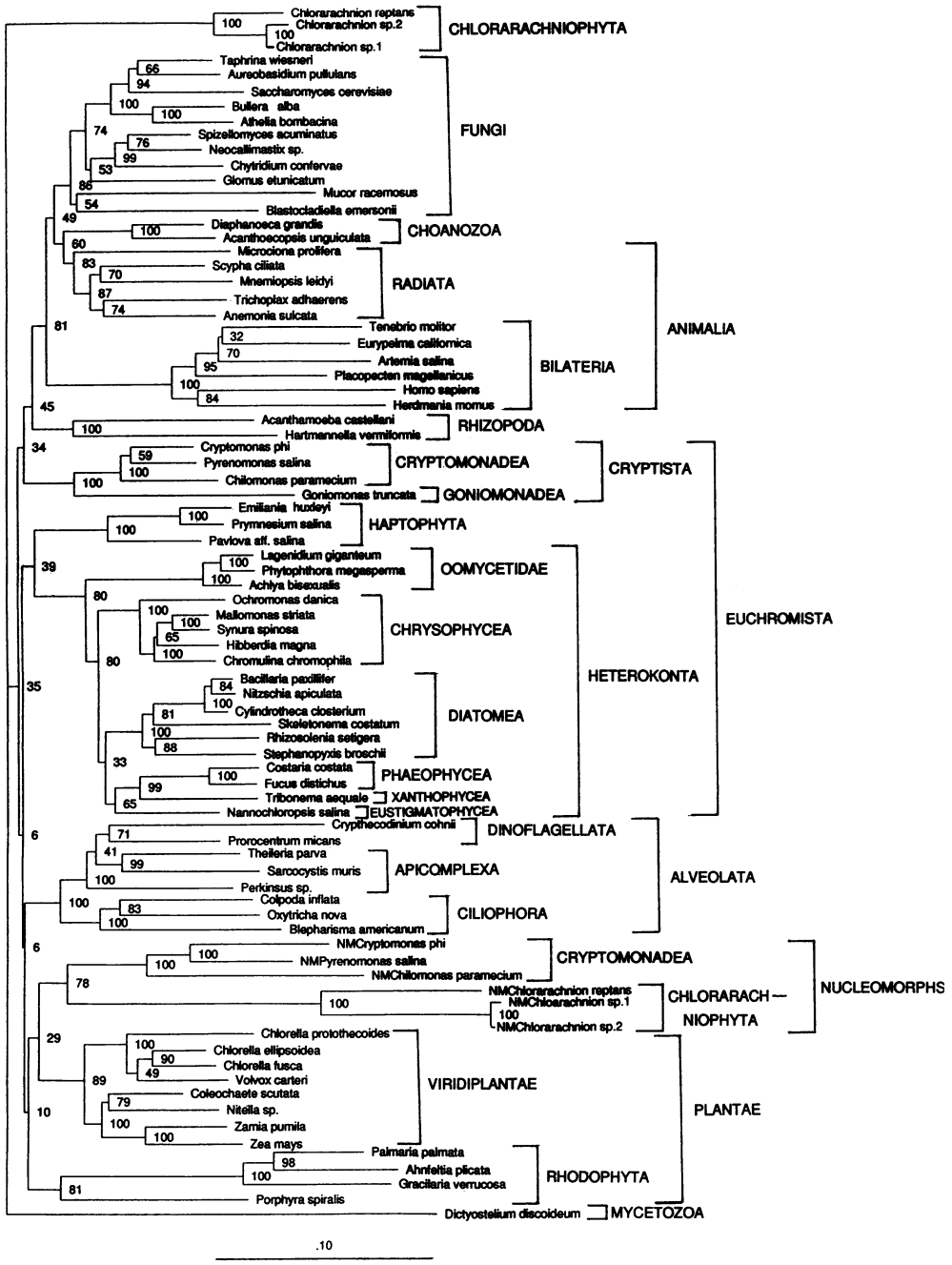

FIG. 2. Neighbor joining (NJ) distance tree for *Chilomonas* and other higher eukaryotes. Bootstrap values for 100 replicates are shown. The same 2172 positions as for Fig. 1 were used to calculate a distance matrix (Jukes–Cantor correction) by the DNADIST program of PHYLIP 3.5; this was used (with the jumble option) to calculate this NJ tree (using the NEIGHBOR program of PHYLIP 3.5) and also a second distance tree using the FITCH program of PHYLIP 3.5. The Fitch tree also showed all six nucleomorph sequences as a single clade [but as sister group of red algae (as in Fig. 1), not green algae (as with NJ)]; but the chlorarachniophyte nuclear clade and Haptophyta interchanged their positions. Moreover, in the Fitch tree Alveolata was the sister group of the Heterokonta/Chlorarachniophyta clade (not the Plantae/nucleomorph clade as with NJ) and Cryptista sister of green plants (not the animal/fungal/choanoflagellate/rhizopod clade as with NJ). Fitch was superior to NJ in showing the monophyly of the kingdom Animalia but did not show Plantae as monophyletic. These major differences in topology mean that distance methods cannot robustly prove or disprove the monophyly of the kingdoms Animalia, Plantae, and Chromista, because their major subtaxa diverged from each other much too rapidly, in a major burst of evolution, the "big bang" of Knoll (20), which fossils suggest occurred 650–700 million years ago (21). But within major clades NJ and Fitch trees were identical, except for two changes in branches with low bootstrap values: (*i*) within Heterokonta, the eustigmatophyte *Nannochloropsis* grouped with Chrysophycea (Fitch), not Phaeophycea/Xanthophycea (NJ); (*ii*) within Alveolata, *Perkinsus* grouped with dinoflagellates (Fitch), not below dinoflagellates plus other Apicomplexa (NJ).

the Fitch tree; Heterokonta and Haptophyta in the NJ tree) group together: the location of Cryptista is indeterminate in distance trees, varying with the algorithm (neighbor joining or Fitch) and the species composition of the alignment. Parsimony analysis gave two shortest trees: one (Fig. 1) grouping all chromists in one clade, the other identical except for the grouping of Cryptista with Alveolata rather than the other chromists. Maximum likelihood (Fig. 3) clearly suggests the paraphyly of the Chromista. It shows Alveolata as sister group to Heterokonta. It has been proposed (4, 25, 26) that dinoflagellates, the only photosynthetic alveolates, obtained their chloroplasts secondarily from a symbiotic chromobiote. Fig. 3 raises the possibility, instead, that the whole dinoflagellate cell including the chloroplasts, which have chlo-

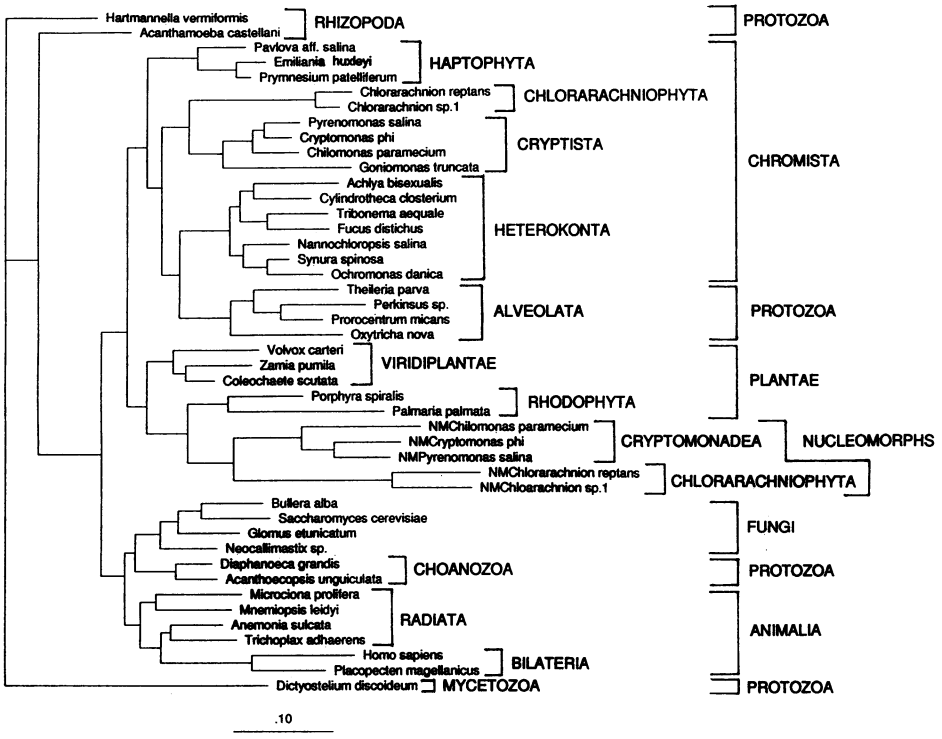

FIG. 3. Maximum likelihood tree for *Chilomonas* nuclear and nucleomorph 18S rRNA and 43 other eukaryotic sequences. Because maximum likelihood calculations would have been impossibly long for 81 sequences, a subselection of 45 was made; 2100 alignment positions were included in the analysis by Olsen's FASTDNAML version 1.0.6 on a Sparc 2 workstation using global rearrangement, allowing the crossing of three branches (4529 trees examined). Three jumbles gave the same tree, with all branch lengths significantly positive ($P < 0.01$); it supports the monophyly of nucleomorphs as well as the animal kingdom (22) and the parvkingdom Alveolata (3). It shows Chromista as paraphyletic—i.e., Alveolata as derived from chromists and the sister group of Heterokonta.

rophyll a plus c as in Chromobiota, might have evolved from a photosynthetic chromobiote. Since the dinoflagellate chloroplast has an envelope of three membranes, this hypothesis implies that the third membrane is homologous with the chromist periplastid membrane and that the chromistan periplastid rough endoplasmic reticulum was lost in the ancestor of dinoflagellates, making Chromophyta (algae with chlorophyll c) monophyletic (27), not polyphyletic as usually thought (2–5, 24–29). This hypothesis requires that all nonphotosynthetic Alveolata (ciliates, apicomplexa, and dinozoans) are secondarily so, but it is simpler than earlier ones (4, 24, 25) in that it does not require a separate symbiotic origin for the dinoflagellate plastid. Apicomplexa have a 35-kb cytoplasmic genome that some interpret as a relic plastid genome (3).

If the ancestral chromist was a photophagotroph, then *Goniomonas* and the oomycete ancestor also must have totally lost plastids (the latter is supported by Fig. 3). Assuming such loss is in some ways more parsimonious than the common assumption of the independent acquisition by the four chromist phyla of chloroplasts (10, 28) and of the novel protein-targeting systems across four separate membranes into them. However, the major weakness in the light of our trees of the theory of a single secondary symbiotic origin for all euchromistan chloroplasts (24) is the fact that within every one of the three phyla the taxa with plastids appear to radiate very much later than the divergence between the three phyla. If the ancestor of each phylum were photosynthetic, we should expect at least one instance within a phylum of an early divergence between two photosynthetic taxa, yet we see none. Moreover, while in principle it is possible that the dinoflagellates obtained their chloroplasts by direct descent from a chromobiote as discussed above, the same problem arises here: the dinoflagellates diverge from chromobiotes much more deeply than do any photosynthetic dinoflagellates from each other. An additional important point is that chromophyte algae (fossils are known only for dinoflagellates, heterokonts, and haptophytes) first appear in the fossil record much later than green and red algae. Given this agreement between the molecular trees and the fossil record we think it quite likely that there were four separate secondary transfers of chloroplasts into phagotrophic protists to create the chromophyte algae (4).

But how can we circumvent the two main obstacles to accepting a polyphyletic (4) rather than a monophyletic (28, 29) origin for the euchromist algae: first, the coincidence that they all have chlorophyll c, rather than b as in green plants, or phycobilisomes as in red algae; and, second, the perceived difficulty (2) of multiple origins of a protein-targeting mechanism across the four euchromist membranes. One way to solve both problems simultaneously is to postulate that one of the three phyla first evolved both chlorophyll c and a protein-import mechanism across four membranes and then became the endosymbiont that was independently implanted into the other two phyla. If the symbiont for the second and third symbiotic events already had nuclear genes encoding the complex protein-import mechanism, and also genes for numerous chloroplast proteins already each bearing the requisite topogenic sequences (31) for import across four membranes, then the simple transfer of these genes into the host nucleus would immediately have provided the new chimeric cell with a fully fledged protein-import mechanism. The major innovations necessary would have been the loss of the symbiont's plasma membrane and the fusion of its nuclear envelope with that of the host: this would simultaneously put its genes in the host nucleus and recreate the characteristic euchromist membrane topology. Membrane loss and fusion are much less mutationally onerous than the origin of a new protein-import system. So the problem raised earlier (2) about multiple origins of complex protein-topogenic mechanisms can be solved more easily than once envisaged. Furthermore, this model readily explains why all euchromists have chlorophyll c. Once one such chimeric cell evolved, it was much better preadapted to serve as an endosymbiont capable of being permanently incorporated in a second or third independent chimeric merger between two eukaryotes than was a green or red alga. Since cryptomonads contain phycobilins (like red algae), possibly they evolved first by incorporating a red algal symbiont and then, after evolving chlorophyll c and a novel protein-import mechanism, served as secondary symbionts to create the next chromophyte—i.e., the first algal chromobiote. For the third chimeric event

that created the second group of algal chromobiotes, either an early cryptomonad or, more likely, the first chromobiote would have served as the endosymbiont.

Since the common stem for the nucleomorph branch on our trees is so short, the ancestors of the cryptomonad and *Chlorarachnion* nucleomorphs probably diverged from each other before either was incorporated as a symbiont. In fact, this common stem is so short that we think the single nucleomorph branch on our present trees may be a long branch attracts artifact (32). Indeed, we find that it can be unstable when the taxa included in the tree are varied. With some combinations of taxa the cryptomonad nucleomorphs branch with red algae and the *Chlorarachnion* ones with the green algae. All of the above considerations make it likely that the traditional view (2, 6) of an independent origin of cryptomonad and *Chlorarachnion* nucleomorphs is correct.

However, we find that the grouping together on our parsimony and maximum likelihood trees of red and green algae [often not demonstrable on distance trees (e.g., refs. 3, 16, and 27)] is reproducible even when we change taxa in the tree. This agrees with recent evidence in red algae for a homologue of the chlorophyll a/b and a/c binding proteins (33), which strongly supports the monophyly of all chloroplasts despite their pigment diversity (24) and the classification of both red and green algae in the kingdom Plantae (1), as does the recognition of red algal transit sequences by green plant plastids (34) and the similarity in glyceraldehyde phosphate dehydrogenase intron positions of red and green algae (35).

By contrast, the concept of a kingdom Chromista (1, 2) now seems less phylogenetically tenable. Instead, it may be preferable to treat Cryptista, Heterokonta, and Haptophyta as separate kingdoms, as proposed much earlier (36, 37). However, their branching order on the rRNA trees is not sufficiently robust for us to rule out the possibility that future evidence may make it desirable to group some of the chromist phyla together [notably Heterokonta and Haptophyta, which share several nonplastid characters (30)], or with another group—e.g., Alveolata—in a higher level taxon. Moreover, if our hypothesis is correct concerning the serial lateral transfer by nuclear fusion between the three original chromist phyla, of hundreds of genes with topogenic sequences for protein import across four membranes, and of the genes for recognizing these, then the algae in these three phyla would indeed share a unique common genetic inheritance: their membrane topology could be regarded as monophyletic despite being transmitted by lateral rather than vertical inheritance. In such a chimeric case classical distinctions between monophyly and polyphyly become problematic. Especially when discussing their shared membrane topology and its molecular biology, we suggest that it will be useful to continue to use the term chromist informally to designate these three phyla.

We thank T. Chappell and L. Chan for typing, G. Matthews for extracting the DNA, the Natural Sciences and Engineering Research Council of Canada for a research grant, G. I. McFadden and U.-G. Maier for supplying sequences and preprints prior to publication, and the referees for helpful comments.

1. Cavalier-Smith, T. (1981) *BioSystems* **14**, 461–481.
2. Cavalier-Smith, T. (1986) in *Progress in Phycological Research*, eds. Round, F. E. & Chapman, D. J. (Biopress, Bristol, U.K.), Vol. 4, pp. 309–347.
3. Cavalier-Smith, T. (1993) *Microbiol. Rev.* **57**, 953–994.
4. Whatley, J. M., John, P. & Whatley, F. R. (1979) *Proc. R. Soc. London B* **204**, 165–187.
5. Cavalier-Smith, T. (1989) in *The Chromophyte Algae: Problems and Perspectives*, eds. Green, J. C., Leadbeater, B. S. C. & Diver, W. C. (Clarendon, Oxford), pp. 379–405.
6. Hibberd, D. J. (1990) in *Handbook of Protoctista*, eds. Margulis, L., Corliss, J. O., Melkonian, M. & Chapman, D. J. (Jones & Bartlett, Boston), pp. 288–292.
7. Maier, U.-G. (1992) *BioSystems* **28**, 69–73.
8. McFadden, G. I. (1992) *Adv. Bot. Res.* **31**, 100–121.
9. Sitte, P. (1993) *Eur. J. Protistol.* **29**, 131–143.
10. Douglas, S. E., Murphy, C. A., Spencer, D. F. & Gray, M. W. (1991) *Nature (London)* **350**, 148–151.
11. Maier, U.-G., Hofmann, C. J. B., Eschbach, S., Wolters, J. & Igloi, G. L. (1991) *Mol. Gen. Genet.* **230**, 155–160.
12. Cavalier-Smith, T. (1991) in *Fundamentals of Medical Cell Biology*, ed. Bittar, G. E. (JAI, Greenwich, CT), Vol. 1, pp. 217–272.
13. Ludwig, M. & Gibbs, S. P. (1989) *J. Phycol.* **25**, 385–394.
14. Cavalier-Smith, T. (1992) *BioSystems* **28**, 91–106.
15. Medlin, L., Elwood, H. J., Stickel, S. & Sogin, M. L. (1988) *Gene* **71**, 491–499.
16. Bhattacharya, D., Elwood, H. J., Goff, L. J. & Sogin, M. L. (1990) *J. Phycol.* **29**, 29–320.
17. McFadden, G. I., Gilson, P. R. & Douglas, S. E. (1994) *J. Cell Sci.* **107**, 649–659.
18. McFadden, G. I., Gilson, P. R. & Hill, D. R. A. (1994) *Eur. J. Phycol.* **29**, 29–32.
19. McFadden, G. I., Gilson, P. R., Hofmann, C. J. B., Adcock, G. J. & Maier, U.-G. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 3690–3694.
20. Knoll, A. H. (1992) *Science* **256**, 622–627.
21. Cavalier-Smith, T. (1990) *Rev. Micropal.* **331**, 145–154.
22. Wainright, P. O., Hinkle, G., Sogin, M. L. & Stickel, S. K. (1993) *Science* **260**, 340–342.
23. Felsenstein, J. (1978) *Syst. Zool.* **27**, 401–410.
24. Cavalier-Smith, T. (1982) *Biol. J. Linn. Soc.* **17**, 289–306.
25. Schnepf, E. (1993) in *Origins of Plastids*, ed. Lewin, R. A. (Chapman & Hall, New York), pp. 53–76.
26. Gibbs, S. P. (1993) in *Origins of Plastids*, ed. Lewin, R. A. (Chapman & Hall, New York), pp. 107–121.
27. Christensen, T. (1989) in *The Chromophyte Algae: Problems and Perspectives*, eds. Green, J. C., Leadbeater, B. S. C. & Diver, W. C. (Clarendon, Oxford), pp. 1–12.
28. Bhattacharya, D., Medlin, L., Wainright, P. O., Ariztia, E. V., Bibeau, C., Stickel, S. K. & Sogin, M. L. (1992) *Evolution* **46**, 1801–1817.
29. Cavalier-Smith, T. (1993) in *Origins of Plastids*, ed. Lewin, R. A. (Chapman & Hall, New York), pp. 291–348.
30. Cavalier-Smith, T. (1994) in *The Haptophyte Algae*, eds. Green, J. C. & Leadbeater, B. S. C. (Clarendon, Oxford), pp. 413–435.
31. Grossman, A. R., Manodori, A. & Snyder, D. (1990) *Mol. Gen. Genet.* **224**, 91–100.
32. Cavalier-Smith, T. (1988) in *McGraw-Hill 1989 Yearbook of Science and Technology* (McGraw-Hill, New York), pp. 175–179.
33. Wolfe, G. R., Cunningham, F. X., Durnford, D., Green, B. R. & Gantt, E. (1994) *Nature (London)* **367**, 566–568.
34. Apt, R. E., Hofmann, W. E. & Grossman, A. (1993) *J. Biol. Chem.* **268**, 16208–16215.
35. Zhou, Y.-M. & Ragan, M. A. (1993) *Curr. Genet.* **23**, 483–489.
36. Leedale, G. F. (1974) *Taxon* **32**, 261–270.
37. Cavalier-Smith, T. (1978) *BioSystems* **10**, 93–114.