J Mol Evol (2000) 51:26–40
DOI: 10.1007/s002390010064

JOURNAL OF MOLECULAR EVOLUTION

Phylogeny of Ultra-Rapidly Evolving Dinoflagellate Chloroplast Genes: A Possible Common Origin for Sporozoan and Dinoflagellate Plastids

Zhaoduo Zhang, Beverley R. Green, Thomas Cavalier-Smith*

Department of Botany, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

Received: 27 December 1999 / Accepted: 24 March 2000

Abstract. Complete chloroplast 23S rRNA and *psbA* genes from five peridinin-containing dinoflagellates (*Heterocapsa pygmaea, Heterocapsa niei, Heterocapsa rotundata, Amphidinium carterae,* and *Protoceratium reticulatum*) were amplified by PCR and sequenced; partial sequences were obtained from *Thoracosphaera heimii* and *Scrippsiella trochoidea.* Comparison with chloroplast 23S rRNA and *psbA* genes of other organisms shows that dinoflagellate chloroplast genes are the most divergent and rapidly evolving of all. Quartet puzzling, maximum likelihood, maximum parsimony, neighbor joining, and LogDet trees were constructed. Intersite rate variation and invariant sites were allowed for with quartet puzzling and neighbor joining. All *psbA* and 23S rRNA trees showed peridinin-containing dinoflagellate chloroplasts as monophyletic. In *psbA* trees they are related to those of chromists and red algae. In 23S rRNA trees, dinoflagellates are always the sisters of Sporozoa (apicomplexans); maximum likelihood analysis of *Heterocapsa triquetra* 16S rRNA also groups the dinoflagellate and sporozoan sequences, but the other methods were inconsistent. Thus, dinoflagellate chloroplasts may actually be related to sporozoan plastids, but the possibility of reproducible long-branch artifacts cannot be strongly ruled out. The results for all three genes fit the idea that dinoflagellate chloroplasts originated from red algae by a secondary endosymbiosis, possibly the same one as for chromists and Sporozoa. The marked disagreement between 16S rRNA trees using different phylogenetic algorithms indicates that this is a rather poor molecule for elucidating overall chloroplast phylogeny. We discuss possible reasons why both plastid and mitochondrial genomes of alveolates (Dinozoa, Sporozoa and Ciliophora) have ultra-rapid substitution rates and a proneness to unique genomic rearrangements.

Key words: Dinoflagellate — Sporozoa — Chloroplast — Ribosomal RNA — *psbA* — Red alga — Chromist — Phylogeny — Alveolates — Mutation pressure

Introduction

Dinoflagellates are a very diverse group of unicellular eukaryotes, about half being photosynthetic, whereas the rest are heterotrophs. Photosynthetic dinoflagellates are important primary producers in marine and freshwater ecosystems, both as free-living algae and as symbionts within corals, while some are notorious for causing toxic "red tides" and killing fish. Typical dinoflagellate chloroplasts are characterized by the presence of chlorophyll c_2 and peridinin (Jeffrey et al. 1975). They differ from all other chloroplasts (except those of euglenoids) in being surrounded by an envelope of three membranes (Dodge 1975), rather than two membranes like red algal, glaucophyte, and green plant chloroplasts, or four membranes as in chromists, chlorarachneans, and the sporozoan *Toxoplasma.* The origin of this unusual membrane topology of typical dinoflagellate chloroplasts has been much debated. Recent phylogenetic analyses (Zhang et

* *Present address:* Department of Zoology, University of Oxford, South Parks Road, Oxford, 0X1 3PS, UK
Correspondence to: T. Cavalier-Smith

al. 1999; Takashita and Uchida 1999) favor the view (Gibbs 1981) that peridinin-containing dinoflagellate chloroplasts are the result of a secondary endosymbiosis like those of chromists (e.g., cryptophytes: Douglas et al. 1991) and chlorarachneans (McFadden et al. 1994), but with the loss of one of the four membranes (Palmer and Delwiche 1998; Cavalier-Smith 1999, 2000).

Peridinin-containing dinoflagellate chloroplasts are also extremely different in genome organization from those of other photosynthetic organisms. Unlike in other algae and higher plants, where chloroplast genomes are single large circular molecules (around 120–200 kb) containing approximately 140 to 250 genes (Palmer 1985; Sugiura 1995; Reith 1995; Turmel et al. 1999a), each chloroplast gene of many peridinin-containing dinoflagellates is on a separate minicircle (Zhang et al 1999). In this respect they differ greatly from their nonphotosynthetic sister group, the parasitic Sporozoa (e.g., malaria parasites like *Plasmodium falciparum*), one subgroup of which (the Coccidiomorpha; Cavalier-Smith 1998) has relict plastids with many different genes on a much reduced single circular genome of around 35 kb (Wilson et al. 1996). Sporozoans (apicomplexans) and dinoflagellates are grouped with the less-well-characterized protalveolate flagellates (e.g., *Perkinsus, Colpodella*) and with ciliates as a major protist group, known as alveolates (Cavalier-Smith 1991, 1993a, 1998), because membrane-bounded cortical alveoli are key shared features. The evolutionary unity of the alveolates is well supported by other ultrastructural characters as well as by nuclear 18S rRNA gene sequences (Gajadhar et al. 1991; Cavalier-Smith 1993a) and actin sequences (Reece et al. 1997). A key question, therefore, is whether the plastids of dinoflagellates and Sporozoa are directly related and were present in their immediate alveolate common ancestor (Palmer 1992; Cavalier-Smith 1999) or were acquired instead by independent endosymbiotic events (Köhler et al. 1997).

Phylogenetic analyses of seven photosynthetic chloroplast proteins of *H. triquetra* suggested that dinoflagellate chloroplasts are related to those of chromists and red algae (Zhang et al. 1999). However, as the relict sporozoan plastids have lost all photosynthetic genes, the relationship of dinoflagellate chloroplasts and sporozoan plastids is still not clear. Several lines of evidence suggest that sporozoan plastids may be related to red algal and chromistan plastids (Williamson et al. 1994; McFadden and Waller 1997; Blanchard and Hicks 1999) like those of dinoflagellates (Zhang et al. 1999; Takashita and Uchida 1999), which favors the view that sporozoan and dinoflagellate plastids may be directly related despite their extremely different genome organization (Palmer 1992; Cavalier-Smith 1999). However, protein synthesis elongation factor *tufA* trees weakly suggested that sporozoan plastids might have originated instead from a green alga (Köhler et al. 1997). Our present phylogenetic analyses of ribosomal RNA and photosystem II *psbA* genes from various dinoflagellates support the conclusion from *H. triquetra* chloroplast proteins that peridinean dinoflagellate plastids are derived from a red alga, like those of chromists. Furthermore, both ribosomal RNA trees suggest that dinoflagellate chloroplasts are related to sporozoan plastids, but the branches are so long that additional data from nuclear-encoded plastid proteins will be needed to eliminate the possibility that this is simply an artifactual attraction of long branches.

Understanding the evolution of dinoflagellate chloroplasts is complicated by the fact that a few aberrant derived dinoflagellate lineages have replaced their ancestral peridinin-containing chloroplasts by differently pigmented chloroplasts acquired secondarily from unrelated algae. Judging by their diverse pigment composition and presence of additional membranes (Palmer and Delwiche 1998), it is probable (but not yet demonstrated) that those replacement plastids derived from haptophytes (e.g., *Gymnodinium breve;* Delwiche 1999; Tengs et al. 2000), cryptomonads (*Dinophysis;* Schnepf and Elbrächter 1988), and prasinophytes (*Lepidodinium;* Watanabe et al. 1990). Some of these (e.g., *Lepidodinium*) seem likely to have become fully integrated with the host cell by acquiring a protein-import mechanism, but this is dubious for those from diatoms (Chesnick et al. 1997) as the diatom nuclei are still present within the endosymbiont. These replacement plastids, however, are not directly relevant to the origin of the more typical peridinin-containing dinoflagellate plastids, which our trees for both 23S rRNA and *psbA* confirm are monophyletic.

Materials and Methods

DNA Extraction. Total DNAs were extracted from the dinoflagellates *Heterocapsa pygmaea* (CCMP 1490), *Heterocapsa niei* (CCMP 447), *Heterocapsa rotundata* (NEPCC D680), *Protoceratium reticulatum* (NEPCC D535, formerly known as *Gonyaulax grindleyi*), *Scrippsiella trochoidea* (NEPCC D602), *Thoracosphaera heimii* (NEPCC D670) as described for *H. triquetra* (Zhang et al. 1999). The same method was used for DNA extraction from *Amphidinium carterae* (CCMP 1314) but without vortexing with glassbeads. Only *A. carterae* DNA was purified by CsCl gradients.

PCR Reactions. Specific dinoflagellate chloroplast 23S rRNA and *psbA* primers were designed based on the *H. triquetra* 23S RNA and *psbA* sequences (Zhang et al. 1999). Degenerate primers were based on all available chloroplast 23S rRNA and *psbA* gene sequences (Table 1; Fig. 1).

Primer pair 23S1/23S2 was used to amplify the most conserved region (~0.7 kb) of 23S rRNA genes from the eight dinoflagellate species. Primers 23S2/23S3 (covering the 23S1/23S2 region) were used to amplify a large fragment (~1.6 kb) from *Heterocapsa* species. Primers 23S1/23S4 and D23S1/D23S2 were used to amplify the rest of the 23S rRNA minicircles, including the noncoding region, from *Heterocapsa* species, *A. carterae,* and *P. reticulatum.* Primer pairs bA6/bA7 and bA1/bA5 were used to amplify *psbA* minicircles from *Heterocapsa* species; bAf1/bAr1, bAf3/bAr1, and bAf2/bAr2 were used to amplify the *psbA* minicircles from other dinoflagellates. PCR reactions were

Table 1. Dinoflagellate chloroplast 23S rRNA and psbA primers

Name	Sequence
23S1	5′ GGCTGTAACTATAACGGTCC3′
23S2	5′ CCATCGTATTGAACCCAGC3′
23S3	5′ ATAAGTGGTTGTAGAAGAAAG3′
23S4	5′ TAATTCTTTCTTCTACAACCAC3′
D23S1	5′ YTACYCWAGGGWTAACAG3′
D23S2	5′ TTMWATSTTTCATGCAGG3′
bA6	5′ GCAAGATCAAGTGGGAAGTTG3′
bA7	5′ GCTCCACCAGTCGATATTG3′
bA1	5′ CCAAGAGCTTCCCAAACTG3′
bA5	5′ CAACTTCCCACTTGATCTTGC3′
bAf3	5′ ATCTTCGCTCCACCAGTTGAYATHGAYGG3′
bAf1	5′ GGTCAAGGTTCTTTCTCTGAYGGNATGCC3′
bAr1	5′ GTTGTGAGCGTTACGTTCRTGCATNACYTC3′
bAf2	5′ GTTAGTACAATGGCTTTCAAYYTNAAYGG3′
bAr2	5′ GGCATACCATCAGAGAATCWNCCYTGNCC3′

carried out for 35 cycles: 94°C for 30 s, 55°C for 30 s followed by 2 min at 72°C in a GeneAmp PCR system 9600 (Perkin-Elmer). Each reaction (50 μl) contained 0.2 m*M* dNTP, 1 × PCR buffer, 0.1–1.0 μg template DNA, 50–200 nmol primer, 2.0 or 2.5 m*M* $MgCl_2$, 1.5–2.5 units Taq polymerase (Sigma or Rose). PCR products were purified from low melting gels or using a purification kit (Amersham-Pharmacia Biotech) and then used for sequencing.

The PCR product of *P. reticulatum* from primer pair D23S1/D23S2 was cloned into pCR-TOPO vector using TOPO TA cloning kit (Invitrogen) according to the manufacturer's instructions.

DNA Sequencing. Sequencing reactions were done in the Perkin-Elmer GeneAmp 9600 using the ABI cycle sequencing protocol: 94°C for 5 s, 50°C for 5 s, 60°C for 4 min for 25 cycles. Each reaction contained 2–3 μl FS-Taq or Bigdye, 20–30 ng DNA (purified PCR product), 3–5 nmol primer, plus distilled water to 10 μl. The sequencing samples were precipitated by adding 1/10 volume sodium acetate (pH 5.2) and 2 volumes 95% ethanol, quenched on ice for 10 min, centrifuged for 20 min, air-dried, and analyzed by an ABI 373 or 377 automatic sequencer. Traces from the sequencer were edited by Staden's Trev program; contigs were generated by Staden's Gap4 program (http://www.mrc-lmb.cam.ac.uk/pubseq).

Phylogenetic Analyses. 23S rRNA and 16S rRNA sequence alignments were from the rRNA database (http://rrna.uia.ac.be). Additional 23S rRNA sequences of *Plasmodium falciparum* (X95275-6), *Plasmodium berghei* (U79732), *Toxoplasma gondii* (U87145), the dinoflagellate *H. triquetra* (CCMP449, AF130039), and seven other dinoflagellates (*H. pygmaea, H. niei, H. rotundata, A. carterae, P. reticulatum, S. trochoidea,* and *T. heimii*) were aligned manually with those from the rRNA database in GDE (ver. 2.2) (Smith et al. 1994). Three masks were used in phylogenetic analyses: mask1 (1316 bp from very conserved regions), mask2 (1885 bp, less stringest than mask1), mask3 (2033 bp, the positions from mask2 plus adjacent regions where a few taxa have small deletions). *psbA* DNA and amino acid sequences were aligned with *psbA* from various photosynthetic organisms. Maximum likelihood trees were constructed with global rearrangement and four different jumbles using fastDNAml (Olsen et al. 1994). Rate heterogeneity (invariable sites and among site rate variation) was taken into account (8 gamma rates + 1 invariable) in constructing quartet puzzling trees using the HKY 85 model of PUZZLE (ver 3.1) with 1000 puzzling steps (Strimmer and von Haeseler 1996). The following trees were constructed using Paup* 4.0 with the heuristic search option (Swofford 1999): the maximum parsimony tree was constructed by optimizing the characters with accelerated transformation (ACCTRAN); the LogDet tree (LogDet/paralinear distance) and neighbor joining tree (HKY 85 distance, gamma distribution rate with shape parameter estimated from PUZZLE) were optimized by minimum evolution, starting tree(s) being obtained by neighbor joining, and branches were swapped by tree-bisection-reconnection (TBR).

Phylogenetic analyses of *psbA* nucleotides (only the first and second nucleotides of the triplet codon were used) and amino acid sequences were carried out by neighbor joining and parsimony using Phylip (ver. 3.5; Felsenstein 1993) and by fastDNAml. Parsimony and maximum likelihood trees were constructed using global rearrangement and neighbor joining trees using the PAM matrix (for protein) or Kimura distances (for DNA). The input order of taxa was jumbled. Bootstrap analyses of both 23S rRNA and *psbA* data sets were based on 500 resamplings.

Results

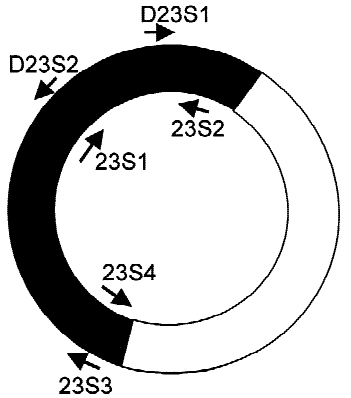
PCR Amplification of 23S rRNA Genes

Primer pair 23S1/23S2 that amplifies the most conserved region of the 23S rRNA gave a product of approximately 0.7 kb from three close relatives of *H. triquetra: H. pygmaea, H. niei, H. rotundata,* and four very distant relatives: *A. carterae, P. reticulatum, T. heimii,* and *S. trochoidea.* Almost complete gene sequences were obtained from the *Heterocapsa*s using primer pair 23S2/23S3. In the inverse PCR reactions for amplifying the rest of the minicircles, primer pairs 23S1/23S4 produced one band in *H. pygmaea, H. niei,* and *H. rotundata,* but degenerate primers D23S1/D23S2 (Fig. 1) were needed to amplify bands from *P. reticulatum* and *A. carterae.* For each of these species, the sequences of PCR products overlapped and generated a single contig that could be circularized to a minicircle. However, neither primer pair 23S1/23S4 nor D23S1/D23S2 yielded any product for *T. heimii* and *S. trochoidea;* this is consistent with evidence from Southern blots (not shown) that their 23S rRNA and *psbA* genes are not organized as minicircles but are on much larger DNA molecules (Zhang et al. in preparation).

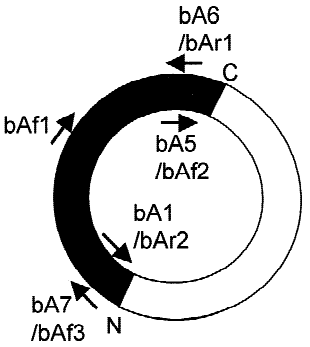
Dinoflagellate chloroplast 23S rRNA genes are extremely divergent, so it is very difficult to align some regions with other chloroplast 23S rRNAs. Close to the 3′ end of 23S rRNA is the most conserved part (~0.7 kb), which is easy to align. The region (~1.3 kb) upstream is very divergent between dinoflagellates and other organisms but very similar among the *Heterocapsa* species. The *P. reticulatum* 23S rRNA gene has a 166-bp insertion (approximately 160 bp upstream of the 3′ terminus) that is not present in any other chloroplast 23S rRNA genes.

The Dinoflagellate 23S rRNA Sequences Are Chloroplast Genes

All maximum likelihood and parsimony trees gave two clearly resolved groups: a cyanobacteria/chloroplast group, and a noncyanobacterial bacteria/mitochondria



H. triquetra 23S rRNA (3,027 bp)



H. triquetra psbA (2,151 bp)

Fig. 1. Structure of *H. triquetra* minicircles encoding 23S rRNA and *psbA*. Coding region, black; noncoding region, white (Zhang et al. 1999). Primers were designed based on *H. triquetra* gene sequences, except for degenerate primers D23S1/D23S2 and bAf/bAr.

group with the mitochondria allied to the α-proteobacteria as expected (Fig. 2). In all other trees (LogDet, neighbor joining, and quartet puzzling), chloroplasts and cyanobacteria grouped together with high bootstrap support, but mitochondria did not group with the α-proteobacteria (not shown). In all trees, the dinoflagellates form a single group within the chloroplast cluster, indicating that dinoflagellate 23S rRNA genes are chloroplast genes, not mitochondrial genes, and that dinoflagellate peridinin-containing plastids are monophyletic. However, the dinoflagellates branches are extremely long, approximately twice that of *Plasmodium* and 1.5 times longer than that of the longest mitochondrial branch, *Neurospora* (Fig. 2). There was no significant difference among the trees constructed using three different masks differing in stringency (see methods).

Figure 2 has two major anomalies inconsistent with previous work; the apparent paraphyly rather than holophyly of Sporozoa and the grouping of euglenoids with them (the alveolates) rather than with the green algae from which their plastids undoubtedly originated (Turmel et al. 1999a). The parsimony tree had the same branching order for the plastids except that heterokonts were also incorrectly paraphyletic. In neighbor joining (gamma distribution, $\alpha = 0.77$), maximum likelihood, and parsimony trees, *Plasmodium* artifactually groups with dinoflagellates, not the other sporozoan *Toxoplasma.* A similar systematic error (grouping of *Plasmodium* with dinoflagellates rather than *Theileria*) was seen on a mitochondrial cytochrome oxidase 1 (*cox1*) maximum likelihood tree (Inagaki et al. 1998). It seemed possible that these errors in tree topology were caused by base composition bias (Lockhart et al. 1994). The AT composition of the *Plasmodium* plastid genome is extremely high (A + T = 86.9%) (Wilson et al. 1996), and the 23S rRNA genes have a higher AT% than those of dinoflagellates, *Toxoplasma* and euglenoids (Table 2), which in turn are higher than those of green plants, red algae and chromists. When a LogDet tree, which has the reputation for correcting base composition bias, was constructed we found that *Plasmodium* and *Toxoplasma* did form a clade with 93% bootstrap support, and the sporozoans formed the sister group of dinoflagellates with 100% bootstrap support (inset on Fig. 1). However, the LogDet tree placed both euglenoids and alveolates below the rest of the chloroplasts and grouped the mitochondria with the plastid/cyanobacterial clade, not the α-proteobacteria. In quartet puzzling trees, the sporozoans also grouped together as a sister group of dinoflagellates.

Errors in tree topology can also result from highly unequal evolutionary rates in different taxa (Felsenstein 1978; Olsen 1987; Hillis et al. 1994). Rate variation among sites is almost universal in molecular evolution (Van de Peer et al. 1996) but was not taken into account by the fastDNAml program used for Fig. 2. To see if this was a problem, quartet puzzling trees were constructed where rate heterogeneity was considered (eight gamma categories). The puzzling tree gave the same topology as the LogDet tree for Sporozoa and dinoflagellates with 52% puzzling step support for the sporozoan clade and 55% for the alveolate grouping of Sporozoa and dinoflagellates, but the euglenoids still incorrectly grouped with the alveolates (not shown). A gamma distribution neighbor joining tree ($\alpha = 0.77$) was overall similar to the maximum likelihood tree of Fig. 2, but worse in that euglenoids and heterokonts also incorrectly appeared paraphyletic.

Relationships Among the Chloroplast 23S rRNA Genes

In an attempt to resolve the topology of the chloroplast cluster more clearly, especially the relationship between Sporozoa and dinoflagellates, phylogenetic trees were constructed using chloroplast and cyanobacterial rRNA sequences only, i.e., excluding the more distant bacterial and mitochondrial sequences which might somewhat perturb the trees using the data set of Fig. 2. In the neighbor joining tree (gamma distribution parameter $\alpha = 0.74$) (Fig. 3) and LogDet tree (Fig. 4), sporozoans form a monophyletic and holophyletic group with bootstrap supports of 84% and 94%, respectively. Sporozoa

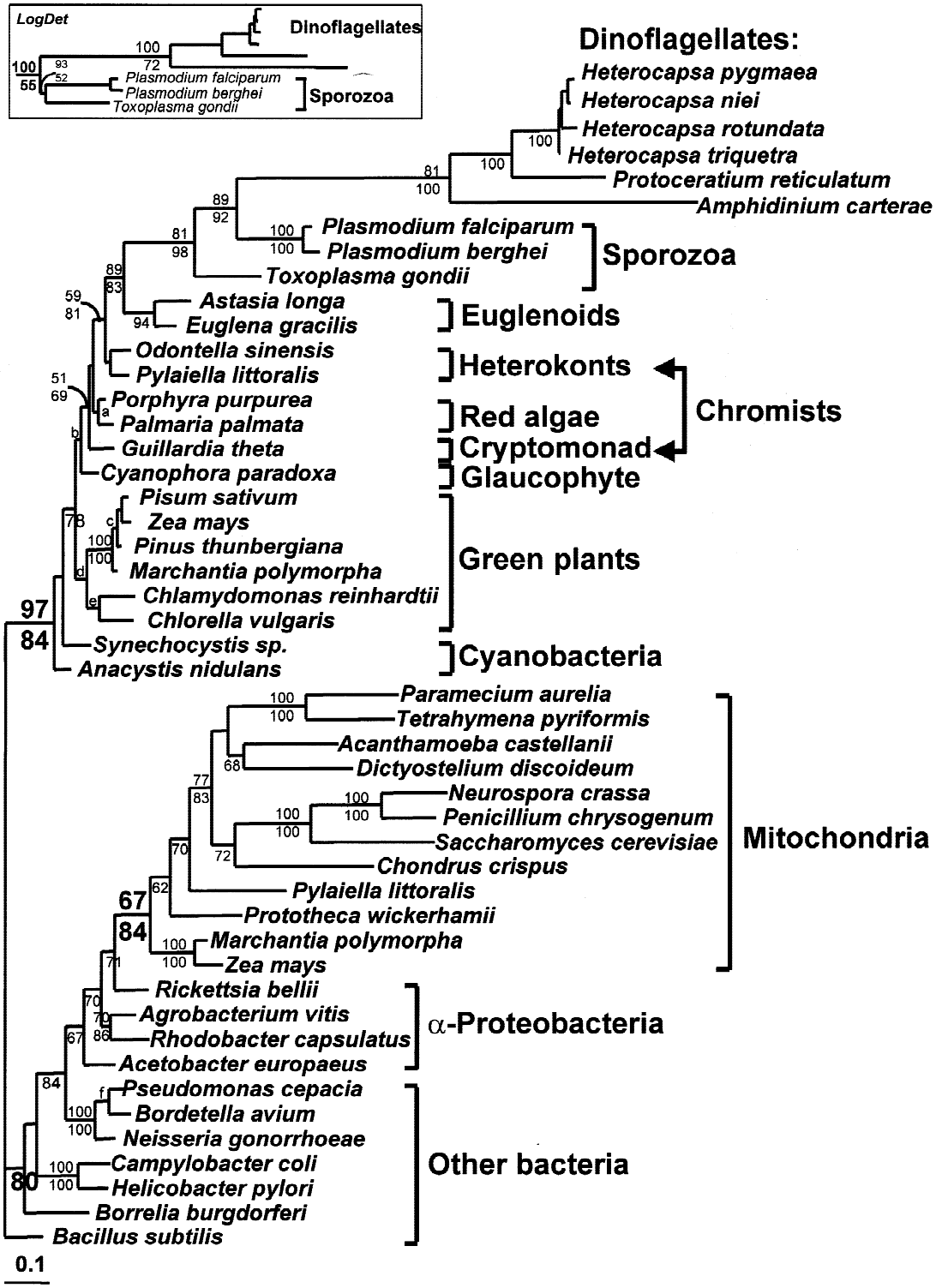


Fig. 2. Maximum likelihood tree of 23S rRNA sequences (1885 bp) of chloroplasts, mitochondria, and bacteria (ln likelihood = −49,685.99168). Bootstrap values of the gamma distribution (α = 0.77) neighbor joining tree (above), and maximum parsimony tree (below) expressed as percentage of 500 resamplings; bootstrap values of a–f: a (/85), b (/69), c (96/94), d (100/99), e (86/99), f (90/94). Bootstrap values below 50% are not shown. Inset shows a sister relationship between a holophyletic Sporozoa and dinoflagellates on the relevant part of a LogDet tree for the same 48 23S rRNA sequences; here the bootstrap values are percentages of 500 resamplings (above) or percentages of 1000 puzzling steps (below). Scale bar indicates 0.1 changes per base pair.

Table 2. Base composition of chloroplast and cyanobacterial 23S rRNA sequences

Species	Sites (bp)	AT%	
Marchantia polymorpha	1878	46.80	Higher plants
Zea mays	1879	45.40	
Pisum sativum	1875	46.20	
Pinus thunbergiana	1880	45.80	
Chlamydomonas reinhardtii	1882	50.00	Green algae
Chlorella vulgaris	1867	49.10	
Porphyra purpurea	1882	49.90	Red algae
Palmaria palmata	1881	50.20	
Cyanophora paradoxa	1883	48.90	Glaucophyte
Plasmodium falciparum	1873	76.30	Sporozoa
Plasmodium berghei	1870	74.70	
Amphidinium carterae	1768	54.60	Dinoflagellates
Protoceratium reticulatum	1853	64.40	
Heterocapsa triquetra	1877	67.80	
Heterocapsa pygmaea	1877	68.00	
Heterocapsa niei	1876	68.20	
Heterocapsa rotundata	1876	67.00	
Pylaiella littoralis	1878	51.90	Heterokonts
Odontella sinensis	1879	52.60	
Guillardia theta	1883	50.60	Cryptomonad
Anacystis nidulans	1879	46.00	Cyanobacteria
Synechocystic sp.	1880	48.00	

is the sister group of dinoflagellates in both trees with ≥98% bootstrap support (Figs. 3, 4). However, maximum likelihood, maximum parsimony, and quartet puzzling trees gave the same artifactual grouping as Fig. 2 (noninset) with respect to the sporozoan *Plasmodium* as the sister group of dinoflagellates (not shown). In quartet puzzling trees in which rate variation among sites was taken into account, the inclusion of invariable sites did not affect tree topology, although the α values were different; in all trees, the euglenoids *Euglena* and *Astasia* are still in the red algae/chromists group, misleadingly strongly supported with high bootstrap values (Figs. 3, 4). On mitochondrial *cox1* trees (Inagaki et al. 1998) Euglenozoa also artifactually cluster with Miozoa (the phylum including dinoflagellates and Sporozoa; Cavalier-Smith 1999). Neighbor joining trees (not shown) were also constructed by Treecon considering transversions only (Van de Peer and de Wachter 1994), but the tree topology was not significantly different from those of Figs. 3 and 4.

Phylogenetic trees (quartet puzzling, neighbor joining, LogDet, and maximum parsimony, not shown) were also constructed for chloroplast 23S rRNAs omitting either dinoflagellates or sporozoans or euglenoids; the remaining two were sister groups (e.g., dinoflagellates and sporozoans) within the chl a/c–red algal cluster. If two of them were omitted, the remaining one was still within the chl a/c–red algal cluster (not shown).

Phylogenetic Analyses of Chloroplast 16S rRNA Sequences

Maximum likelihood analysis using the dinoflagellate *H. triquetra* 16S rRNA gene (Zhang et al. 1999) also showed artifactual grouping of *H. triquetra* with *Plasmodium* to the exclusion of *Toxoplasma* (Fig. 5). As for 23S rRNA, LogDet correctly showed the sporozoan clade, but the long branch *Prototheca* moved to become sister to Sporozoa. *Prototheca,* which is an aberrant colorless green alga, should have grouped with the *Chlorella/Nanochlorum* clade, but it also remained with Sporozoa on gamma distribution neighbor joining trees. In such gamma trees, *H. triquetra* grouped with the long branch chlorarachnean sequences, but the position of their artifactual clade was wildly unstable, depending on whether the heuristic search option (Paup 4.0) was used or not. The relative position of the three chromist groups (cryptomonads, heterokonts, and haptophytes) differed between the trees. Because of their lack of robustness and the systematically biased grouping of long branch *Prototheca,* alveolate, euglenoid, and chlorarachnean clades, 16S rRNA trees cannot be relied on for the accurate reconstruction of chloroplast phylogeny, contrary to what is sometimes assumed (Medlin et al. 1995).

psbA *Protein Trees*

psbA genes were amplified from *H. pygmaea, H. niei,* and *H. rotundata* using primer pairs bA6/bA7 and bA1/bA5 (Fig. 1). Primer pair bA6/bA7 gave one band from the three species; bA1/bA5 yielded one band from *H. niei* and *H. rotundata,* but two bands from *H. pygmaea* with a 0.5-kb difference in size (Zhang et al. 1999). Heterogeneity of the noncoding region of the *psbA* minicircles in *H. pygmaea* was originally revealed by the presence of two bands in a blot of electrophoretically separated uncut total DNA blot hybridized with a spinach *psbA* probe (Zhang et al. 1999). Direct sequencing of the two PCR products revealed that the size difference was caused by short indels in the noncoding region, showing that *H. pygmaea* has two kinds of *psbA* minicircles, identical in the coding region but different in the noncoding region (Zhang et al. in preparation). Neither bA6/bA7 nor bA1/bA5 gave products for *A. carterae* or *P. reticulatum.* Degenerate primer pairs bAf1/bAr1 and bAf3/bAr1 did amplify *psbA* from *A. carterae, T. heimii,* and *S. trochoidea,* but only bAf1/bAr1 gave a product (0.5 kb) from *P. reticulatum* (Fig. 1). Direct sequencing of the PCR products indicated that the *psbA* gene is on a minicircle in *H. pygmaea, H. niei, H. rotundata,* and *A. carterae.* In contrast to the 23S rRNA genes, dinoflagellate *psbA* genes are well conserved in DNA and protein sequence and were easily aligned with *psbA* sequences of other organisms.

Neighbor joining and maximum parsimony trees of *psbA* protein (309 amino acids) showed two clusters: a

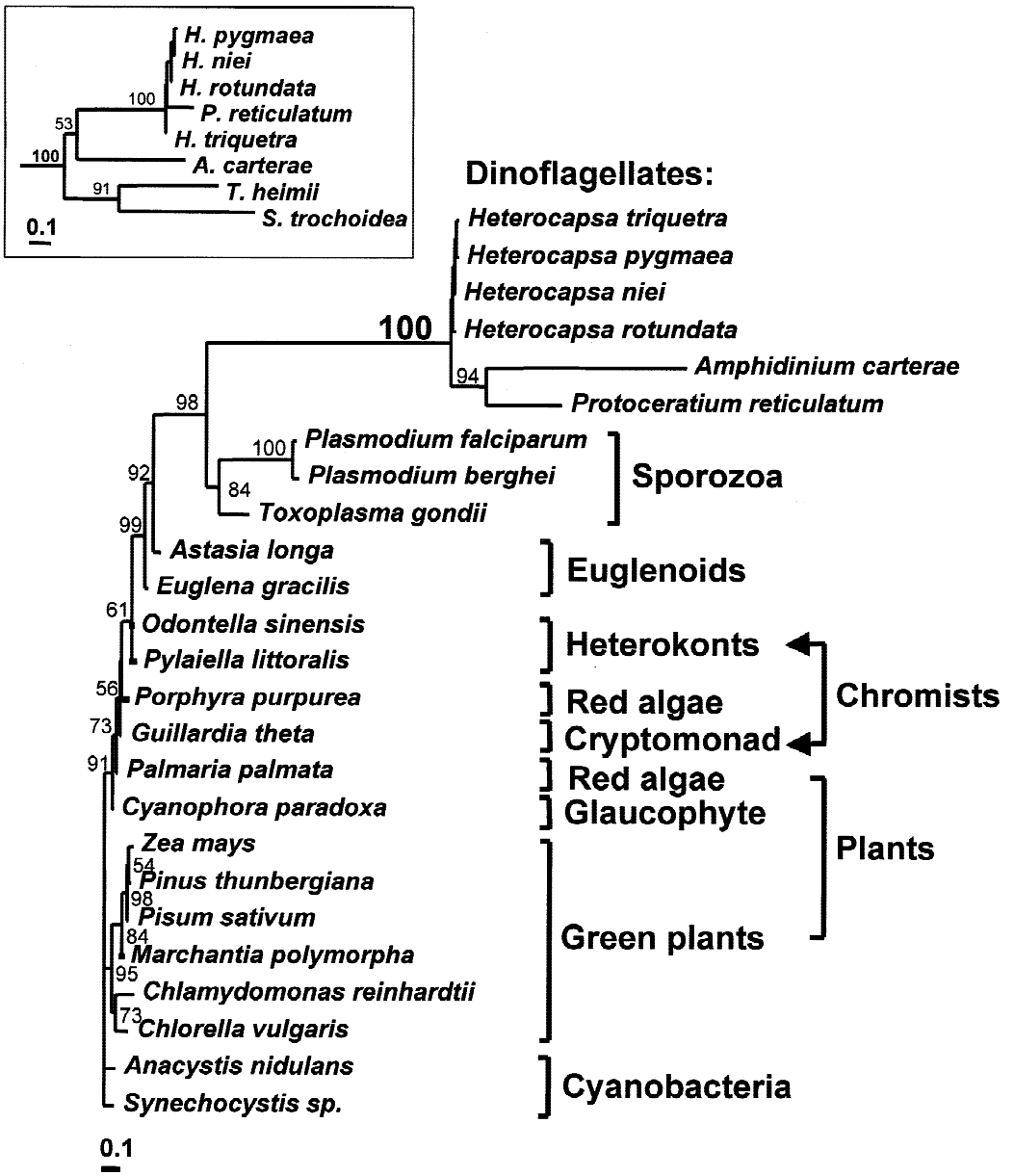


Fig. 3. Gamma distribution ($\alpha = 0.74$, 1 invariable + 8 gamma categories) neighbor joining tree of chloroplast 23S rRNA sequences (1885 bp) with bootstrap values as percentages of 500 resamplings. Inset is a gamma distribution ($\alpha = 0.57$) neighbor joining tree of partial chloroplast 23S rRNA sequences (601 bp) showing the topology of the eight dinoflagellate species. Scale bar corresponds to 0.1 changes per base pair.

green plant/euglenoid (chl a/b) cluster, and a red alga/chromist/dinoflagellate cluster (Fig. 6). In the latter, dinoflagellates are clearly monophyletic and weakly sisters to heterokonts (with low [53%] bootstrap support). *Euglena* groups with *Chlorella* in the green plant cluster. This is consistent with the phylogenetic conclusions about the topology of dinoflagellates and *Euglena* drawn from trees made with seven concatenated protein sequences (Zhang et al. 1999), i.e., it supports the idea that dinoflagellate chloroplasts are related to red algal and chromist chloroplasts, whereas *Euglena* chloroplasts are related to green algal chloroplasts. The identical C-terminal truncation of the PsbA protein precursor in *Euglena* and most dinoflagellates (Takishita and Ushida 1999; Zhang et al. 1999) compared with other eukaryotes must therefore be convergent.

The glaucophyte *Cyanophora* groups weakly with the green clade, in contrast to the 23S rRNA tree, where it is the sister to the red algal/chl a/c clade, and the 16S rRNA tree where it is basal to both as in concatenated protein trees (Martin et al. 1998; Zhang et al. 1999; Turmel et al. 1999a). Thus, the branching order of glaucophytes relative to the red and green algal clade remains uncertain.

Relationships Among the Dinoflagellates

As our chloroplast gene trees have many fewer dinoflagellate taxa than those for nuclear 18S rRNA (Saunders

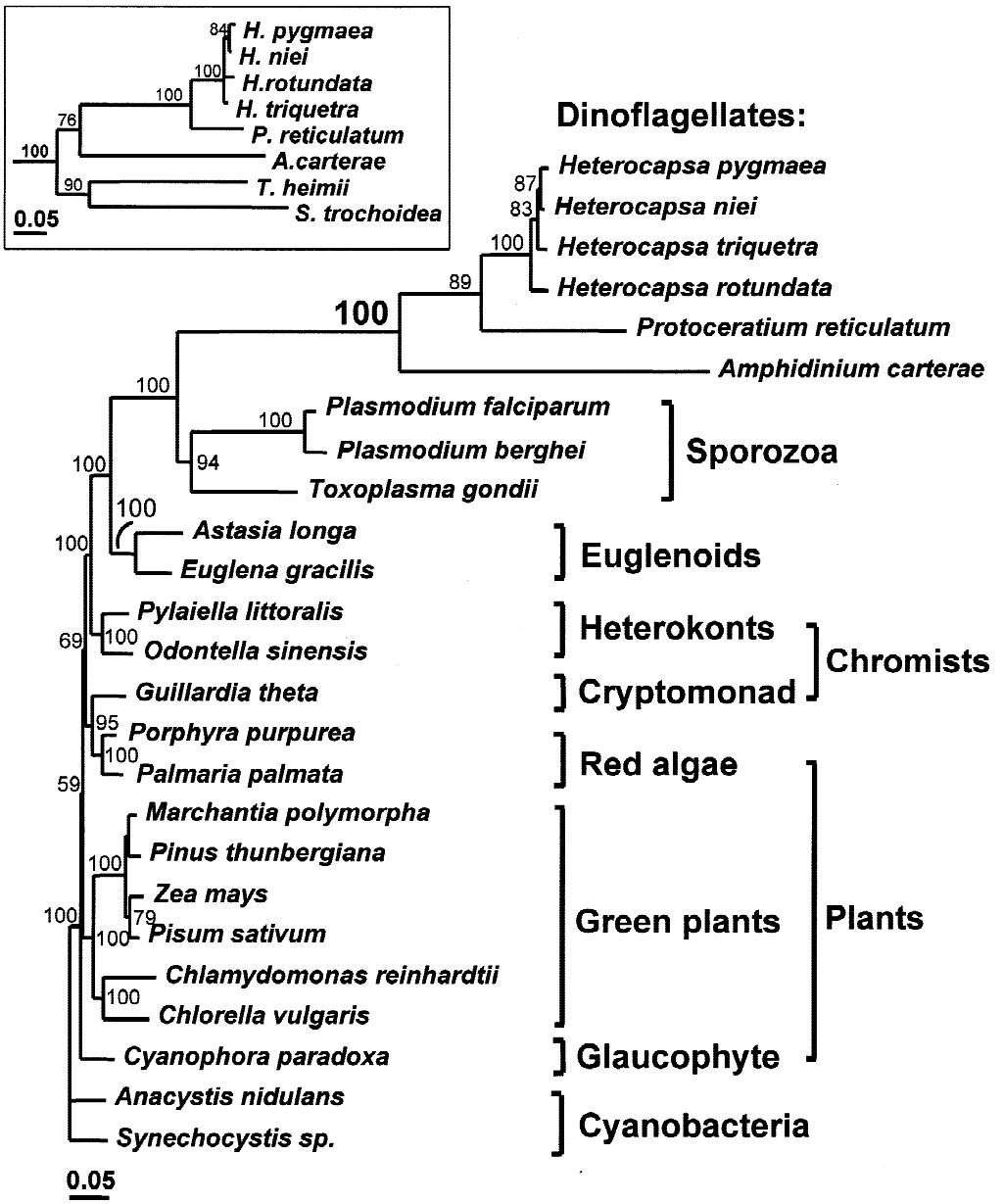


Fig. 4. LogDet tree of chloroplast 23S rRNA. Bootstrap values are percentages of 500 resamplings. Inset is from a LogDet tree of partial chloroplast 23S rRNA sequences (601 bp) showing the topology of the eight dinoflagellate species. Scale bar corresponds to 0.05 changes per base pair.

et al. 1997), they are much less satisfactory for resolving the branching order within the dinoflagellates. Nevertheless, the branching order in the gamma distribution neighbor joining tree (Fig. 3) is identical to that for 18S rRNA. Most other trees, however, suffer from the long branch problem: in Figs. 2 and 4 the extra-long branch *Amphidinium* is incorrectly at the base and Fig. 6 orders dinoflagellate taxa simply according to their branch lengths, not their known affinities. Thus in addition to dinoflagellate plastid genes having a systematically ultra-rapid evolutionary rate, there are marked rate differences between taxa that make them less suitable than nuclear genes for accurately reconstructing phylogeny.

Discussion

Ancestral Peridinean Dinoflagellates Probably Had Chloroplasts with Peridinin and Minicircles

Phylogenetic analyses of 23S rRNA and *psbA* protein confirm the dinoflagellates are a monophyletic group and indicate that peridinin-containing dinoflagellate chloroplasts all have a common ancestry. Our data set includes representatives of the two major lineages that diverged at the base of the Peridinea (Saunders et al. 1997): *Amphidinium/Protoceratium* on the one hand and *Hetero-*

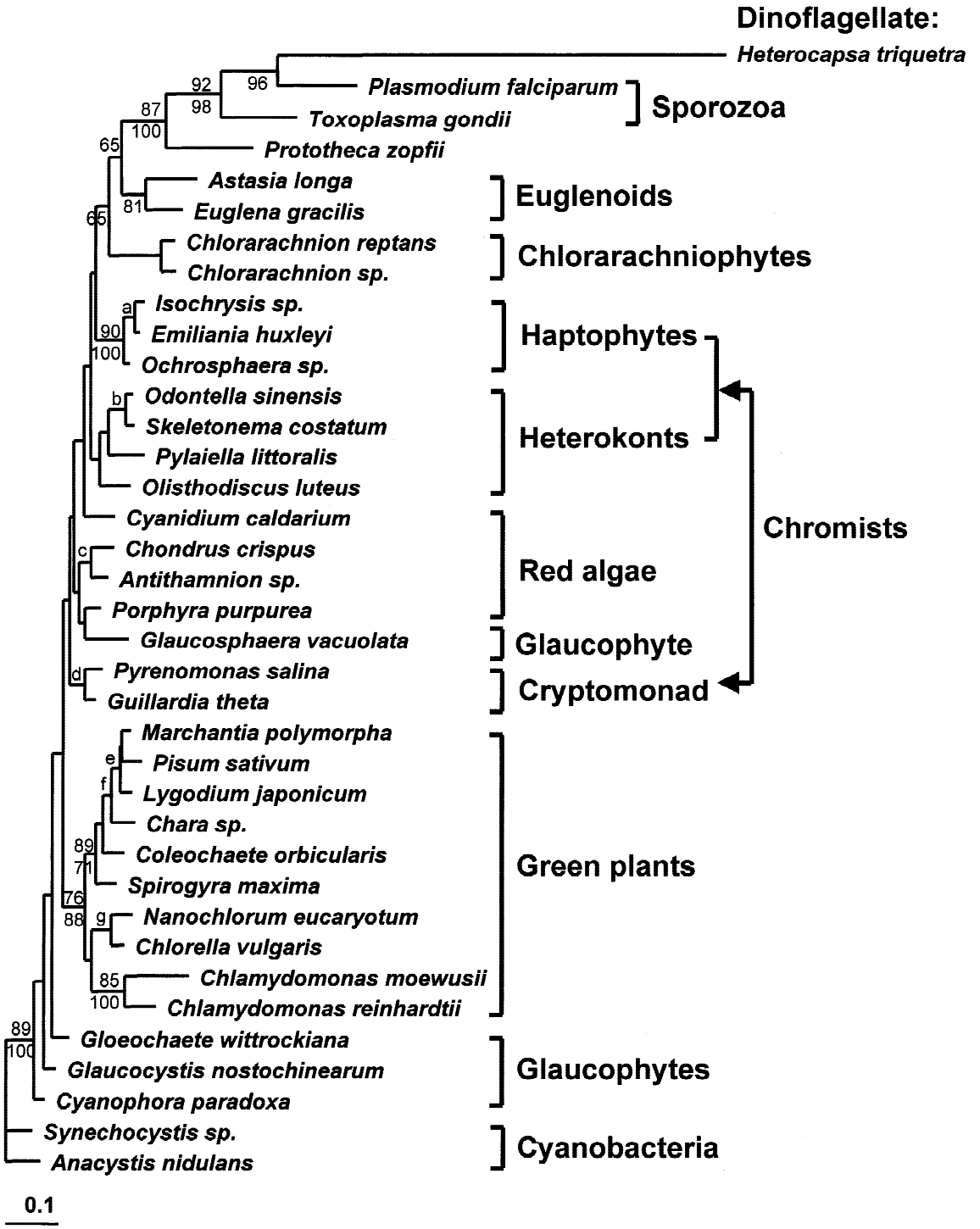


Fig. 5. Maximum likelihood tree of chloroplast 16S rRNA (1247 bp; ln likelihood = −8490.22911). Bootstrap values are the percentage of 500 replicates for separate gamma distribution neighbor ($\alpha = 0.46$, above) and maximum parsimony trees (below); bootstrap values of a–g: a (92/99), b (87/100), c (81/88), d (97/98), e (99/95), f (74/51), g (96/100). Scale bar corresponds to 0.1 changes per base pair.

capsa/Thoracosphaera/Scrippsiella on the other (Fig. 7). The monophyly of peridinin-containing dinoflagellate plastids, robustly shown by both our *psbA* and 23S rRNA trees as well as by the *psbA* data of Takashita and Uchida (1999), suggests, if the 18S rRNA tree (Saunders et al. 1997) is correct, that the latest common ancestor of peridinean dinoflagellates had a peridinin-containing plastid and that all heterotrophic peridinean dinoflagellates have secondarily lost plastids, contrary to the widespread assumption that the ancestral peridinean was nonphotosynthetic. However, since on 18S rRNA trees the two other nonphotosynthetic dinoflagellate groups (*Noctiluca* and Syndinea) diverge from photosynthetic peridineans before the split between the *Amphidinium* and *Heterocapsa*

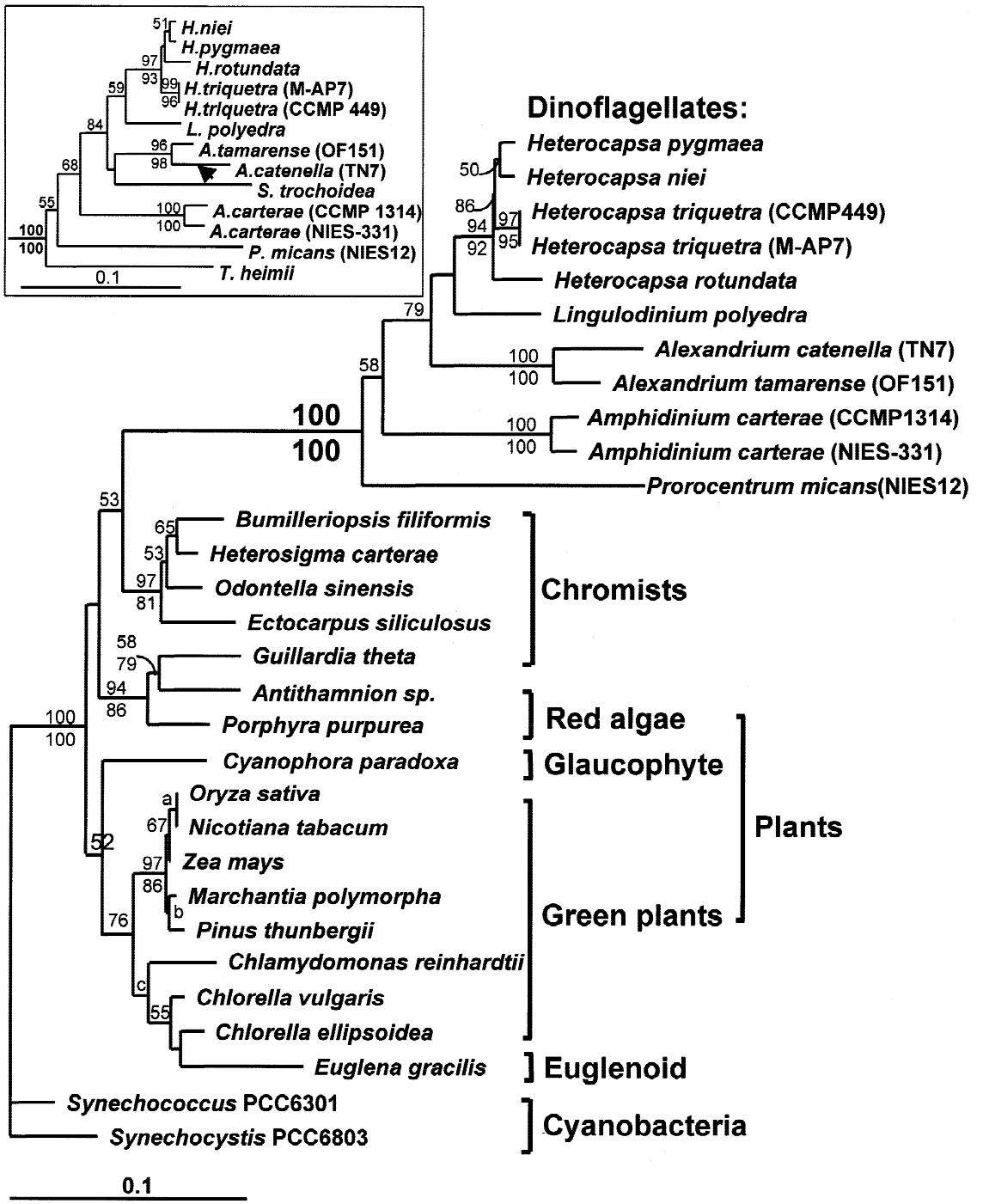


Fig. 6. Neighbor joining tree of *psbA* protein sequences (309 amino acids). Bootstrap values expressed as percentage of 500 resamplings for neighbor joining (above) and maximum parsimony (below); bootstrap values of a–c: a (74/67), b (63/), c (61/52). The following dinoflagellate *psbA* protein sequences are from Genbank: *Heterocapsa triquetra* (CCMP 449, AF130033), *Heterocapsa triquetra* (M-AP7, ABO25587), *Alexandrium catenella* (TN7, ABO25590), *Alexandrium tamarense* (OF151, ABO25589), *Amphidinium carterae* (NIES-331, ABO25586), *Prorocentrum micans* (NIES-12, ABO25585), and *Lingulodinium polyedra* (ABO25588). Inset is a neighbor joining tree of partial *psbA* protein sequences (261 amino acids) showing phylogenetic relationships among the dinoflagellates, the arrow indicates the position of *P. reticulatum* when its partial *psbA* protein sequence of (151 amino acids) is included. Scale bar corresponds to 0.1 changes per amino acid.

lineages (Saunders et al. 1997; Gunderson et al. 1999), we cannot yet conclude that the latest common ancestor of all dinoflagellates was photosynthetic, as one of us recently proposed (Cavalier-Smith 1999). As shown in Fig. 7 and deduced previously (Zhang et al. 1999), the chloroplast genome was probably fragmented into single gene minicircles in the ancestral peridinean dinoflagellate; if, as our preliminary results for *Thoracosphaera* and *Scrippsiella* suggest, some dinoflagellates eventually turn out to lack typical chloroplast DNA minicircles, but

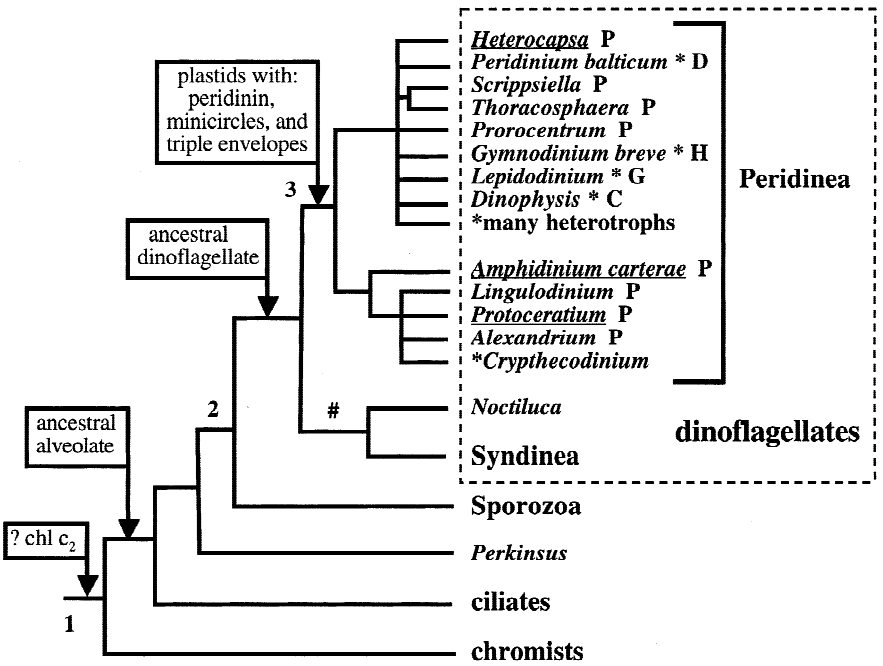


Fig. 7. Schematic relationship between selected dinoflagellates and other alveolates. The branching order follows nuclear 18S rRNA trees (Saunders et al. 1997; Gunderson et al. 1998; and our unpublished maximum likelihood analysis suggesting that *Noctiluca* and the syndinean *Amoebophrya* are sisters), except for *Lingulodininium* and *Protoceratium* for which molecular data are not yet published; however, their thecal morphology unambiguously shows them to be related to the other gonyaulacaleans, *Crypthecodinium* and *Alexandrium*. Dinoflagellates with peridinin-containing plastids are marked P; those that have lost them are shown by an asterisk; several replaced them by differently pigmented plastids from other eukaryotic algae (replacement plastid sources are: D = diatoms; H = haptophytes; G = green algae; C = cryptomonads). Dinoflagellates known to have single gene circles are underlined. Chloroplasts with both peridinin and minicircles were almost certainly present in the ancestral peridinean (position 3), but the symbiogenesis that implanted a red algal plastid into their ancestors may have taken place earlier (at positions 2 or 1). If, as we argue, plastids of red algal origin were probably already present in the common ancestor of dinoflagellates and Sporozoa (position 2), they must also have been lost independently in two further lineages: i.e., in the common ancestor (#) of the nonperidinean dinoflagellate classes (Noctilucea and Syndinea) and within Sporozoa in the *Cryptosporidium* lineage (which lacks plastids; Zhu et al. 2000). If they were acquired even earlier (position 1) in a single symbiogenetic event in the common ancestor of alveolates and chromists, in which chlorophyll c_2 was postulated to have evolved once only (Cavalier-Smith 1999), then the ancestors of ciliates and of *Perkinsus* must also have lost plastids.

have substantially larger DNA molecules instead, such larger size must be a secondarily derived property (unless minicircles evolved independently in the *Amphidinium/Protoceratium* and *Heterocapsa* lineages, which is highly unlikely). The fragmentation of a typical large chloroplast genome with scores of genes and a single replication origin into 10 (or more) separate single gene minicircles, each bearing closely related noncoding sequences, would have been a complex and inherently improbable succession of events; therefore it is much less likely to have occurred even twice independently in different dinoflagellate lineages than several increases in molecular size, whether by expansion of minicircles by duplications, recombination between them, or by gene transfer to the nucleus.

A Common Photosynthetic Ancestor for Dinoflagellates and Sporozoa?

A second key question is whether the common ancestor of dinoflagellates and Sporozoa was photosynthetic (Palmer 1992; Cavalier-Smith 1999) or nonphotosynthetic (Köhler et al. 1997); in other words, did sporozoan and peridinean dinoflagellate plastids have a common or an independent secondary origin? Unfortunately, dinoflagellate branches are extremely long in both 23S and 16S rRNA and *psbA* trees. Although dinoflagellates group with sporozoans (joined by euglenoids at the base) within the red algal/chl a/c group in all our 23S rRNA trees with strong bootstrap support, we cannot rule out the possibility that this grouping of dinoflagellates and sporozoans is a "long branch artifact" (Felsenstein 1978; Philippe and Laurent 1999). Because some mitochondrial branches are longer than those of sporozoan plastids (Fig. 2), but these mitochondria did not group with dinoflagellate plastids that have the longest branches, it is possible that the grouping of dinoflagellates with Sporozoa in both 23S and 16S rRNA trees is not purely artifactual but may reflect a genuine common origin for their plastids. Dinoflagellate chloroplast sequences do group with those long branching mitochondrial sequences in neighbor joining trees if rate variation among sites is not taken into account (not shown). This illustrates the importance of using phylogenetic methods incorporating such rate variation (Yang 1996). However, taking into account rate variation among sites (including variant sites) in puzzling analyses did not change the sister relationship of dinoflagellates and sporozoans.

Distantly related species with similar base compositions tend to cluster (Lockhart et al. 1994, 1996), but LogDet analysis that reputedly corrects for base composition bias did not change the grouping of dinoflagellates

and sporozoans. However, in all the rRNA trees the dinoflagellate and sporozoan branches are so long that we should be cautious when interpreting them, because none of the methods tried could move the euglenoids and *Prototheca* to the green group, and so did not eliminate every long branch problem.

In almost all the trees, the dinoflagellates and sporozoans are within the red algae/chromists group, while green plants and glaucophytes are most distant. The only tree where this is not true (a LogDet tree of 23S rRNA sequences, not shown) does not group them with green plants or glaucophytes either but at the base. These trees are consistent with the evidence from conserved gene clusters that sporozoan plastids originated from outside the green plastid lineage (Blanchard and Hicks 1999), probably from a red alga as revealed by the similarity of ORF 470 (Williamson et al. 1994) and ribosomal protein clusters (Stoebe and Kowallik 1999) between the sporozoan *Plasmodium falciparum* and a red alga. The congruence between our trees and the gene-cluster data make unlikely a green algal origin of sporozoan plastids, as was inferred from phylogenetic analyses of the *tufA* gene (Köhler et al. 1997). The bootstrap support of the *tufA* topology is weaker than that for a red algal ancestry on our present trees and very much weaker than that for a red algal ancestry on the concatenated protein gene trees (Zhang et al. 1999). Because of the presence of an envelope of three membranes surrounding the dinoflagellate plastid, it was suggested that dinoflagellate chloroplasts originated from secondary endosymbioses (Gibbs 1981). Although the presence of three membranes could in principle have been explained in terms of a primary endosymbiosis in which the food vacuole membrane surrounding the double cyanobacterial envelope was retained to form a third membrane (Cavalier-Smith 1982), that hypothesis would predict dinoflagellate plastid sequences to branch lower on the trees than sequences from plastids with two membranes (glaucophyte, red algae, and green plants, i.e., the kingdom Plantae). With the exception of the LogDet tree (not shown), all our present trees, and the earlier concatenated tree (Zhang et al. 1999) rule out such an early divergence with reasonably high confidence; therefore, we conclude that peridinin-containing dinoflagellate plastids most probably evolved by secondary symbiogenesis, not directly from a cyanobacterium.

Our phylogenetic analysis of numerous *psbA* genes clearly supports the idea that dinoflagellate chloroplasts probably originated from a red alga by secondary endosymbiosis and are therefore related to chromistan chloroplasts (Zhang et al. 1999; Takishita and Uchida 1999). The evidence from the rRNA genes is also broadly consistent with this, but is rather more ambiguous as long branch lineages properly belonging in the chlorophyll a/b clade group with the alveolates. However, the fact that when these long branch taxa are excluded, dinoflagellates and Sporozoa still branch with the red algal/chl a/c clade favors a red algal rather than a green algal ancestry for both dinoflagellate and sporozoan plastids. The fact that dinoflagellates and Sporozoa were at the same position on both the ribosomal RNA trees, irrespective of whether only one or both groups were included in the tree, indicates that their overall position is not simply a consequence of their two long branches attracting each other.

A direct common ancestry for dinoflagellate and sporozoan plastids (Palmer 1992; Cavalier-Smith 1999) therefore remains the most parsimonious working hypothesis because of their common relationship to red algae and similarity of protein-targeting mechanisms (Cavalier-Smith 1999), but it requires more rigorous testing by sequencing genes for nuclear-encoded plastid proteins. If these do not suffer from the extremely elevated evolutionary rates of the minicircle genes, they may be able more decisively to answer the question.

Ultra-Rapid Evolution of Alveolate Organelle Genomes

The question arises: Why do dinoflagellate and sporozoan chloroplast genes evolve so exceptionally rapidly? One possibility is that this is an indirect consequence of their apparently much smaller gene content than other chloroplasts (Wilson et al. 1996; Zhang et al. 1999). A smaller plastid gene content might be expected to reduce the strength of stabilizing selection for the retention of accurate replication and efficient DNA repair of the residual plastid DNA, because fewer different proteins would be affected by slight random drift away from an optimal sequence. A comparison with exceptionally high evolutionary rates of mitochondrial DNA is interesting in this connection: in trees for the mitochondrial gene cytochrome oxidase 1 (*cox1*) the branches for the alveolates (dinoflagellates, sporozoans, and ciliates) are also exceedingly long, with those of ciliates by far the longest (Inagaki et al. 1998). This interesting similarity to our chloroplast trees raises the possibility that there may be shared features of the DNA replication or repair machinery of both mitochondria and chloroplasts (neither organelle encodes its own machinery and all is nuclear encoded) and that one or more common elements were lost or modified in such a way as to drastically increase the net replication error rate in both genomes in the ancestral alveolate.

Gene reduction is also a prominent feature of such alveolate mitochondria. The sporozoan mitochondrial genome is exceptionally reduced in gene content, having just three polypeptides (Feagin 1992) and the possibility exists that this is also true for their dinoflagellate sisters in which no other protein genes have yet been found (Norman and Gray 1997; personal communication).

However, as the ciliate mitochondrial genome is not exceptionally reduced in gene content (Pritchard et al. 1990), such reduction cannot be an obligatory prerequisite for the origin of ultra-rapid evolutionary rates in organelle genomes. Possibly therefore the bias toward ultra-rapid organelle mutation arose purely accidentally in the ancestral alveolate.

There is however a frequent, albeit nonabsolute correlation between reduction in gene numbers, unusual genome organization, and unusually rapid nucleotide and amino acid substitution in both organelle genomes. Thus, the second longest branches on the *cox1* tree are the Euglenozoa, which have very unusual mitochondrial DNA that is well characterized only in kinetoplastid flagellates with their remarkable kinetoplast DNA and RNA pan-editing (Maslov et al. 1994). As mentioned above, on the mitochondrial *cox1* tree (Inagaki et al. 1998) their very long branch causes Euglenozoa to be placed artifactually within the alveolates and group as sisters to the Miozoa (dinoflagellates and Sporozoa) just as in most of our chloroplast rRNA trees (Figs. 2, 3–5). In further support of a connection between abnormal genomes and long branches, the third longest branch on the *cox1* tree and on three-gene and seven-gene trees (Turmel et al. 1999b) is a chlorophycean subgroup of green algae (incorrectly separated from other green plants on the tree because of its unusual branch length; Inagaki et al. 1998) that includes *Chlamydomonas*, which has a dwarfed and linearized mitochondrial genome with fragmented and jumbled genes (Boer and Gray 1988). The frequency with which such radically derived organellar genes are found in lineages with ultra-rapid substitution rates (e.g., Turmel et al. 1999b) suggests that radical genome rearrangements and ultra-rapid substitution may have a common underlying evolutionary cause: possibly an elevated rate of mutation caused by sloppier DNA replication, recombination, or repair. Any such relationship must be rather loose and stochastic rather than deterministic because the genomic novelties (e.g., minicircles in dinoflagellates alone among alveolates and kinetoplast DNA with maxicircles and minicircles in kinetoplastids alone among Euglenozoa) are exceedingly rare (unique) evolutionary accidents, probably of no adaptive significance (Cavalier-Smith 1993b). It is probably not realistic to think that mutation pressure plays the dominant role in the evolution of coding sequences (Kimura 1983; Ayala 1999), though this is probably essentially true for noncoding DNA sequences. The general inequality of branch lengths and erratic distribution of exceptionally long branches on molecular trees for different genes suggests instead that there is a frequently changing balance between mutation pressure and stabilizing (or, more rarely, directional) selection that can be much more dramatically shifted by unpredictable evolutionary accidents affecting the DNA manipulating machinery. Although the particular mutational changes that affect substitution rates might not be the same as those involved in genomic rearrangements, both would have more chance of fixation by random drift if stabilizing selection became weaker. Therefore it may not be entirely coincidental that dinoflagellates have a more rapid substitution rate in their chloroplast genes as well as a more aberrant genome organization than do Sporozoa. An early substantial reduction in dinoflagellate chloroplast genome size by extensive gene transfer to the nucleus and/or loss of genes, as their unusually restricted gene composition (Zhang et al. 1999) favors, could significantly have weakened stabilizing selection for accuracy of the chloroplast DNA manipulating machinery, and thereby led to the ultra-rapid substitution rates that we observe here for rRNA and *psbA*, and indeed for all their protein-coding genes yet studied (Zhang et al. 1999).

Acknowledgments. We thank K. Ishida for helpful discussion and J. Saldarriaga for *H. rotundata*, *T. heimii*, *P. reticulatum* total DNA. B.R.G. and T.C-S thank NSERC (Canada) for research grants and T.C-S thanks the Canadian Institute for Advanced Research for fellowship support. Note: The sequences reported here are in the GenBank sequence database under accession numbers AF 130040, 130041, 206701–206712.

References

Ayala FJ (1999) Molecular clock mirages. BioEssays 21:71–75

Blanchard JL, Hicks JS (1999) The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. J Eukaryot Microbiol 46:367–375

Boer PH, Gray MW (1988) Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. Cell 55:399–411

Cavalier-Smith T (1982) The origins of plastids. Biol J Linn Soc 17: 289–306

Cavalier-Smith T (1991) Cell diversification in heterotrophic flagellates. In: Patterson DJ, Larsen J (eds) The biology of free living heterotrophic flagellates. Oxford University Press, Oxford, pp 113–131

Cavalier-Smith T (1993a) Kingdom Protozoa and its 18 phyla. Microbiol Rev 57:953–994

Cavalier-Smith T (1993b) Evolution of the eukaryotic genome. In: Broda P, Oliver SG, Sims P (eds) The eukaryotic genome. Cambridge University Press, Cambridge, pp 333–385

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 Eukaryot Microbiol 46:347–366

Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. Tr Pl Sci 5:174–182

Chesnick JM, Kooistra WH, Wellbrock U, Medlin LK (1997) Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). J Euk Microbiol 44:314–320

Delwiche CF (1999) Tracing the thread of plastid diversity through the tapestry of life. Am Nat 154:S164–S177

Dodge JD (1975) A survey of chloroplast ultrastructure in Dinophyceae. Phycologia 14:253–263

Douglas SE, Murphy CA, Spencer DF, Gray MW (1991) Cryptomonad algae are evolutionary chimeras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148–151

Feagin JE (1992) The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. Mol Biochem Parasitol 52:145–148

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

Felsenstein J (1993) Phylip (phylogeny inference package) version 3.57c. Distributed by the author, Department of Genetics, University of Washington, Seattle, WA

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

Gibbs SP (1981) The chloroplasts of some groups of algae may have evolved from endosymbiotic eukaryotic algae. Ann NY Acad Sci 361:193–207

Gunderson JH, Goss SH, Coats DW (1999) The phylogenetic position of *Amoebophrya* sp. infecting *Gymnodinium sanguineum.* J Eukaryot Microbiol 46:194–197

Hillis DM, Huelsenbeck JP, Cunningham CW (1994) Application and accuracy of molecular phylogenies. Science 264:671–677

Inagaki Y, Ehara M, Watanabe KI, Hayashi-Ishimaru Y, Ohama T (1998) Directionally evolving genetic code: the UGA codon from stop to tryptophan in mitochondria. J Mol Evol 47:378–384

Jeffrey SW et al. (1975) Chloroplast pigment patterns in dinoflagellates. J Phycol 111:374–384

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Köhler S, Delwiche CF, Denny PW, Tilney LG, Webster P, Wilson RJ, Palmer JD, Roos DS (1997) A plastid of probable green algal origin in apicomplexan parasites. Science 275:336–342

Lockhart PJ, Steel MA, Hendy MD, Penny D (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. Mol Biol Evol 11:605–612

Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, Penny D (1996) Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc Natl Acad Sci USA 93: 1930–1934

Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393:162–165

Maslov DA, Avila HA, Lake JA, Simpson L (1994) Evolution of RNA editing in kinetoplastid protozoa. Nature 368:694

McFadden GI, Waller RF (1997) Plastids in parasites of humans. BioEssays 19:1033–1040

McFadden GI, Gilson PR, Hofmann CJ, Adcock GJ, Maier UG (1994) Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. Proc Natl Acad Sci USA 91:3690–3694

Medlin LK, Cooper A, Hill C, Wrieden S, Wellbrock U (1995) Phylogenetic position of the Chromista plastids based on small subunit rRNA coding regions. Curr Genet 28:560–565

Norman JE, Gray MW (1997) The cytochrome oxidase subunit 1 gene (*cox1*) from the dinoflagellate, *Crypthecodinium cohnii.* FEBS Lett 413:333–338

Olsen GJ (1987) Earliest phylogenetic branching: comparing rRNA-based evolutionary trees inferred with various techniques. Cold Spring Harbor Symp Quant Biol 52:825–837

Olsen GJ, Matsuda H, Hagstrom R, Overbeek R (1994) fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comp App Biosci 10:41–48

Palmer JD (1985) Comparative organization of chloroplast genomes. Ann Rev Genet 19:325–354

Palmer JD (1992) Green ancestry of malarial parasites? Curr Biol 2: 318–320

Palmer JD, Delwiche CF (1998) The origin and evolution of plastids and their genomes. In: Soltis DE, Soltis PS, Doyle JJ (eds) Molecular systematics of plants II. Kluwer, Norwall, MA, pp 375–409

Philippe H, Laurent J (1999) How good are deep phylogenetic trees? Curr Opin Genet Dev 8:616–623

Pritchard AE, Seilhamer JJ, Mahalingam R, Sable CL, Venuti SE (1990) Nucleotide sequence of the mitochondrial genome of *Paramecium.* Nucleic Acids Res 18:173–180

Reece KS, Siddall ME, Burreson EM, Graves JE (1997) Phylogenetic analysis of *Perkinsus* based on actin gene sequences. J Parasitol 83:417–423

Reith M (1995) Molecular biology of rhodophyte and chromophyte plastids. Ann Rev Plant Physiol Plant Mol Biol 46:549–575

Saunders GW, Hill DRA, Sexton JP, Andersen RA (1997) Small-subunit ribosomal RNA sequences from selected dinoflagellates: testing classical evolutionary hypotheses with molecular systematic methods. Plant Syst Evol [Suppl] 11:237–259

Schnepf E, Elbrächter M (1988) Cryptophycean-like double membrane-bound chloroplast in the dinoflagellate, *Dinophysis* Ehrenb: evolutionary, phylogenetic and toxicological implications. Botanica Acta 101:196–203

Smith SW, Overbeek R, Woese CR, Gilbert W, Gillevet PM (1994) The genetic data environment an expandable GUI for multiple sequence analysis. Comp App Biosci 10:671–675

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

Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Sugiura M (1995) The chloroplast genome. Essays Biochem 30:49–57

Swofford DL (1999) Phylogenetic analysis using parsimony (and other methods) (PAUP* 4.0b) (test version). Sinauer, Sunderland (Massachusetts)

Takishita K, Uchida A (1999) Molecular cloning and nucleotide sequence analysis of *psbA* from the dinoflagellates: origin of the dinoflagellate plastid. Phycol Res 47:207–216

Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Delwiche C, Klaveness D, Rudi K, Jakobsen KS (2000) Dinoflagellates of tertiary origin—the haptophyte origin of 19′ hexanoyl-fucoxanthin containing chloroplasts. Mol Biol Evol (in press)

Turmel M, Otis C, Lemieux C (1999a) The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea:* insights into the architecture of ancestral chloroplast genomes. Proc Natl Acad Sci USA 96:10248–10253

Turmel M, Lemieux C, Burger G, Lang BF, Otis C, Plante I, Gray MW (1999b) The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor:* two radically different evolutionary patterns within green algae. Plant Cell 11:1717–1729

Van de Peer Y, de Wachter R (1994) TREECON for windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comp Appl Biosci 10:569–570

Van de Peer Y, Van de Auwera G, De Wachter R (1996) The evolution of stramenopiles and alveolates as derived by "substitution rate calibration" of small ribosomal subunit RNA. J Mol Evol 42:201–210

Watanabe MM, Suda S, Inouye I, Sawaguchi T, Chihara M (1990) *Lepidodinium viride* gen sp. nov. (Gymnodiniales, Dinophyta), a

green dinoflagellate with chlorophyll a- and b-containing endosymbiont. J Phycol 26:741–751

Williamson DH, Gardner MJ, Preiser P, More DJ, Rangachari K, Wilson RJM (1994) The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum:* new evidence supports a possible rhodophyte ancestry. Mol Gen Genet 243:249–252

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

Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. Tree 11:367–372

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

Zhu G, Marchewka MJ, Keithly JS (2000) *Cryptosporidium parvum* appears to lack a plastid genome. Microbiology 146:315–321