A review of the evolution of the diatoms – a total approach using molecules, morphology and geology

LINDA K. MEDLIN[1*], WIEBE C. H. F. KOOISTRA[1,3] & ANNA-MARIA M. SCHMID[2]

Abstract. The evolution of the diatoms has been compared using sequence data from several genes: 18S rRNA, 16S rRNA, large sub-unit of RUBISCO, and tufA. Results of each analysis supports the division of the diatoms into two clades, neither of which corresponds to the present day classification system of the diatoms. Both the araphid and centric diatoms are paraphyletic. Plotting onto the rRNA tree of certain cytoplasmic characters, such as the arrangement of the Golgi bodies, supports the division into the two clades, whereas other characters normally used to classify the diatoms at higher taxonomic levels, e.g. at the class or subclass level do not. We also plotted morphological data from the best-preserved and earliest fossil diatom deposit onto the rRNA tree and the presence or absence of a central structure in the valve and details of the linking structures between cells support the two clades recovered in the molecular analysis. Molecular clocks calculated from a linearized tree representing each gene suggest that the average date of the origin of the diatoms is unlikely to have occurred before 200 Ma ago and the pigmented heterokonts likely arose at, or about, the Permian – Triassic boundary.

Key words: diatoms, endosymbiosis, fossils, Golgi bodies, molecular clock, phylogeny, rRNA sequences.

[1] *Alfred-Wegener-Institute, Am Handelshafen 12, D-27570 Bremerhaven, Germany*
[2] *Institut für Pflanzenphysiologie, Universität Salzburg, A-5020 Salzburg, Austria*
[3] *Present Address: Department of Marine Biology, Biological Centre, P.O. Box 14, 9750 AA Haren (GN), The Netherlands*
* *corresponding author: lmedlin@awi-bremerhaven.de*

INTRODUCTION

Use of Molecular Data

The diatoms (Bacillariophyta) are one of the most successful groups of microalgae, occurring in both aquatic and terrestrial habitats with over 10,000 described species

A. Witkowski and J. Siemińska (eds), The origin and early evolution of the diatoms: fossil, molecular and biogeographical approaches.
W. Szafer Institute of Botany, Polish Academy of Sciences, Cracow, 2000

and potentially many more cryptic species (Mann 1999). They are distinctly well known for among other features, their architecturally complex siliceous cell walls (valves and girdle bands) and the unusual pattern of reduction in cell size of one of the daughter cells following mitosis (see review in Mann & Marchant 1989). The intricate designs of their cell wall have been the basis for their classification system since the last century. However, it is well known that comparison of morphometric characters can be misleading and can lead to groups being mistakenly grouped by convergent evolution, possibly because of adaptation to similar ecological niches. Cladistic approaches in diatom morphology suffer from two drawbacks. Before morphological structures can be used in a meaningful way, one has to clarify homology among them; distinctly different structures may have a common origin. In contrast, common morphology does not need to be shared exclusively by common descent. Not surprisingly, several recent molecular phylogenetic reappraisals of taxonomies in other algal groups have demonstrated numerous cases of parallelism and convergence. This is particularly true for morphologically simple organisms or for those whose elaborate morphology is a function of just a few parameters.

Small-subunit ribosomal RNA (SSU rRNA) sequence comparison has proven to be a powerful alternative for resolving phylogenetic relationships at all taxonomic levels because it is a functionally stable evolutionary marker that evolved independently of morphology (Woese 1987, Bhattacharya *et al.* 1992). The rRNA genes are functionally conserved in all cells with no documented evidence of lateral gene transfer (Woese 1987). SSU rRNA sequence comparisons (Bhattacharya *et al.* 1992, Medlin *et al.* 1993) support a monophyletic origin of the diatoms within the pigmented protistan heterokont lineage (all organisms bearing tubular mastigonemes or flagellar hairs on their leading flagellum). Recent rRNA analyses have shown a new algal class, the Bolidophyceae, i.e., picoplanktonic algae with a reduced cellular organization, to be the sister group to the diatoms (Guillou *et al.* 1999).

Origin of the Diatom Plastid

As mentioned above, the diatoms belong to the heterokont algae, i.e., chlorophyll *a*+*c* algae with two heterodynamic flagella, of which the apically inserted flagellum bears tripartite mastigonemes and the laterally inserted flagellum is smooth. Although the flagellated stages of the diatoms are restricted to the spermatozoids of the oogamous genera (Manton & von Stosch 1966) and they lack the smooth posterior flagellum and the second basal body, they still belong to the heterokonts because they share pigment composition and plastid ultrastructure. Because both heterokonts and the haptophytes possess chlorophyll *a*+*c* and have complex plastids, e.g., plastids that lie within the so-called chloroplast endoplasmic reticulum, which is a special localized differentiation of the endoplasmic reticulum that surrounds the plastid and as the nuclear envelope the nucleus, it has been assumed that these groups of algae are also closely related. The questions of the monophyly of the host cells and the number of endosymbioses giving rise to their plastids confound relationships between these two groups. Medlin *et al.* (1997) formu-

lated 4 scenarios/hypotheses to explain the evolution of the heterokonts and the haptophytes host cells and their plastids. Briefly, these are:

Hypothesis 1a: The heterokonts and haptophytes form a monophyletic group that gained their plastids from a single secondary endosymbiotic event. That is, the host cells and the plastids for the two groups will have similar phylogenies because they share the same evolutionary history. Thus in molecular analyses, both the two host cells and the two plastids should be each other's sister group, respectively.

Hypothesis 1b: The heterokonts and the haptophytes are a monophyletic group, but after their divergence, each acquired its plastid through independent endosymbioses. Thus, in the molecular analyses, the host cells are each other's sister group but their plastids are not.

Hypothesis 1c: The heterokonts and the haptophytes are not a monophyletic group. However, they both engulfed and retained a similar eukaryotic cell as their plastid. Thus in the molecular analyses, the host cells are not each other's sister group but their plastids are.

Hypothesis 1d: The heterokonts and the haptophytes are not a monophyletic group. Each gained their plastids from separate secondary endosymbiotic events. Therefore, neither the host cells nor the plastids for the two groups will have similar phylogenies because each has had an independent evolutionary history.

Hypotheses 1a, 1c, & 1d can accommodate the possibility of plastid gain and subsequent loss in the heterokont lineage.

The endosymbiotic hypothesis of plastid evolution maintains that chloroplasts were acquired by primitive eukaryotic heterotrophs through the engulfing and maintenance of photosynthetic prokaryotes (Lauterborn 1895, Mereschkowsky 1905, Margulis 1981). The plastids of the rhodophyte, chlorophyte and glaucocystophyte algae and the higher plants have only two membraned-plastids and are assumed to have resulted from a primary endosymbiotic event in which a eukaryotic host engulfed a prokaryotic cell. The host organisms associated with the primary endosymbiosis appear to have arisen as independent plastid-bearing lineages within the crown group radiation of the eukaryotes (Bhattacharya & Medlin 1995, Fig. 1).

The algae with 3–4 membraned plastids are thought to have arisen through a secondary endosymbiotic event(s) in which a heterotrophic eukaryote host engulfed a photosynthetic eukaryote cell and converted it into a plastid. The additional membranes around the plastid are remnants of the endosymbiotic event (i.e., the host cell's endocytic vacuole and the plasmalemma of the endosymbiont, see review in Gibbs (1993)). Algae resulting from the secondary endosymbiosis are the euglenophytes and chlororachniophytes, which contain chlorophyll *a+b*, as well as the heterokont chromophytes, haptophytes, dinoflagellates and cryptophytes, predominately with chlorophyll *a+c* (Gibbs 1978, 1981, Cavalier-Smith 1982, Jeffrey 1989, Rowan 1989, Kowallik 1993, Sitte 1993, Valentin *et al.* 1993).

Whereas current evidence from molecular and morphological/ biochemical data suggest that the primary endosymbiotic event occurred only once, the secondary endosymbiotic event may have occurred several times (see review in Bhattacharya & Medlin

1995, Melkonian 1996, and Delwiche & Palmer 1996). The host organisms associated with the secondary endosymbioses, (viz., the euglenoids, cryptomonads, chlororachniophytes, dinoflagellates and heterokont chromophytes, with 3–4 membraned plastids) do not share a common ancestry, and thus a more likely hypothesis for their emergence as pigmented lineages is that each lineage has acquired its plastids through an independent secondary endosymbiosis, i.e. multiple secondary endosymbioses rather than through a single endosymbiotic event (see Cavalier-Smith 1982). The recognition that the vestigial nucleus (nucleomorph) in the plastids of the cryptophytes and chlororachniophytes is associated with the red algae and green algae, respectively, provides direct evidence using the nuclear genome of the endosymbiont that multiple secondary endosymbioses have occurred (Gibbs 1981, McFadden *et al.* 1997).

Many of the host lineages believed to have arisen from secondary endosymbiosis event(s) also have heterotrophic taxa as sister groups or as early divergences in their lineages (viz., heterokonts, cryptomonads, euglenoids, alveolates, chlororachniophytes, see Fig. 1 in Medlin *et al.* (1997) and in Bhattacharya & Medlin (1995)). Either these lineages were originally photosynthetic and these heterotrophic taxa lost their plastids, or the lineages gained their plastids later in their evolution through independent, secondary endosymbioses. The phylogenies of the plastids arising from secondary endosymbioses must be examined and compared with those of their host cells to infer the likely source of the photosynthetic eukaryote that was transformed into the plastids. Medlin *et al.* (1997) compared the host and plastid phylogenies from 1 nuclear and 3 plastid genes between the two groups.

Molecular Support for Lower Taxonomic Levels

Thus, molecular data has now identified the true sister group of the diatoms as the bi-flagellated Bolidophyceae (Guillou *et al.* 1999), and that they and the diatoms are correctly placed as the first photosynthetic lineage in the heterokont algae (Medlin *et al.* 1997). The branching order within the Heterokonta and the true sister group of the diatoms has been a highly controversial point (Medlin *et al.* 1995, 1997). Despite this recent knowledge, relationships at higher taxonomic levels in the diatoms are still virtually unexplored with modern phylogenetic and morphometric tools. This likely reflects the unique, highly derived morphologies of the diatom orders. Central to diatom taxonomy and phylogeny has been an assumption that the diatoms contain two groups, which can be distinguished from one another by their mode of sexual reproduction, pattern centers or symmetry, and plastid number and structure (Round *et al.* 1990). Familiar to most aquatic and cell biologists are the oogamous, radially symmetrical centric diatoms with numerous discoid plastids, which are distinct from the isogamous, bilaterally symmetrical pennate diatoms with fewer plate-like plastids. These characters have led most diatomists to separate historically the centric and pennate diatoms into two distinct classes. It is generally agreed that the pennate diatoms evolved from the centric forms because of their later appearance in the fossil record. In a recent systematic revision (Round *et al.* 1990), the raphid pennate diatoms (those with a slit opening [raphe] in the cell wall

for movement) and the araphid pennate diatoms (those without this slit) were given equal taxonomic ranking. Thus, the present classification system of the diatoms recognizes three classes: Coscinodiscophyceae (centric diatoms), Fragilariophyceae (araphid pennate diatoms), and Bacillariophyceae (raphid pennate diatoms). Classification of the diatoms at the order level and below is almost exclusively based on morphological features of the siliceous cell wall (Round *et al.* 1990). Key characters limiting centric diatom orders are valve structure and symmetry, and the type, number, and arrangement of tubular openings (labiate or strutted processes) through the valve. These features, plus the presence or absence of the raphe and its structure, e.g., presence or absence of fibulae, shape of the raphe endings, presence or absence of ribs, etc., figure prominently in the subdivision of raphid and araphid taxa. The raphe and the labiate or strutted processes through the cell wall are probably the most important features used to infer phylogenetic relationships among the diatoms. The labiate process normally has elaborate lip-like internal extensions, but some types of labiate processes are reduced, with only a simple slit-like internal opening. The slit scales in the auxospore wall of *Coscindiscus granii* are most likely the ontogenic predecessors of the labiate processes in this species (Schmid 1994b). The function of the labiate process is not well established, although in a restricted number of centric taxa, mucilage is secreted through the labiate process for movement (Medlin *et al.* 1986, Pickett-Heaps *et al.* 1986) or for some other unknown function (Crawford 1973). Unusually, in the absence of a labiate process and a raphe, *Ardissonia* moves by secreting mucilage through lateral grooves at the valve-girdle junction (Pickett-Heaps *et al.* 1991). The internal part of the labiate process can be used as a cytological anchor for the nucleus during interphase and new valve formation (Schmid 1987, 1994a, b). It has been hypothesized that the labiate process evolved into the raphe of the pennate diatoms (Hasle 1974), which has to be substantiated with the evolution of the macro-labiate process as an intermediate stage (see discussions in Schmid 1994a). The strutted process is restricted to the centric order Thalassiosirales and is a simple tube buttressed with struts and with adjacent pores. Chitan threads are secreted through the main pore for chain formation and flotation (Herth 1979, and references in Round *et al.* 1990). The cytoplasmic organization beneath the strutted process is more complicated than that beneath the labiate processes (compare figures in Herth 1979 to fig. 11 in Schmid 1994a).

Using rRNA sequence comparisons, Medlin *et al.* (1993) presented the first evidence that suggested centric and araphid pennate diatoms had a paraphyletic origin. One centric diatom, *Skeletonema costatum,* a member of the Thalassiosirales, was most closely related to the pennate diatoms, with high bootstrap support in phylogenetic analyses; the araphid taxa were a paraphyletic lineage within the pennates. An analysis of partial sequences from the 28S large-subunit (LSU) rRNA coding region from eight diatoms also showed that centric and araphid taxa were paraphyletic (Sörhannus *et al.* 1995). All subsequent analyses have supported these preliminary data and suggest that the currently used higher level diatom systematics do not reflect their evolutionary history (Medlin *et al.* 1995, 1996a, b, 1997). Thus the different reactions of dividing cells of *Coscinodiscus* (Clade I sensu Medlin *et al.* 1996b) and of members of the family Thalassiosiraceae and

of the raphid pennates (both of Clade 2 sensu Medlin *et al.* 1996b) support the uniqueness of the family Thalasiosiraceae (Schmid *et al.* 1981 p. 82–83, Schmid 1984b).

Calculations of a Molecular Clock

Because the diatoms have a good fossil record, the sequence data available from the extant diversity can be used to calibrate a molecular clock. Molecular data are normally used to reconstruct the phylogenetic history of extant organisms. Ideally, as organisms diverge, their genomes accumulate base substitutions in a stochastic, but clock-like manner. It is now widely recognized that a universal molecular clock does not exist and that the base substitution rate varies within lineages and genes. Nevertheless, if potential errors are identified with a relative rate test and corrected by eliminating the "fast" taxa and by linearising the rate of evolution, it is then possible to use molecular data to estimate divergence times.

This paper reviews the most recent evolutionary relationships among the diatoms as determined by the molecular data. Evidence from cytoplasmic detail and fossil material are used to find supporting evidence for the molecular clades. The most recent molecular clock interpretations for the origin of the diatoms and their plastids are also reviewed.

MATERIALS AND METHODS

DNA Methods. Nucleic acids were extracted as described in Medlin *et al.* (1988) or with a 3% CTAB (hexadecyltrimethylammonium bromide) procedure (Doyle & Doyle 1990) from cultures (Medlin *et al.* 1986) representing the three classes of diatoms and most major orders of centric taxa. Voucher slides are available for examination of those isolates no longer in culture. Small subunit rRNA coding regions were amplified using the polymerase chain reaction (PCR (Saiki *et al.* 1988)) and cloned as described in Medlin *et al.* (1988). Single-stranded templates were also produced for direct sequencing using biotin-labelled eukaryote-specific primers (Dynabeads M-280 Streptavidin, DYNAL A. S. Oslo, Norway). No fewer than 6 PCR reactions in each orientation were pooled for direct sequencing of the single-stranded templates. Both coding and non-coding strands were completely sequenced (Sanger *et al.* 1977, Elwood *et al.* 1985). Direct sequencing of the double stranded PCR products was also done using a LiCor sequencer with infra-red-labelled sequencing primers.

Sequence analysis. Previously published rRNA sequences from diatoms (Bhattacharya *et al.* 1992, Medlin *et al.* 1991, 1993, 1996a) and other chromophytes/oomycetes, dinoflagellates, and prymnesiophytes (Medlin *et al.* 1996a, Neefs *et al.* 1991, Andersen *et al.* 1993) were used to align the ssu rRNA sequences using maximum primary and secondary structural similarity. Positional homology was assumed for 1739 positions; 528 of these were informative and used in the maximum parsimony analysis. The final data set contained 89 taxa. We rooted the trees with one oomycete and three chromophytes because the use of multiple outgroups improves the analysis.

Maximum parsimony analyses were implemented with the PAUP program (Swofford 1993). Introduced gaps were treated as missing data, and informative characters were treated as multistate and unordered. Unweighted maximum parsimony trees were obtained using the tree-bisection-reconnection (TBR) branch swapping option and a heuristic search with random additions of the taxa. The most parsimonious trees (MPT) and the data matrix were loaded into the MacClade computer program (Maddison & Maddison 1992) to infer a weighted data set in which the frequency of nucleotide substitutions at each position was inversely related to the number of changes at that position (Scale 1–100). The type of substitution, i.e., purine –> purine, purine –> pyrimidine, (i, j), at each position was also weighted (K_{ij}) as the cost of going from one state to another: ($K_{ij} = - \ln (X_{ij}/X_i)$ where X_{ij} is the number of i–>j changes, X_i is the number of changes from i to any state, and X is the number of changes on the tree (Scale 1–100). These weightings greatly enhance the ability of the maximum parsimony analyses to recover the correct tree when multiple substitutions have occurred (Hillis & Moritz 1990). These constraints were used to generate weighted maximum parsimony trees. Stability of the branching order was estimated using bootstrap analysis for 100 replicates (Felsenstein 1985).

Distance analysis was performed using PHYLIP (Felsenstein 1993). Dissimilarity values (Fitch & Margoliash 1967), based on pairwise comparisons of sequences, were transformed into distances using the Kimura-two-parameter-model (Kimura 1980). Distance matrices were converted into trees using the neighbor-joining option in PHYLIP. Branching order stability was estimated by bootstrap analysis as above. A log-det transformation was used to calculate a distance matrix in which variable types of base substitution were permitted (Lockhart *et al.* 1994).

Molecular Clock Calculations: A relative rate or branch length test was calculated in which the evolutionary rate of all pair-wise combinations of taxa was compared to several outgroups (Wu & Li 1985, Takezaki *et al.* 1995) to identify taxa not evolving within a stochastic model of base substitution. A range of taxa with varying degrees of distance from one another was selected to construct a linearized neighbor-joining tree in which rate variation between the taxa is assumed to be eliminated (Takezaki *et al.* 1995). A linearized neighbor joining (NJ) tree was constructed from the nuclear-encoded SSU rRNA genes and from the plastid-encoded SSU rRNA, *rbc*L and the *tuf*A genes.

A regression of first appearance dates of diatom and coccolithophorid taxa from the fossil record against branch lengths of extant representatives of these taxa in all trees was performed. The average age of any undated node was determined by multiplying the length of its median or average branch by the regression coefficient. The earliest possible age for any undated node is the upper 95% confidence limit around the age estimate given the distance of its median or average branch (Hillis & Moritz, 1990). For the linearized trees, standard errors (SE's) associated with the Kimura-2-parameter distances between neighbor lineages were calculated according to Li and Bousquet (1992), using the next nearest neighbor's taxon with the median branch length as reference. The SE's associated with the height of internal nodes in the trees were obtained by dividing each SE by 2. The 95% upper and lower limits around the average estimate of the origin were

obtained by multiplying the SE/2 by the appropriate critical value of the Student's *t*-distribution corresponding to n-1 taxa in that lineage.

RESULTS AND DISCUSSION

Molecular data

18S small subunit rRNA sequences for all taxa are approximately 1800 nucleotides in length with the exception of *Cymatosira belgica* and *Aulacoseira ambigua*, which were longer with several inserted helices (see Medlin *et al.* 1996b). A phylogeny of the diatoms inferred from the maximum parsimony method is presented in Fig. 1. Bootstrap values greater than 60% supporting the recovered branches in either the distance or the weighted maximum parsimony trees are placed at the major internal nodes of the tree. In all analyses, the diatoms diverge initially into two clades (Fig. 1). This division does not correspond to centrics or pennates (see discussion in Simonsen 1972, 1979) or to the three classes of diatoms in Round *et al.* (1990). Beyond this initial separation, the relationships inferred from our analysis correspond best to ordinal level in current diatom systematics (Fig. 1).

Members of the Paraliales, Coscinodiscales, Rhizosoleniales, Corethrales, Leptocylindrales, Melosirales, and Aulacoseirales were found in the first clade recovered in each of the two analyses (Fig. 1). In general these taxa have labiate processes located **peripherally** around the cell wall (Fig. 3). Exceptions to this do occur, e.g., (1) in large cells of *Coscinodiscus wailesii* there are at least 50 to 100 labiate processes on the valve face in addition to the marginal ones and (2) in *Azpeitia* there is a single central labiate process, but we believe the labiate process moved to the valve center of this genus after the evolution of the lineage because *Azpeitia* first appears in the Eocene, although other members of the Stellarimaceae and the closely related Benetoraceae occur in the lower Cretaceous (Nikolaev & Harwood 1994, 1997, Sims 1994). Clearly, extant members of the Stellarimaceae should be targeted for rRNA analysis to investigate this hypothesis. Also, in many *Rhizosolenia* species, the labiate process appears to be centrally located.

Within this first clade are several lineages (Fig. 1). *Paralia* roots the entire Clade 1. Following that, there is a divergence into two lineages. Taxa with large, elaborate labiate processes (Hasle & Sims 1990) belong to both lineages (orders Coscinodiscales and Rhizosoleniales). In the first lineage are the orders Corethrales and Rhizosoleniales. Fryxell and Hasle (1977) suggested a close relationship between *Rhizosolenia* and *Corethron* because of similar girdle band architecture. Round *et al.* (1990) erected the order Corethrales because *Corethron* lacks any kind of processes, but this separation may not be justified. The Leptocylindrales and *Proboscia* root the second lineage, notably the latter is not affiliated with the Rhizosoleniales, a feature that is supported by their morphology and certain details of the cytoplasm during the cell cycle (Pickett-Heaps, pers. comm.). Some taxa in the second lineage (order Melosirales) have only small labiate processes (Round *et al.* 1990).

In all analyses, Clade 2 comprises several groups. *Lampriscus* and *Biddulphiopsis* root Clade 2 diatoms, but there is stronger bootstrap support if these two taxa are excluded from Clade 2(66). Following that there are divergences into several groups. The order Thalassiosirales, with (usually) one large **marginal** labiate process but with one to several **central** strutted processes, together with the Triceratales and Lithodesmiales, form one group. The bi-(multi) polar centric diatoms of the orders Cymatosirales, and Chaetocerotales are another group, which is sister to the pennate diatoms. These diatoms have a small, but **centrally** located labiate process. The pennate diatoms comprise the third lineage and have a **central** rib or raphe. The labiate processes of the araphid taxa are usually small and commonly located at the apices. Low bootstrap values indicate weak support for the branching order within this major second diatom lineage, which may be evidence for a rapid diversification within this clade.

Among the pennate diatoms the araphid taxa are paraphyletic. The freshwater and marine *Fragilaria* spp. do not form a monophyletic group and should be separated into distinct genera. The raphid diatoms are monophyletic. The naviculoid raphid diatoms have evolved into two separate lineages. The nitzschioid diatoms appear monophyletic; although fibulae have arisen twice, e.g., in the Bacillariaceae and in *Entomoneis* and in the Surirellales. The relationship recovered in the rRNA tree suggest that the unique raphe system of the surirelloid diatoms is likely to have its origins in the amphoroid diatoms.

Because the two major clades do not correspond to any higher taxonomic level of the diatoms, Medlin *et al.* (1996a) performed additional analyses to test for biases in the data set, which could give us an erroneous tree. User-defined trees afford the possibility of testing alternative evolutionary hypotheses. Medlin *et al.* (1996a) constructed three alternative topologies. In that analysis, to test for the monophyly of the centrics, Medlin *et al.* (1996a) exchanged the pennates with clade 1 radial centrics. To test for number of groups of centric diatoms, Medlin *et al.* (1996a) maintained the first tree rearrangement and then grouped the centrics as follows: the radial centrics, the bipolar centrics and the Thalassiosirales. To test for the uniqueness of the Thalassiosirales, Medlin *et al.* (1996a) kept the original tree topology found in the maximum likelihood analysis but removed the Lithodesmiales to a separate branch so that the pennates and the Thalassiosirales were sister clades. These trees were used to constrain new maximum likelihood and maximum parsimony analyses. Statistics associated with each analysis were compared to those obtained in the original analyses, which were assumed to represent the best tree. All three alternative topologies required more steps for the most parsimonious tree than the original analysis, and the log-likelihoods from the constrained maximum likelihood analyses were significantly worse than that found for the best tree in the original maximum likelihood analysis (Table 1). These results indicate that a dichotomy into centric and pennate taxa is unlikely. Significantly, the removal of Lithodesmiales from the base of the Thalassiosiralean lineage yields a less robust tree.

Our 18S rRNA analysis reveals a dichotomy in the diatoms that is difficult to explain if morphological characters that currently define diatom classes, such as pattern centers, reproduction and plastid morphology are used to characterize the molecular tree. Further-

Table 1. Summary of tree statistics obtained when three-user defined trees were used to constrain maximum likelihood and maximum parsimony analyses using the Kishino & Hasegawa (1989) model. Taken from Medlin *et al.* 1996.

Trees	Log-Likelihood	Difference in Log-Likelihood	Standard Deviation	Significantly Worse	Number of steps in MPT
1. Best tree	14443.799	–	–	–	2258
2. Monophyly of the Centrics	14479.662	-35.862	17.877	Yes	2270
3. Monophyly plus 3 centric clades	14482.788	-38.989	18.980	Yes	2269
4. Thalassiosirales and Pennates as sister clades	14469.000	-25.201	10.723	Yes	2268

more, valve structures and tubular openings, such as labiate and strutted processes used to characterize many diatom orders, also fail to describe the two clades recovered in our molecular tree. However, these structures can describe younger branches in the molecular tree at the order level in diatom taxonomy.

Cytoplasmic Support for Molecular Data

We next turned to internal structures of the cytoplasm to explain the phylogeny recovered in our tree. Among the diatoms there appear to be at least three, possibly four types of arrangement of the Golgi apparatus (Fig. 2). In the first arrangement, the Golgi body is associated with a cisterna of endoplasmic reticulum and a mitochondrion in the genera *Coscinodiscus* (Schmid 1984a, 1988), *Stephanopyxis* (Wurzinger & Schmid, unpubl.) and *Ellerbeckia* (similar to *Paralia*) (Schmid & Crawford 2000). Near the nucleus, the mitochondrion lies closer to this organelle, whereas in the cortical cytoplasm, the mitochondrion faces the plasmalemma. This association is termed the G-ER-M unit and is also found in the oomycetes and the red algae (Schmid 1988), and we refer to it as Type 1 variation (Fig. 2a). Type 2 Golgi variation can be found in most pennates and in the Thalassiosirales where the Golgi bodies encircle the nucleus to form a peri-nuclear shell or ring in profile (Schmid 1989, Fig. 2b). Also in *Synedra ulna*, the nuclear envelope, together with nuclear matrix and RNA , but no DNA, is pulled out into 2 long tentacles reaching the cell poles (Schmid 1989). Golgi bodies are aligned along one side of the tubes, i.e., the Plattenband *sensu* Geitler (Schmid 1989), whereas the other side of the tentacle bears nuclear pores. Thus we consider this type to be a variation of type 2 (Fig. 2, variation 2.1), which became necessary to cope with a space problem in a diatom 8 by 12 µm wide and 500 µm long. Another variation of type 2 (Variation 2.2 in Fig. 2)

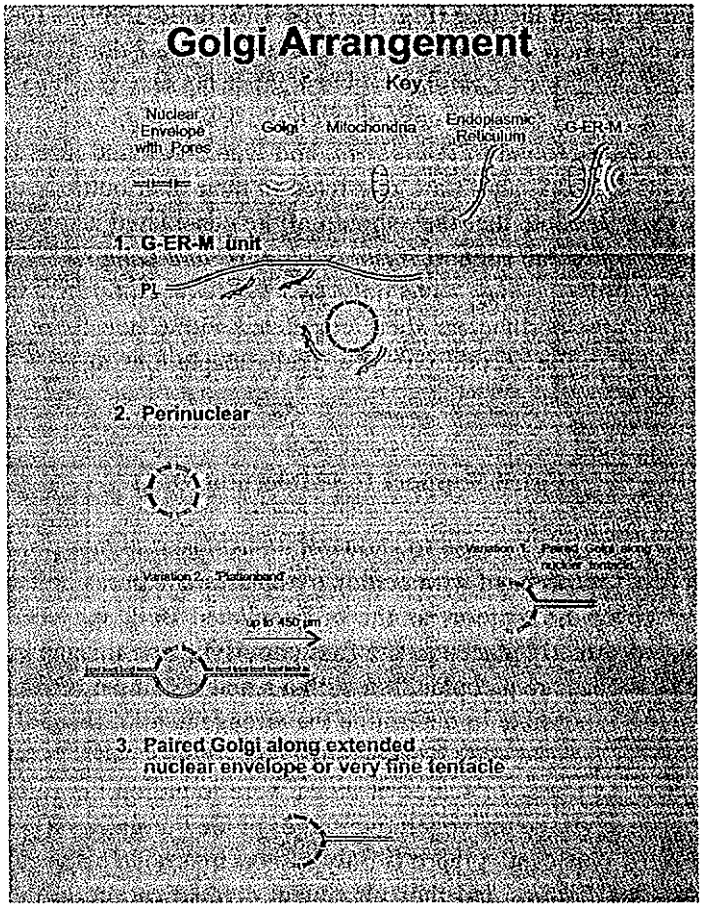


Fig. 2. A schematic drawing of the three types of Golgi body arrangements found so far in the diatoms.

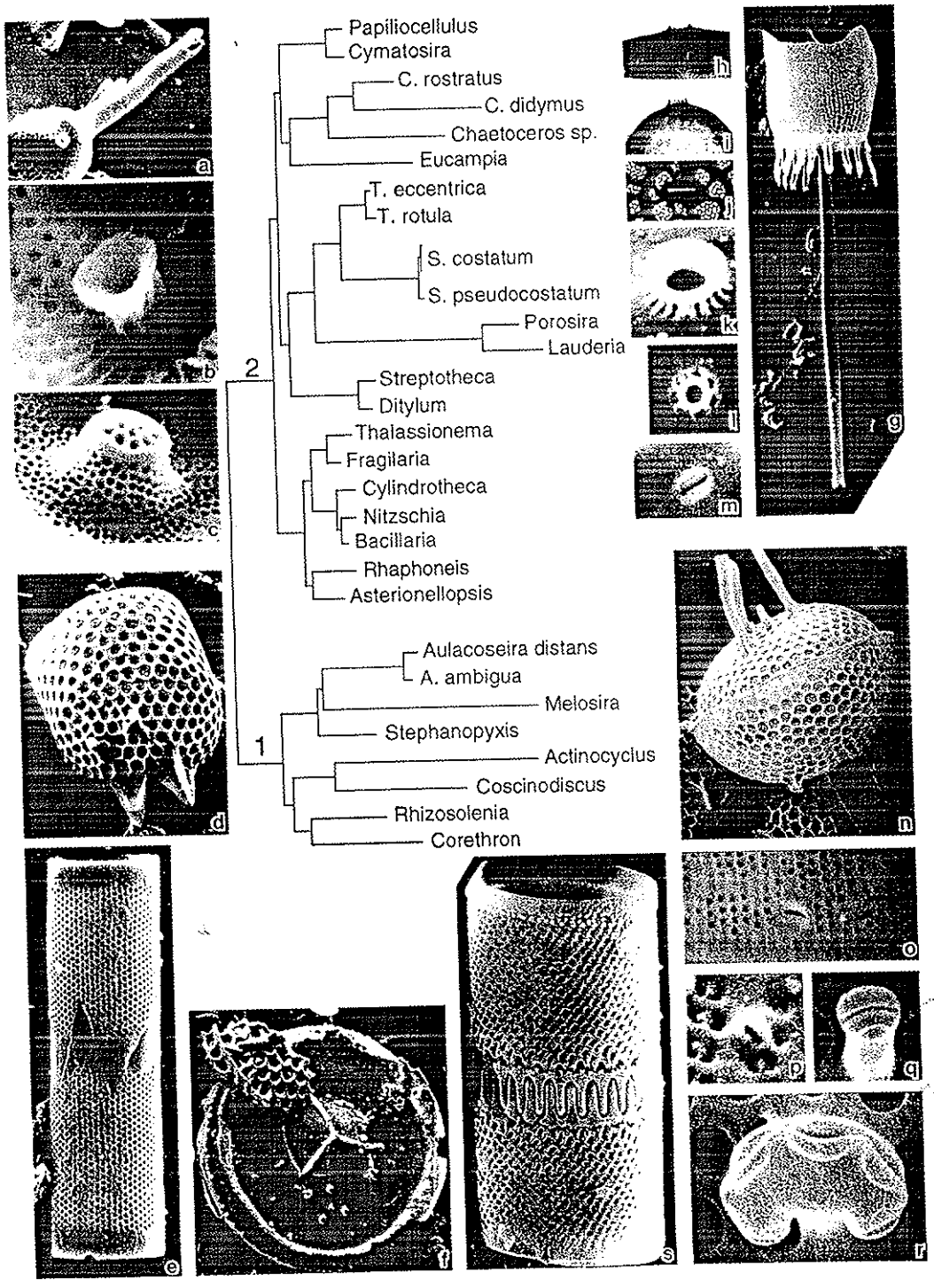
can be found in *Pinnularia*, where nuclear tentacles are "pulled out" not only in 2, but in multipolar directions, because of more available space around the nucleus as compared to *Synedra ulna*. In this case the tentacles are much narrower already at their base, leaving enough space on the nuclear envelope for pores, and thin out distally into ER-cisternae. Golgi-bodies are paired along these nuclear-protrusions ("Doppelplättchen"), i.e., on both sides of the tentacle. Finally, these arrangements are functionally topped by the Golgi-type 3 variation found in *Biddulphiopsis titiana*, where only the outer membrane of the nuclear envelope forms filose tentacles along which Golgi-bodies are paired, whereas the remaining envelope around the large nucleus bears a porous pattern of the same density as seen in *Coscinodiscus*. Although, the GERM-unit of *Coscinodiscus* can be found also along ER cisterna in the cortical cytoplasm, the Type 3 Golgi-arrangement has been found only along radial ER-cisternae (Schmid unpubl.). If these four variations are plotted onto the rRNA tree, then is quickly becomes apparent that they are well corre-

lated with certain clades in the tree (Fig. 1). Type 1 (Schmid 1989) is found primarily in Clade 1 with one exception *Odontella sinensis* whereas type 2 dominates in Clade 2. Interestingly, other members of *Odontella* investigated possess the type 2 arrangement, (Pickett-Heaps *et al.* 1986, Schmid unpubl., Schmid 1989, Fig. 2). Type 2.1 is found in *Synedra ulna,* type 2.2 is found in *Pinnularia*, and type 3 is found in *Biddulphiopsis*. As Golgi arrangements 2.1, 2.2, and 3 are restricted in their distribution and are embedded in Clade 2, it is then possible that they should all be considered a variant of type 2, but clade 2 has stronger bootstrap support if *Biddulphiopsis and Lampriscus* are not included in this clade. In reality we may have only two types of Golgi arrangement: (1) the G-ER-M unit and (2) the perinuclear shell, correlated with Clade 1 and Clade 2, respectively. The variations in the second type are likely adaptations to cope with large cells and extra space.

Fossil Support for Molecular Data

Thus, molecular data can be used to show relationships at higher taxonomic levels in the diatoms and is best correlated with certain cytoplasmic features. It has also identified the true sister group of the diatoms as the bi-flagellated Bolidophyceae, and that they and the diatoms are correctly placed as the first photosynthetic lineage in the heterokont algae. The Bolidophyceae are picoplankton with a reduced cellular organization (Guillou *et al.* 1999). It was not possible to determine the arrangement of the Golgi body and the single mitochondrion in these cells because the cells are so small and all the organelles are closely aligned. No traces of scales or silica were found on the cells. However, the possibility that these might be the vegetative stages of the Parmales was noted (see Mann & Marchant 1989). Significantly, there is only a short internode between the divergence of Bolidophyceae and diatoms and the subsequent diversification of the diatoms themselves. Therefore, we believe that the earliest fossil diatoms are (relatives of) progenitors of extant clades and that the origin of the extant diversity of the diatoms must predate

Fig. 3. Phylogeny inferred with the neighbor-joining method using a LogDet matrix. Clades 1 and 2 are labeled. Scanning electron micrographs (SEM) 2a-f are of ODP Site 693 fossil taxa. SEM and transmission electron micrographs 2g-s are of modern taxa. Figs. 2a-c. Group 2 fossil taxa (34), suggested as ancestors of Clade 2 modern taxa. a) *Rhynchopyxis*, ×1500. b) central structure of *Gladiopsis*, ×11,000. c) central structure of *Praethalassiosiropsis*, ×4,800. Figs. 2d-f Groups 1 and 4 and Group 3 fossil taxa (34), the former suggested as ancestors of Clade 1 modern taxa. d) *Amblypyrgus*, Group 1, ×1,900. e) *Archepyrgus*, Group 4, ×2,200. f) *Trochus*, slits arrowed, Group 3, ×1,600. Figs. 2g-m. Clade 2 taxa. g) *Ditylum*, ×2,000. h) side view of central labiate process of *Arcocellulus*, ×8,000. i) side view of central labiate process of *Chaetoceros*, ×9,400. j) central labiate process of *Eucampia*, ×6,300. k) central strutted process of *Thalassiosira wittiana,* ×20,000. l) central strutted process of *Thalassiosira fraga,* ×12,000. m) apical labiate process in *Tabularia,* ×32,000. Figs 2n-s. Clade 1 taxa. n) *Stephanopyxis*, ×1,300. o) peripheral labiate process of *Stephanopyxis*, ×8,800. p) peripheral labiate process of *Aulacoseira*, ×13,200. q) peripheral small labiate process of *Coscinodiscus,* ×19,300. r) peripheral large labiate process of *Coscinodiscus,* ×16,600. s) *Aulacoseira*, ×4,200. Dr. R. Gersonde kindly provided SEM micrographs of the fossil taxa.


Papiliocellulus
Cymatosira
C. rostratus
C. didymus
Chaetoceros sp.
Eucampia
T. eccentrica
T. rotula
S. costatum
S. pseudocostatum
Porosira
Lauderia
Streptotheca
Ditylum
Thalassionema
Fragilaria
Cylindrotheca
Nitzschia
Bacillaria
Rhaphoneis
Asterionellopsis
Aulacoseira distans
A. ambigua
Melosira
Stephanopyxis
Actinocyclus
Coscinodiscus
Rhizosolenia
Corethron
2
1
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s

their first appearance in the fossil record. The short internode leaves too little time for the existence of a more ancient diversification and subsequent extinction.

If we then use the morphological features present in the earliest exceptionally well-preserved Lower Cretaceous diatom fossil material from Ocean Drilling Program (ODP) Leg 113, Site 693, Antarctica deposit to characterize the molecular tree, then we can also recover the clades in our molecular analysis (Medlin *et al.* 1995) (Fig. 3). Diatoms in this deposit were placed into four groups by their valve structure and type of linking apparatus (Gersonde & Harwood 1990). None of the diatoms in this fossil deposit possess labiate or strutted processes, as currently defined, but possess other types of processes, instead. If we use the presence or absence of a central invagination that develops into a central tube in the valve face of the diatoms of this deposit and details of the elaborate linking structures between cells to support our rRNA tree, then three of the four groups defined from this fossil diatom deposit correlate with the two clades recovered in our rRNA tree (Medlin *et al.* 1995, Fig. 3).

The absence of a central structure/tube and the presence of robust linking structures are characteristic of Group 1 and Group 4 taxa of Gersonde and Harwood (1990), and we have proposed that these are the features characteristic for the ancestral stock of Clade 1 of our molecular tree (Medlin *et al.* 1995, Fig. 3 d-f, n-s). Linking structures in *Stephanopyxis* are similar to those of Group 1 taxa of Gersonde and Harwood (1990); the structures holding the external tubes of the labiate processes are formed from part of the areolar wall (Fig. 3d, and see Fig. 6 in Haga 1997), whereas in *Aulacoseira*, the linking structures are formed from the main part of the wall like the Group 4 taxa of Gersonde and Harwood (1990, Fig. 3e, s). Significantly, modern taxa from Clade 1 **lack**, in general, any central structure or tube in the valve but do possess one of two types of peripherally located tubes, the labiate processes (Fig. 3p, q, r). These labiate processes must have evolved multiple times independently since the initial divergence of the two clades because no labiate processes or tubes can be found in the presumed ancestral stock for Clade 1 from the Lower Cretaceous deposit. This implies that not all labiate processes are homologous structures, and cladistic studies coding them as different states of the same structure may be erroneous.

The presence of a central tube-like structure and a reduction in linking mechanisms are characteristic of Group 2 taxa of Gersonde and Harwood (1990). We have suggested that this group forms the ancestral stock for Clade 2 of our rRNA tree (Medlin *et al.* 1995) (Fig. 3a-c, g-m). Morphological changes in the valve center of these Lower Cretaceous diatoms can be hypothesized to have occurred from a valve with an invagination (e.g. the uvular process of *Archaegladiopsis*, see Nikolaev & Harwood 1994, 1997) to a valve with a central tube whose internal opening is covered with a perforate plate (e.g. the perforate process of *Praethalassiosiropsis*) (Gersonde & Harwood 1990, Fig. 3c). The next step is for the perforate plate to be lost, then the annular process or multi-strutted processes of *Gladiopsis speciosa* (Schulz) Gers. & Harw. (Sims 1994) and *Thalassiosiropsis* (Hasle & Syvertsen 1985) are formed as the pores in the wall of the central tube coalesce. If the perforations in the wall of the central tube are further reduced, then the sipho-shaped process (*Gladiopsis ellipsoidea* Gers. & Harw.) is formed. Finally a

valve with an open central tube without perforations in the tube wall (e.g., the rhyncho-shaped process of *Rhynchopyxis* (Gersonde & Harwood 1990, Nikolaev & Harwood 1994, 1997, and R. Gersonde & D. Harwood, pers. comm., Fig. 3a, b) is produced. As originally defined, *Gladiopsis* contains taxa with at least three or more types of central valve structures/tubes: the uvular process, the annular or multi-strutted process, and the siphoshaped process (Nikolaev & Harwood 1994, 1997). Nikolaev and Harwood (1997) have proposed one new genus, *Archaegladiopsis* for those taxa formerly placed in *Gladiopsis* with an uvular process. The line of development from the perforate process of *Praethalassiosiropsis* into the multi-strutted (annular) process of *Thalassiosiropsis* (Harwood, pers. comm.) appears to stop at the Palaeocene (Sims 1994). Although true *Thalassiosira* spp. do not appear until the Eocene, a search for more direct ancestors spanning the gap between the last *Thalassiosiropsis* spp. existed and modern Thalassiosirales should perhaps be directed toward taxa with a central structure/tube more closely resembling that of *Ditylum* or *Rhynchopyxis*. Such evolutionary sequences are supported by the phylogenetic relationships inferred from the rRNA tree. The removal of *Ditylum* from the base of the Thalassiosirales in our rRNA tree yields a significantly less robust tree and also implies that the morphological changes inferred from the fossil material have merit. This suggests that the satellite pores of true Thalassiosirales may be a later addition to the central tube or strutted process in this lineage. In an analysis of partial large subunit (28S) rRNA sequences, Sörhannus *et al.* (1995) recovered the Thalassiosirales as outgroup to *Lithodesmium* and the pennates. Clearly, there is a close association between the Thalassiosirales and the Lithodesmiales. Significantly, modern taxa in Clade 2 have a central structure in the valve, namely a labiate or strutted process or a rib or raphe (Fig. 3g-m).

Finally, from our rRNA analysis and from the Golgi arrangement, it can be shown that *Coscinodiscus* is not related to *Thalassiosira*. Historically, many *Thalassiosira* species were identified as *Coscinodiscus*, and today, many taxonomist likely still confuse the two genera. Any scenario that reconstructs a recent evolutionary history between the orders to which these taxa belong would be in serious conflict with several lines of evidence. Not only do they belong to two entirely different clades of diatoms as shown by our rRNA tree, but they also have completely different valve structures and tube processes (Fryxell & Hasle 1972, Herth 1979) and have different responses to microtubule poisons (Schmid 1981, 1984b).

Molecular Clock Calculations

Using the method of Hillis and Moritz (1990) Medlin *et al.* (1997) have estimated from the nuclear and plastid genes: the time of origin (1) for the diatoms, (2) for the heterokont chromophytes, (3) for the haptophytes and the divergence of their two subclasses and (4) for the timing of the secondary endosymbiotic events for the pigmented heterokonts and the haptophytes. Plastid genes from haptophytes and heterokont chromophytes were never sister taxa in the analyses of Medlin *et al.* (1997), and the heterokont chromophyte plastid genes were always embedded within the red algal plastid lineage. Changes in gene order and composition also support this molecular evidence

(Kowallik *et al.* 1993, Martin *et al.* 1998). The convergence of the 16S rRNA and RUBISCO large subunit plastid gene phylogenies suggests that the heterokont chromophytes were likely to have engulfed a primitive red alga, whereas the cryptophytes and the haptophytes are more likely to have taken in an ancestor of the red algae. This suggests that the Chromista, which are the chlorophyll *a+c* algae, are not monophyletic and that haptophyte, pigmented heterokont and cryptophyte plastids arose from separate endosymbiotic events. These data support Hypotheses 1c and d. Medlin *et al.* (1997) initially calculated a relative rate or branch length test in which the evolutionary rate of all pair-wise combinations of taxa was compared to several outgroups (Wu & Li 1985, Takezaki *et al.* 1995). In this manner Medlin *et al.* (1997) identified taxa not evolving within a stochastic model of base substitution. Medlin *et al.* (1997) then selected a range of taxa with varying degrees of distance from one another to be used for the construction of a linearized neighbor-joining tree in which rate variation between the taxa was assumed to be eliminated (Takezaki *et al.* 1995). A linearized neighbor joining (NJ) tree was constructed from the nuclear-encoded SSU rRNA genes and from the plastid-encoded SSU rRNA, *rbc*L and the *tuf*A genes (Fig. 4). First appearance dates of diatom and coccolithophorid taxa with a fossil record were regressed against estimated branch lengths of lineages in each tree to construct a molecular clock for each gene or group of organisms (Fig. 5). First appearance dates of taxa immediately after a gap in the diatom fossil record were predated to the middle of the gap or before it, if potential ancestors of the extant taxa could be identified in well-preserved diatom deposits before the gap (Gersonde & Harwood 1990).

For each linearized tree, Kooistra and Medlin (1996) and Medlin *et al.* (1997) estimated an average age of the clade and its earliest probable age (p = 95%). The average age of any undated node was determined by multiplying the length of its median or average lineage by the regression coefficient. The earliest probable age for any undated node was taken from the upper 95% confidence limit around the time estimate for the undated node in the tree given the length of its median or average lineage.

Fossil dates may seriously underestimate the first appearance date of any lineage (Wray & al. 1996) because molecular divergence will nearly always precede morphological divergence and the first fossil record will always be younger than the first appearance. Thus, time estimates based on the average age for the lineage, given that the fossil dates may be later than first appearances, are also likely to underestimate the origin of groups. The molecular clock method of Hillis and Moritz estimates an upper (older) and a lower (younger) 95% confidence limit for the origin of any undated node with the high probability time range around the average value. Thus, using this calculation, the actual time of origin of any undated node in the tree, should realistically lie somewhere between the average age determined from the regression line and the earliest probable age determined from the upper 95% confidence limit. Dates between the lower 95% limit and the first appearance of fossil taxa are, of course, nullified because these fossils prove unambiguously the existence of that lineage at that time. However, with few fossil-dating points, the 95% confidence interval can be quite broad, pushing the earliest probable age farther back in time.

From the 18S rRNA clock calculated for the pigmented heterokonts, Medlin *et al.* (1997) have estimated the average age of the brown algae, the diatom and the pigmented heterokonts (see also Kooistra & Medlin 1996 and Fig. 5). The recent appearance of the brown algae is well in agreement with other molecular, morphological and biogeographic evidence (see Saunders & Druehl 1992), but contrasts with some earlier putative brown algal fossils (Taggart & Parker 1976, Xiao *et al.* 1998). The average age of the diatoms is very close to their first fossil record (Medlin *et al.* 1996a, Rothpletz 1896). Using our average and earliest probable dates for the origin of the pigmented heterokonts (170–270 Ma), Medlin *et al.* (1997) concluded that the pigmented heterokonts are unlikely to have existed much before the Permian-Triassic boundary. Clearly, the heterokont lineage is much older but the pigmented heterokonts are a later divergence within this lineage.

By constructing a molecular clock from our plastid gene sequence data, Medlin *et al.* (1997) attempted to date the timing of the endosymbiotic event leading to the transfer of photosynthetic capacity to the heterokont and haptophyte lineages. Thus, the ages derived from plastid genes were compared to those of the origins of their host cells. If Hypothesis 1a is true, then the dates for the divergence of the haptophyte and heterokont algae should be younger than the dates for the endosymbiotic event leading to the transfer of photosynthetic capacity to the haptophyte and heterokont lineages.

Hypotheses 1a, 1c and 1d can support the possibility that the early heterotrophic divergences in the heterokont lineage are the result of plastid loss. If that is true, then the endosymbioses should predate the origin of the hosts. If the converse is true, then the origin of the heterokont algae should coincide with the timing of the secondary endosymbiotic event leading to the transfer of photosynthetic capacity to heterokont organisms. If the haptophytes obtained their plastids at the origin of their lineage, then the timing of their symbiosis should be widely disparate from that of the pigmented heterokonts assuming no plastid loss.

The average date for the origin of the pigmented heterokonts calculated from the 18S rRNA gene (170 Ma) is close to the mean of the average date for a secondary endosymbiotic origin of the heterokont plastid estimated from three plastid genes (274 Ma) (Medlin *et al.* 1997, Table 2). This provides support for Hypotheses 1c and 1d over the remaining two hypotheses. The consistent separation of the haptophytes from the heterokonts in all of the plastid phylogenies provides evidence to support Hypothesis 1d over 1c.

The estimated average date for the origin of the haptophyte plastid is considerably younger than the origin of the host lineage. This would suggest that early members of the haptophytes were not photosynthetic and that the endosymbiosis occurred somewhere along the internode leading to the diversification of the haptophytes. The divergence of the two subclasses of the haptophytes, as estimated from the 18S rRNA and RUBISCO large subunit gene, is very close to the origin of the haptophyte plastid (Table 1). Medlin *et al.* (1997) hypothesized that the endosymbiotic event in the haptophyte lineage occurred just prior to the divergence of the two subclasses. Therefore, it follows that all early ancestors in the haptophyte lineage were heterotrophic and are either extinct, or are under-sampled. Novel haptophyte lineages have been found from clone li-

Table 2. Estimated time of origin of the host cells and their plastids from the heterokont chromophytes, diatoms and haptophytes taken from Medlin *et al.* (1997).

	Average time of origin in millions of years according to:				
	Host Cells		Plastids		
Gene	18S rRNA	16S rRNA	*tuf*A	RUBISCO	Mean of three Plastid Genes
Heterokont chromophytes	170	293	190	337	274
Diatoms	135	249	160	190	200
Haptophytes					
Group origin	850				
Subclass divergence	420			177	
plastid origin		263		322	293

braries with no knowledge of the cell's morphology to which the novel sequences were obtained (Edvardsen *et al.* 2000).

The estimated times for the two secondary endosymbiotic events in haptophytes and heterokonts, respectively, are also remarkably close. The mean divergence time of the two groups estimated from 3 genes is 281 Ma (n=5, Table 2). Thus, the transfer of the photosynthetic capacity to these lineages occurred approximately at the same time or before the Permian-Triassic boundary (250 Ma). Medlin *et al.* (1996b) have presented evidence for the correlation of the Permian-Triassic mass extinction with the re-radiation of the modern phytoplankton following this event. Both the pigmented heterokonts and the haptophytes comprise the bulk of today's eukaryotic phytoplankton in the oceans. Our molecular clock calculations would support a theory that the Permian-Triassic extinction opened many niches in the world's ocean and that those organisms capable of engulfing and maintaining a photo-autotroph had an adaptive advantage. This suggests that multiple secondary endosymbioses could have occurred at a similar time.

Molecular clock-calculated dates suggest that the haptophyte host cell lineage is relatively ancient (Proterozoic-Paleozoic) but that the haptophyte plastid was acquired more recently (Mesozoic) (Medlin *et al.* 1997). The molecular clock-calculated dates for the origin of the heterokont chromophytes are more recent (Mesozoic), both with respect to the host cell and the plastid. Perhaps coincidentally, the estimated dates for the origin of plastids in both haptophytes and heterokont chromophytes are nearly identical. Data suggest both groups first became photosynthetic at, or shortly before, the Permian-Triassic boundary. This would support a hypothesis that secondary endosymbioses, which represent a major evolutionary step in the advancement of the algae, may be associated with the major climatic changes at the end Permian and the mass extinctions that followed (Erwin 1994). The origin of the diatoms between 250 Ma and 180 MA as estimated by

our molecular clock is also consistent with the draw down of silica in the silica cycle at the time in the geological history of the earth (Siever 1991). Medlin *et al.* (1997) concluded that there were multiple secondary endosymbiotic events and that the timing of these events coincided primarily with the Permian-Triassic boundary.

CONCLUSIONS

It is clear that the diatoms belong to the pigmented heterokont algae, and with their sister group, the Bolidophyceae, represent the first divergence of the pigmented heterokont algae. Molecular clock calculations based on linearized trees from both nuclear and plastid genes and first appearance dates of the diatoms indicate that it is unlikely that the diatoms existed before the Permian Triassic boundary. The phylogeny of the diatoms recovered with these genes does not agree with their current classification system, which is based on the features of the siliceous cell wall and gross cytoplasmic details, such as number of plastids and type of sexual reproduction. The deeper clades recovered in the molecular tree are best supported by the arrangement and structure of the Golgi apparatus in extant taxa and by the cell wall morphology in the best preserved fossil deposits of diatoms from the early Cretaceous, whereas younger clades in the tree are best correlated with siliceous features of the cell wall.

REFERENCES

Andersen R. A., Saunders G. W., Paskind M. P. & Sexton J. P. 1993. Ultrastructure and 18S rRNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the description of a new algal class, the Pelagophyceae classis nov. J. Phycol. **29**: 701–715.

Bhattacharya D. & Medlin L. 1995. The phylogeny of plastids: A review based on comparison of small subunit ribosomal RNA coding regions. J. Phycol. **31**: 489–498

Bhattacharya D., Medlin L., Wainwright P. O., Ariztia E. V., Bibeau C., Stickel S. K., & Sogin M. L. 1992. Algae containing chlorophylls *a* + *c* are paraphyletic: molecular evolutionary analysis of the Chromophyta. Evol. **46**: 1808–1817. Errata Evol. **47**: 986.

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

Crawford R. M. 1973. The protoplasmic ultrastructure of the vegetative cell of *Melosira varians* C. A. Agardh. J. Phycol. **9**: 50–61.

Doyle J. J. & Doyle J. L. 1990. Isolation of plant DNA from fresh tissue. Focus **12**: 13–15.

Delwiche C. F. & Palmer J. D. 1996. Rampant horizontal transfer and duplication of Rubisco genes in Eubacteria and plastids. Mol. Biol. Evol. **13**: 873–882.

Edvardsen B., Eikrem W., Green J. C., Andersen R. A., Moon-Van Der Staay S. Y. & Medlin L. K. 2000. Phylogenetic reconstructions of the Haptophyta inferred from SSU rRNA sequence analyses. Phycologia, in press.

Elwood H. J., Olsen G. J. & Sogin M. L. 1985. The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata*. Mol. Biol. Evol. **2**: 399–410.

Erwin D. H. 1994. The Permo-Triassic extinction. Nature **367**: 231–236.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evol. **39**: 783–791.

Felsenstein J. 1993. "PHYLIP manual, Version 3.5", Department of Genetics, Univ. Washington, Seattle.

Fitch W. M. & Margoliash E. 1967. Construction of phylogenetic trees: a method based on mutation distances as estimated from cytochome *c* sequences is of general applicability. Science **155**: 279–284.

Fryxell G. A. & Hasle G. R. 1972. *Thalassiosira eccentrica* sp. nov. and some related centric diatoms. J. Phycol. 8: 297–317.

Fryxell G. A. & Hasle G. R. 1977. The genus *Thalassiosira*: some species with a modified ring of central strutted processes. Nova Hedwigia Beih. **54**: 67–98.

Gersonde R. & Harwood D. M. 1990. Lower Cretaceous diatoms from ODP Leg 113 site 693 (Weddell Sea). I: Vegetative cells. In: **P. F. Barker, J. P. Kennett** *et al.* (eds) *Proceedings* of the Ocean Drilling Program, Scientific Results, pp. 365–402, Ocean Drilling Program, College Station, Tx 113.

Gibbs S. P. 1978. The chloroplasts of *Euglena* may have evolved from symbiotic green algae. Can. J. Bot. **56**: 2882–2889.

Gibbs S. P. 1981. The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. Ann. N. Y. Acad. Sci. **361**: 193–208.

Gibbs S. P. 1993. The origin of algal chloroplasts. In: Lewin R. A. (ed.) Origins of plastids, pp. 107–121. New York: Chapman and Hall.

Guillou L., Chretiennot-Dinot M-J., Medlin L. K., Claustre H., Loiseaux-de Goer S. & Vaulot D. 1999. *Bolidomonas*, a new genus with two species belonging to a new algal class: The Bolidophyceae. J. Phycol. **35**: 368–381.

Haga M. 1997. Morphology of vegetative and resting spore valves of *Stephanopyxis nipponica*. Diatom Res. **12**: 217–228.

Hasle G. R. 1974. The mucilage pore of pennate diatoms. Nova Hedw. Beih. **45**: 167–186.

Hasle G. R. & Sims P. A. 1990. The diatom genus *Coscinodiscus* Ehrenb.: *C. argus* Ehrenb. and *C. radiatus* Ehrenb. Bot. Mar. **29**: 305–318.

Hasle G. R. & Syvertsen E. E. 1985. *Thalassiosiropsis*, a new diatom genus from the fossil record. Micropaleo. **31**: 82–91.

Herth W. 1979. The site of β-chitin fibril formation in diatoms. I. The chitin forming cytoplasmic structures. J. Ultrastr. Res. **68**: 16–27.

Hillis D. M. & Moritz C. 1990. An overview of applications of molecular systematics. In : D. M. Hillis & C. Moritz, (eds), Molecular Systematics, pp. 502–515. Sinauer Associates, Sunderland, Massachusetts.

Jeffrey S. W. 1989. Chlorophyll *c* pigments and their distribution in the chromophyte algae. In: **J. C. Green, B. S. C. Leadbeater, W. L. Diver,** (Eds): The Chromophyte Algae: Problems and Perspectives, pp. 13–36. Clarendon Press: Oxford.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**: 111–120

Kishino H. & Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the Hominoidea. J. Mol. Evol. **29**: 170–179.

Kooistra W. H. C. F. & Medlin L. K. 1996. Evolution of the diatoms (Bacillariophyta). IV. A reconstruction of their age from small subunit rRNA coding regions and the fossil record. Mol. Phylog. Evol. **6**: 391–407.

Kowallik K. V. 1993. Origin and evolution of plastids from chlorophyll *a+c* containing algae: Suggested ancestral relationships to red and green algal plastids. In: R. A. Lewin (ed.) Origin of Plastids, pp. 223–263. Chapman & Hall, New York.

Lauterborn R. 1895. Protozoenstudien II. *Paulinella chromatophora* nov. gen. nov. spec., Ein beschalter Rhizopode des Süsswassers mit blaugrünen chromatophorenartigen Einschlüssen. Zeit. Wiss. Zool. **59**: 537–547.

Li P. & Bousquet J. (1992). Relative rate test for nucleotide substitutions between two lineages. Mol. Biol. Evol. **9**: 1185–1189.

Lockhart P. J., Howe C. J., Bryant D. A., Beanland M. D. & Penny D. 1994. Substitutional bias confounds inference of cyanelle origins from sequence data. J. Mol. Evol. **34**: 153–162.

Maddison W. P. & Maddison D. R. 1992. "MacClade. Analysis of phylogeny and character evolution. Version 3," Sinauer Associates, Sunderland, Massachusetts.

Mann D. G. 1999. The species concept in diatoms. Phycologia **38**: 437–495.

Mann D. G. & Marchant H. J. 1989. The origin of the diatom and its life cycle. In: **J. C. Green, B. S. C. Leadbeater & W. L. Diver** (eds) The Chromophyte Algae: Problems and Perspectives, pp. 307–323, Clarendon Press, Oxford.

Manton I. & von Stosch H. A. 1966. Observations on the fine structure of the male gamete of the marine centric diatom *Lithodesmium undulatum*. J. Roy. Micr. Soc. **85**: 119–134.

Margulis L. 1981. Symbiosis in Cell Evolution. San Francisco, Freeman.

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

McFadden G. I., Gilson P. R. & Hoffman C. J. B. 1997. Division Chlorarachniophyta. In: **D. Bhattacharya** (ed.) Origins of the algae and their plastids. Pl. Sys. Evol. (Suppl.) **11**: 175–186.

Medlin L. K., Cooper A., Hill C., Wrieden-Prigge S. & Wellbrock U. 1995. Phylogenetic position of the Chromista plastids from 16S rDNA coding regions. Cur. Gen. **28**: 560–565.

Medlin L. K., Crawford R. M. & Andersen R. A. 1986. Histochemical and ultrastructural evidence for the function of the labiate process in the movement of centric diatoms. Br. phycol. J. **21**: 297–301.

Medlin L. K., Elwood H. J., Stickel S. & Sogin M. L. 1988. The characterization of enzymatically amplified eukaryotic 18S rRNA-coding regions. Gene **71**: 491–499.

Medlin L. K., Elwood H. J., Stickel S. & Sogin M. L. 1991. Morphological and genetic variation within the diatom *Skeletonema costatum* (Bacillariophyta) evidence for a new species, *Skeletonema pseudocostatum*. J. Phycol. **27**: 514–524.

Medlin L. K., Gersonde R., Kooistra W. H. C. F. & Wellbrock U. 1996a. Evolution of the diatoms (Bacillariophyta II. Nuclear-encoded small-subunit rRNA sequence comparisons confirm a paraphyletic origin for the centric diatoms. Mol. Biol. Evol. **13**: 67–75.

Medlin L. K., Kooistra W. H. C. F., Gersonde R. & Wellbrock U. 1996b. Evolution of the diatoms (Bacillariophyta). III. Molecular evidence for the origin of the Thalassiosirales. Nova Hedwigia **11**: 221–234.

Medlin L. K., Kooistra W. H. C. F., Potter D., Saunders G. W. & Andersen R. a. 1997. Phylogenetic relationships of the 'golden algae' (haptophytes, heterokont chromophytes) and their plastids. In: D. Bhattacharya (ed.) Origins of the algae and their plastids. Pl. Sys. Evol. (Suppl.) **11**: 187–219.

Medlin L. K., Williams D. M. & Sims P. A. 1993. The evolution of the diatoms (Bacillariophyta). I. Origin of the group and assessment of the monophyly of its major divisions. Eur. J. Phycol. **28**: 261–275.

Melkonian M. 1996. Phylogeny of photosynthetic protists and their plastids. Verh. Dtsch. Zool. Ges. **89**: 71–96.

Mereschkowsky C. 1905. Über Natur und Ursprung der Chromotaphoren im Pflanzenreiche. Biol. Cent. **25**: 593–604.

Neefs J-M., Van de Peer Y., De Rijk P., Goris A. & De Wachter R. 1991. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res. **19**(supp): 1987–2015.

Nikolaev V. A. & Harwood D. M. 1994. Morphology and taxonomic position of Cretaceous diatom genus *Pomphodiscus* Barker & Meakin. 13th International Diatom Symposium. Abstract pp. 75.

Nikolaev V. A. & Harwood D. M. 1997. New process, genus and family of Lower Cretaceous diatoms from Australia. Diatom Res. 12: 43–53.

Pickett-Heaps J., Hill D. R. A. & Weatherbee R. 1986. Cellular movement in the centric diatom *Odontella sinensis*. J. Phycol **22**: 234–239.

Pickett-Heaps J., Hill D. R. A. & Blazé K. L. 1991. Active gliding motility in an araphid marine diatom, *Ardissonea* (formerly *Synedra*) *crystallina*. J. Phycol. **27**: 718–725.

Rothpletz A. 1896. Über die Flysch-Fucoiden und einige andere fossile Algen, sowie über liasische, Diatomeen führende Hornschwämme. Zeitschr. Deutsche Geol. Ges. **48**: 954–914.

Round F. E., Crawford R. M. & Mann D. G. 1990. The Diatoms: Biology and Morphology of the Genera, 747 pp. Cambridge University Press, Cambridge.

Rowan K. S. 1989. Photosynthetic pigments of algae. Cambridge, Cambridge University Press.

Saiki R., Gelfand D. H., Stoffel S., Scharf S. J., Higuchi R., Horn G., Mullis T. K. B. & Erlich H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. Science, **239**: 487–491.

Sanger F., Nicklen S. & Coulsen A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463–5467.

Saunders G. W. & Druehl D. 1992. Nucleotide sequences of the small-subunit ribosomal RNA genes from selected Laminariales (Phaeophyta) implications for kelp evolution. J. Phycol. **28**: 544–549.

Schmid A-M. M.1984a. Valve morphogenesis in diatoms. In: K. Bach & B. Burckhardt (eds). Diatoms I. Shells in nature and technics. Communications of the Inst. Lightweight Structure IL 28, pp. 300–317, Cramer, Stuttgart,

Schmid A-M. M. 1984b. Wall morphogenesis in *Thalassiosira eccentrica*: comparison of auxospore formation and the effects of MT-inhibitors. In: D. G. Mann (ed.). Proceedings of the 7th International Diatom Symposium, Philadelphia 1982, pp. 42–70. O. Koeltz, Koenigstein.

Schmid A-M. M. 1987. Morphogenetic forces in diatom cell wall formation. In: J. Bereiter-Hahn, O. R. Anderson, W. E. Reif (eds) Cytomechanics, pp. 183–199. Springer, Berlin, Heidelberg, New York, Tokyo.

Schmid A-M. M. 1988. The special Golgi-ER-mitochondrium unit in the diatom genus *Coscinodiscus*. Pl. Syst. Evol. **158**: 211–223.

Schmid A-M. M. 1989. Geitler's "Plattenband" in the diatom *Synedra* cf. *ulna* in the light of TEM investigations. Pl. Syst. Evol. **164**: 239–252.

Schmid A-M. M. 1994a. Aspects of morphogenesis and function of diatom cell walls with implications for taxonomy. Protoplasma **181**: 43–60.

Schmid A-M. M. 1994b. Slit scales in the auxospore scale case of *Coscinodiscus granii*: the rudiments of rimoportulae? Diatom Res. **9**: 371–375.

Schmid A-M. M., Borowitzka M. A. & Volcani B. E. 1981. Morphogenesis and biochemistry of diatom cell walls. In: O. Kiermayer (ed.) Cytomorphogenesis in plants, pp. 63–97. Springer, Wien, New York.

Schmid A. M. & Crawford R. M. *Ellerbeckia arenaria* (Bacillariophyta): Formation of the auxospores and initial cells. Eur. J. Phycology, in press.

Siever R. 1991. Silica in the Oceans: Biological-Geochemical Interplay. In: Scientists on Gaia. pp. 287–295, Cambridge Ma. MIT Press.

Sims P. A. 1994. *Benetorus*, *Gladiopsis* and related genera from the Cretaceous. Diatom Res. **9**: 165–187.

Simonsen R. 1972. Ideas for a more natural system of the centric diatoms. Nova Hedwigia **29**: 37–54.

Simonsen R. 1979. The diatom system: ideas on phylogeny. Bacillaria **2**: 9–71.

Sitte P. 1993. Symbiogenetic evolution of complex cells and complex plastids. Europ. J. Protistol. **29**: 131–143.

Sörhannus U., Gasse F., Perasso R. & Baroin-Tourancheau A. 1995. A preliminary phylogeny of diatoms based on 28S ribosomal RNA sequence data. Phycologia **34**: 65–73.

Swofford D. L. 1993. PAUP, Phylogenetic Analysis Using Parsimony, Version 3.1, program and documentation. Champaign, IL: Illinois Natural History Survey, University of Illinois.

Taggart R. E. & Parker L. R. 1976. A new fossil alga from the Silurian of Michigan. Amer. J. Bot. **63**: 1390–1392.

Takezaki N., Rzhetsky A. & Nei M. 1995. Phylogenetic test of the molecular clock and linearized trees. – Mol. Biol. Evol. **12**: 823–833.

Valentin K., Cattolico R. A. & Zetsche K. 1993. Phylogenetic origin of the plastids. In: R. A. Lewin (ed.), Origins of plastids pp. 193–22. New York, Chapman and Hall.

Woese C. R. 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271.

Wray G. A., Levington J. S. & Shapiro L. H. 1996. Molecular evidence for deep Precambrian divergences among Metazoan phyla. – Science **274**: 568–573.

Wu C.-I. & Li W-H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82**: 1741–1745.

Xiao S., Knoll A. H. & Yuan X. 1998. Morphological reconstruction of *Miaohephyton bifurcatum*, a possible brown alga from the neoproterozoic Douchantuo Formation, South China. J. Paleont. **72**: 1072–1086.

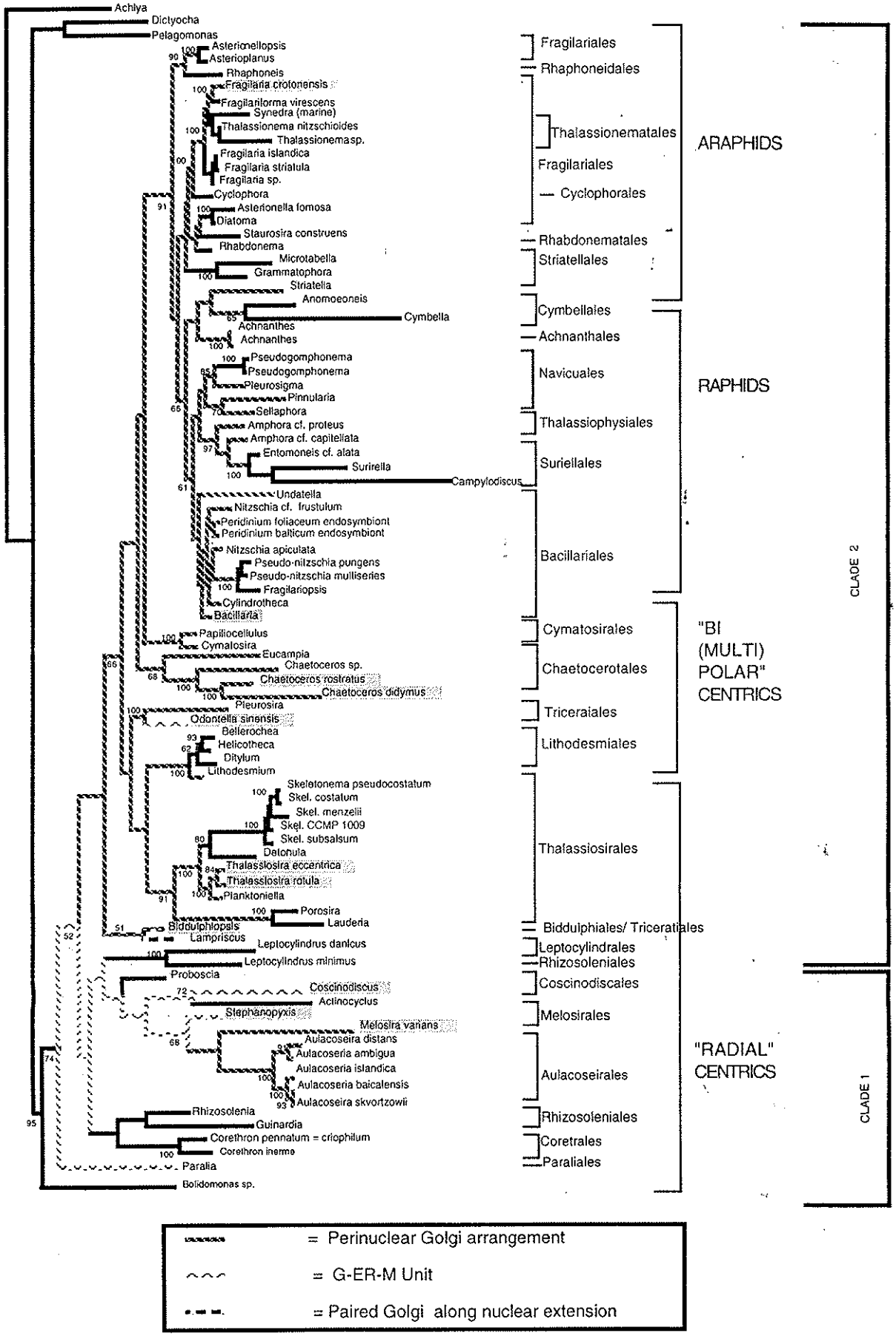


Fig. 1. Phylogeny inferred with the maximum parsimony analysis. The numbers to the left of the nodes are the bootstrap values obtained from 100 replications of the data set using a neighbor joining analysis (Felsenstein 1989). The type of Golgi apparatus possessed by the diatoms is also plotted onto the tree. The branch length is pattern-coded (see legend in tree) for three of the four types of Golgi body arrangements found so far in the diatoms (Fig. 2.). If the genus or family is known to have one particular Golgi body arrangement, then the branch length leading to the terminal taxa is pattern-coded. If the actual species sequenced has been investigated and the Golgi body arrangement known, then the name of the taxon is highlighted. Diatom orders are bracketed in the tree and as well as the major clades recovered in the analysis.

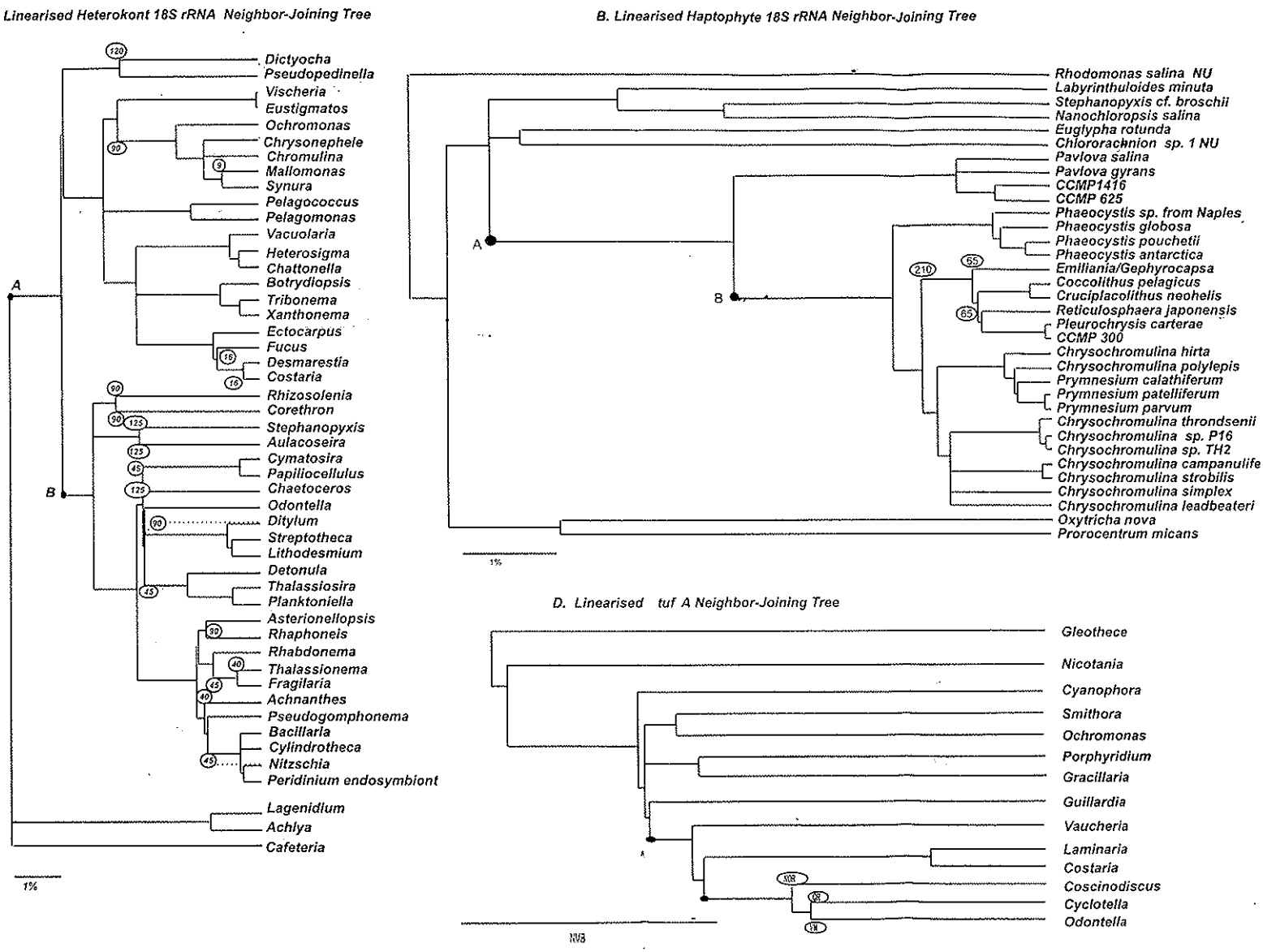


Fig. 4. Linearized neighbor-joining trees constructed from unweighted Kimura-2-parameter distances or a gamma distribution and from unlinearized NJ trees according to Takezaki *et al.* (1995) for A, the SSU nucle the Haptophyta, C, the *tuf*A protein coding gene, D, the large subunit of RUBISCO, and E, the SSU plastid-encoded rRNA gene. The linearization process collapsed some nodes into polytomies. Encircled numbe which the taxa to the right belong. In tree A, point A marks the divergence of the pigmented lineages from the heterotrophic lineages, point B is the divergence of the diatoms from other pigmented heterokonts. subclasses. In tree C, point A marks the origin of the heterokont plastid; B is the origin of the diatom plastid. In trees D & E, point A marks the origin of the heterokont plastid; B is the origin of the diatom plastid; C i

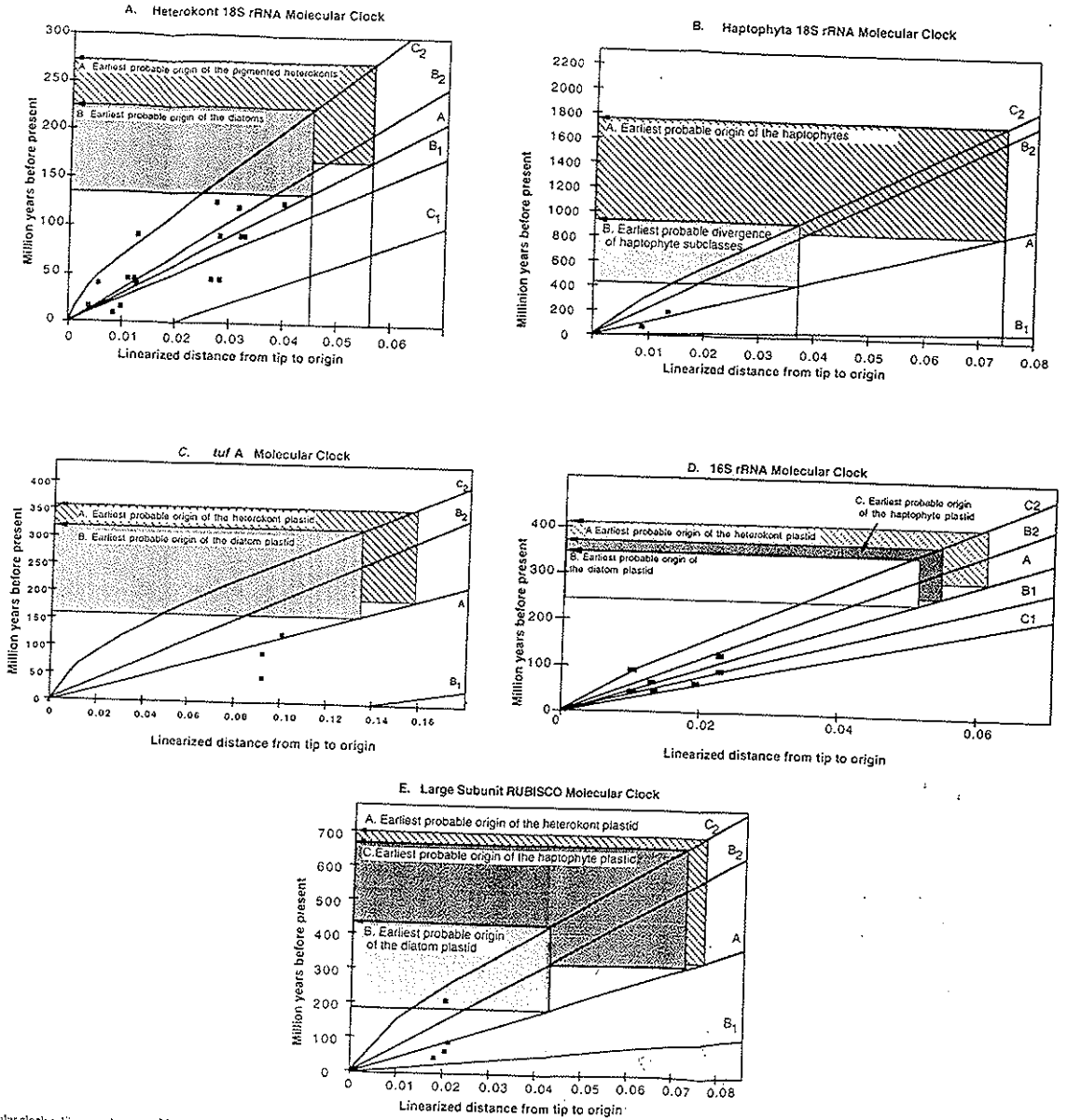


Fig. 5. Molecular clock calibration for each of the linearized trees in Fig. 4. A, from the SSU nuclear-encoded rRNA gene from the Heterokonta B, from the SSU nuclear-encoded rRNA gene from the Haptophyta C, from the *tuf*A protein coding gene D, from the large subunit of RUBISCO, and E from the SSU plastid-encoded rRNA gene. First appearances of haptophytes or heterokonts from the fossil record were regressed against measured branch lengths from the linearized NJ trees in Fig. 1. For each clock: A is the regression line, constrained through the origin. Lines B1 and B2 are the 95% confidence limits around the regression line. Lines C1 and C2 are the 95% confidence limits for a new predicted value of time given the length of an undated node[2]. Lower confidence limits, below zero, are reset at zero. Blocks of time are shown for each group whose origin has been estimated from the molecular clock. The block spans the time of the average age of the group (A) to the earliest possible time of origin based on the upper 95% confidence interval (C_2) of an undated node. Redrawn from Medlin *et al.* 1997.