A Time line of the Environmental Genetics of the Haptophytes

Hui Liu,[1,2] Stéphane Aris-Brosou,*[,3] Ian Probert,[2] and Colomban de Vargas[2]

[1]Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ

[2]CNRS-UPMC, Station Biologique de Roscoff, UMR 7144, Evolution du Plancton et PaléoOcéans, Roscoff, France

[3]Department of Biology and Mathematics and Statistics, Centre for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

***Corresponding author:** E-mail: sarisbro@uottawa.ca.

Associate editor: Andrew Roger

Abstract

The use of genomic data and the rise of phylogenomics have radically changed our view of the eukaryotic tree of life at a high taxonomic level by identifying 4–6 "supergroups." Yet, our understanding of the evolution of key innovations within each of these supergroups is limited because of poor species sampling relative to the massive diversity encompassed by each supergroup. Here we apply a multigene approach that incorporates a wide taxonomic diversity to infer the time line of the emergence of strategic evolutionary transitions in the haptophytes, a group of ecologically and biogeochemically significant marine protists that belong to the Chromalveolata supergroup. Four genes (*SSU*, *LSU*, *tufA*, and *rbcL*) were extensively analyzed under several Bayesian models to assess the robustness of the phylogeny, particularly with respect to 1) data partitioning; 2) the origin of the genes (host vs. endosymbiont); 3) across-site rate variation; and 4) across-lineage rate variation. We show with a relaxed clock analysis that the origin of haptophytes dates back to 824 million years ago (Ma) (95% highest probability density 1,031–637 Ma). Our dating results show that the ability to calcify evolved earlier than previously thought, between 329 and 291 Ma, in the Carboniferous period and that the transition from mixotrophy to autotrophy occurred during the same time period. Although these two transitions precede a habitat change of major diversities from coastal/neritic waters to the pelagic realm (291–243 Ma, around the Permian/Triassic boundary event), the emergence of calcification, full autotrophy, and oceanic lifestyle seem mutually independent.

Key words: Haptophyta, divergence times, ancestral characters, Bayes, maximum likelihood.

Research article

Introduction

Eukaryotes are provisionally subdivided into six supergroups (Opisthokonta, Amoebozoa, Archaeplastida, Excavata, Rhizaria, and Chromalveolata) whose phylogenetic relationships are slowly emerging (e.g., Lane and Archibald 2008). The Chromalveolata, 1 of the 6 eukaryotic supergroups, comprise a disputed assemblage made of eukaryotes with red algal–derived plastids that originate ultimately from a common secondary endosymbiosis (Yoon et al. 2004). This potentially paraphyletic or even polyphyletic supergroup is composed of the alveolates (dinoflagellates, apicomplexans, and ciliates) and the chromists (stramenopiles, cryptophytes, and haptophytes) and accounts for about half of the described diversity of protists (Cavalier-Smith 2004). Recent studies found that 4 of these 6 lineages (apicomplexans, ciliates, dinoflagellates, and stramenopiles) consistently form a monophyletic assemblage, whereas the remaining two lineages (cryptophytes and haptophytes) form a weakly supported group that remains to be substantiated (e.g., Harper et al. 2005; Hackett et al. 2007; Hampl et al. 2009; but see Rice and Palmer 2006; Patron et al. 2007). These studies provide important insights into the basal relationships between these lineages, but they do not have the taxonomic coverage that would allow us to infer the emergence of key evolutionary transitions within each lineage, in particularly within the haptophytes.

The present study focuses on the haptophytes, one of the most abundant groups of oceanic phytoplankton and significant primary producers (Thomsen et al. 1994; Field et al. 1998). Haptophytes, or Haptophyta, differ from other eukaryotes by possessing a unique flagellum-like organelle, the haptonema, that is thought to play a role in prey capture in some species (Kawachi et al. 1991). Another unique feature found in the coccolithophores or Calcihaptophycideae, the best-known members of this division, is the presence of a calcified exoskeleton consisting of minute, intracellularly formed, calcite platelets (coccoliths) that sediment to the ocean floor upon death of cells, resulting in the formation of limestone and chalk deposits over geological time. Based on morphology, the division Haptophyta is classically subdivided into two classes: the Pavlovophyceae, asymmetrical cells covered by organic knoblike scales and with anisokont (unequal length) flagella, and the Prymnesiophyceae, symmetrical cells covered by organic plate scales (that serve as the matrix for calcification in the coccolithophores) and with isokont flagella. The coccolithophores are presently responsible for the bulk of oceanic calcification (Milliman 1993). Consequently, they heavily influence the marine carbonate system and have a major impact on the global carbon cycle. The fossil archive of the coccolithophores is probably the most complete of any protist lineage, with 20–30% of species leaving a fossil

record (Young et al. 2005), and this archive has been intensively studied by biostratigraphers (e.g., Bown 1998). Certain haptophytes, such as members of the genera *Emiliania*, *Gephyrocapsa*, *Phaeocystis*, *Chrysochromulina*, and *Prymnesium*, are responsible for extensive blooms that have major biogeochemical, ecological, and economic impacts (Brown and Yodar 1994; Robertson et al. 1994; Edvardsen and Paasche 1998; Lancelot et al. 1998). For example, massive blooms of the coccolithophore *Emiliania huxleyi* are thought to affect global climate by increasing water albedo through dimethylsulfide production and also drive large fluxes of calcium carbonate out of surface waters (Tyrrel and Merico 2004). As focus is increasingly falling on the impacts of rising anthropogenic CO_2 on the carbonate system in the ocean, a better understanding of the diversification of the haptophytes and how this diversification has correlated with past environmental conditions may help predict how these species will react to future environmental change (Fabry 2008). However, despite their ecological, biogeochemical, and geological role, our knowledge of the diversification of this division is still limited to what is known from the coccolithophores, the only members of the haptophytes that leave traces in the fossil record; yet, coccolithophores represent less than half of the existing diversity of the haptophytes (Young et al. 2005).

The molecular studies that pioneered the reconstruction of the diversification of the haptophytes used either a single slowly evolving nuclear gene such as the 18S rDNA (*SSU*: Medlin et al. 1997; Simon et al. 1997; Edvardsen et al. 2000) or faster evolving plastid genes such as *rbcL* (Fujiwara et al. 1994; Daugbjerg and Andersen 1997; Inouye 1997; Fujiwara et al. 2001) or *tufA* (Sáez et al. 2003). These early studies supported the morphological taxonomy by dividing haptophytes into two main clades: the Prymnesiophyceae and the Pavlovophyceae. Yet, phylogenetic resolution beyond this taxonomic level was still limited. In combination with morphological, physiological, and ecological data, more recent molecular approaches further recognized four major clades (Prymnesiales, Coccosphaerales, Isochrysidales, and Phaeocystales) within the Prymnesiophyceae (Edvardsen et al. 2000). However, the resolution of these molecular studies remained poor, particularly within the coccolithophore clade. The most comprehensive molecular phylogenetic reconstructions of the Haptophyta to date are those of Medlin et al. (2008) using sequences of the nuclear *SSU* and plastid *tufA* genes from ca. 60 cultured species.

With increasing molecular phylogenetic resolution and an outstanding fossil record for the past 220 million years (My) (Bown 1998), the haptophytes are an ideal group for applying molecular clock methods to date key transitions and unravel the tempo of their evolution. The inadequacy of the strict molecular clock is no longer controversial for most modern data sets, but known limitations can be alleviated by meeting four general conditions (Soltis et al. 2002; Yoon et al. 2004): 1) use of a well-supported and accurate tree that resolves all important nodes (normally entailing the use of large multigene data sets); 2) use of reliable fossil calibrations; 3) use of methods that account for substitution rate heterogeneity within and across lineages; and 4) broad taxon sampling. An early strict molecular clock study based on an *SSU* phylogeny estimated that the haptophytes diverged from other chromists between 1,750 and 850 million years ago (Ma; Medlin et al. 1997). Two subsequent studies that did not assume a strict molecular clock, one based on six plastid genes (Yoon et al. 2004) and the other one on a single ribosomal gene (*SSU*; Berney and Pawlowski 2006), narrowed down the previous estimate to ~1,100–900 Ma. A more recent molecular study based on two genes, *SSU* and *tufA*, did include more representatives of the haptophytes (Medlin et al. 2008) and dated the origin of the Haptophyta at ca. 1,200 Ma. However, this latter study 1) did not discuss the use of multiple gene partitions to estimate the tree used for dating, 2) assumed that the two genomes, nuclear and plastid, share the same history, 3) assumed that this phylogeny is known with an absolute certainty in order to estimate divergence times, and 4) was still based on a strict molecular clock that limited the analysis to the only gene (*SSU*) following approximately this strict clock hypothesis. Apart from the dating controversy, it was also suggested that extant coccolithophores diversified from a few lineages that survived the major extinction at the Cretaceous/Tertiary (K/T) boundary, whereas noncalcifying haptophytes were not affected by the K/T extinction (Medlin et al. 2008). The adaptation of noncalcifying haptophytes to eutrophic coastal environments and their ability to switch nutrition modes from autotrophy (photosynthesis only) to mixotrophy (photosynthesis + particle grazing, which requires some phagocytic ability) were posited as possible explanations for their survival during this abrupt global change event. Such a parsimony-driven reconstruction of character states from their observed distribution in contemporary organisms highlights the possibility of using ancestral reconstructions to glimpse the past by discovering how nonfossilizable traits evolved. Statistically robust computational methods are available to reconstruct ancestral characters or states, even in the presence of uncertainty in estimates of the tree and its branch lengths (e.g., Pagel et al. 2004).

Here we resolve the mode and tempo of the diversification of the haptophytes using an extensive multigene analysis that includes both nuclear (*SSU* and *LSU*) and plastid (*tufA* and *rbcL*) gene sequences for a total of 5,006 bp. Our species sampling includes 34 representative taxa from the Pavlovophyceae and the Prymnesiophyceae, the latter including members of all formally described extant orders. Our analyses show that 1) the haptophytes evolved ca. 824 Ma (1,031–637 Ma), 2) the nuclear and plastid genomes share the same history within the haptophytes, and 3) the reconstruction of this history is not plagued by artifacts such as long-branch attraction (LBA) due to general model misspecification. Moreover, we reconstruct and date four key transitions: the evolution of calcification and organic scales and the switches from coastal to oceanic dwelling as well as from mixotrophic to autotrophic nutrition mode. The timing of these key evolutionary transitions is interpreted in an ecological and geological context.

Table 1. Accession Numbers of the Sequences Included in This Study.

Species	*LSU*	*SSU*	*tufA*	*RbcL*
Coccolithus pelagicus	EU729464[a]	AJ246261	AJ544128	X
Calcidiscus quadriperforatus	EU502878[a]	AJ544115	AJ544124	X
Calcidiscus leptoporus	EU729460[a]	AJ544116	AJ544126	AB043690
Umbilicosphaera hulburtiana	EU729463[a]	AM490993	AM502981	X
Umbilicosphaera foliosa	EU729462[a]	AJ544119	AJ544130	AB043629
Umbilicosphaera sibogae	EU729461[a]	AJ544118	AJ544129	AB043691
Cruciplacolithus neohelis	EU729467[a]	AB058348	X	AB043689
Calyptrosphaera sphaeroidea	EU729466[a]	AM490990	X	AB043628
Helladosphaera sp.	EU729465[a]	AB183607	X	X
Hymenomonas coronata	EU819083	AM490981	X	X
Jomonlithus litoralis	EU502875[a]	AM490979	X	X
Hymenomonas globosa	EU502872	AM490982	X	X
Ochrosphaera neapolitana	EU729469[a]	AM490980	X	X
Ochrosphaera sp.	EU819082	AB183615	X	X
Pleurochrysis carterae	EU819084	AJ246263	AJ544131	D11140
Pleurochrysis dentata	EU729468[a]	AJ544121	AJ544132	AB043688
Gephyrocapsa oceanica	EU729476[a]	AB058360	AF545609	D45844
Isochrysis galbana	EU729474[a]	AJ246266	AF545610	AB043693
Isochrysis littoralis	EU819085	AM490996	X	X
Algirosphaera robusta	EU729470[a]	AM490985	AM502934	X
Coronosphaera mediterranea	EU729471[a]	AM490986	AM502941	X
Syracosphaera pulchra	EU502879[a]	AM490987	X	X
Helicosphaera carteri	EU729473[a]	AM490983	AJ544134	X
Scyphosphaera apsteinii	EU729472[a]	AM490984	X	X
Prymnesium patelliferum	AF289038	L34671	X	X
Prymnesium parvum	EU729443[a]	AJ246269	X	AB043698
Prymnesium sp.	EU729445[a]	U40923	X	X
Platychrysis pigra	EU729458[a]	AM491003	X	X
Imantonia rotunda	EU729457[a]	AJ246267	X	X
Phaeocystis sp.	EU729477[a]	X77475	X	X
Pavlova virescens	EU729477[a]	AJ515248	AF545612	X
Pavlova pinguis	EU502883[a]	AF106047	X	X
Rebecca salina	EU729478[a]	L34669	X	AB043633
Exanthemachrysis gayraliae	EU729479[a]	AF106060	X	X
Vaucheria bursata	AF409127	U41646	U09448	AF476940
Tribonema aequale	Y07979	M55286	AF038002	AF084611
Undaria pinnatifida	AY851528	AF319007	AF038003	AY851535
Costaria costata	AY851522	AB022819	U09429	AY851541
Heterosigma akashiwo	AF409124	AB183667	AF545613	AB176660
Skeletonema costatum	EF433522	AY684947	AF015569	AF545615

NOTE.—X, missing genes.

[a] Accession numbers of the sequences obtained in this study (see supplementary table S1, Supplementary Material online, for corresponding RCC identifiers).

Materials and Methods

Taxonomic Sampling and Culture Conditions

About 430 clonal culture strains of haptophytes were isolated and maintained as described in Probert and Houdan (2004). The majority of these strains are available from the Roscoff Culture Collection (RCC: http://www.sb-roscoff.fr/Phyto/RCC/). Taxonomic identification of cultures was based on transmission electron microscopy observation of body scale morphology for nonmineralized taxa and scanning electron microscopy observation of coccolith morphology for mineralized taxa. Taxonomic concepts used here follow those of Young et al. (2003) and Jordan et al. (2004). Partial *LSU* sequences for ca. 300 strains were obtained over the course of this study (see supplementary table S1, Supplementary Material online, for a list of all sequenced strains). We included four gene sequences (*SSU*, *LSU*, *tufA*, and *rbcL*) from each of 34 species of haptophyte and 6 nonhaptophyte taxa in our analysis (table 1). Species sampling within the haptophytes included 2–3 representatives of all genera available from the RCC, and for each species, sequences of at least 3 of the 4 genes included in the analysis (*SSU*, *LSU*, and *tufA*) originated from the exact same culture strain. Our choice of nonhaptophyte taxa to root the tree was guided by the availability of the four gene sequences in GenBank. When this data set was assembled (October 2007), the closest outgroup sequences found by BlastN searches were from six stramenopiles (table 1).

LSU Gene Sequencing

Exponential phase cultures were harvested by centrifugation (1,000 rpm for 5 min) and 100 μl of GITC* DNA extraction buffer (4 M guanidine thiocyanate, 50 mM Tris–HCl [pH 7.6], 2% *N*-lauroyl-sarcosine, 0.1 M β-mercaptoethanol) were added to the cell pellet. Cells in buffer were stored at −20 °C until analysis. Total DNA was extracted using the DNeasy Plant MiniKit (Qiagen, Hilden, Germany) following the instructions from the manufacturer. A

Table 2. Specification of CCs Used in BEAST.

	Root	Node 47	Node 57	Node 62	Node 63	Node 77	Node 79
4 CCs	LN(0.0, 0.5) + 0.5	Ø	Ø	E(0.01) + 0.065	E(0.01) + 0.024	E(0.01) + 0.055	E(0.01) + 0.025
5 CCs	LN(0.0, 0.5) + 0.5	Ø	E(0.1) + 0.220	E(0.01) + 0.065	E(0.01) + 0.024	E(0.01) + 0.055	E(0.01) + 0.025
6 CCs	LN(0.0, 0.5) + 0.5	E(0.2) + 0.350	E(0.1) + 0.220	E(0.01) + 0.065	E(0.01) + 0.024	E(0.01) + 0.055	E(0.01) + 0.025

NOTE.—Node identifiers represent the following divergences (see supplementary fig. S1, Supplementary Material online, for details): *Exanthemachrysis gayraliae* and *Helicosphaera carteri* (node 47); *Coccolithus pelagicus* and *H. carteri* (node57); *C. pelagicus* and *Umbilicosphaera hulburtiana* (node 62); *Calcidiscus leptoporus* and *Umbilicosphaera foliosa* (node 63); *Coronosphaera mediterranea* and *Scyphosphaera apsteinii* (node 77); and *H. carteri* and *S. apsteinii* (node 79). LN(*x*,*y*), lognormal distribution with mean *x* and variance *y*; E(*x*), exponential distribution with parameter *x*; Ø, no CC specified. Times are in billion years. A lognormal process was assumed for modeling the evolution of rates of evolution across lineages. Plus signs (+) indicate the offset applied to each prior (minimum age setting).

nuclear *LSU* rDNA fragment of 941 bp containing the D1 and D2 domains was polymerase chain reaction (PCR) amplified using a set of eukaryotic primers in forward: Leuk2 (5′-acccgctgaacttaagcatatcact-3′) and in reverse: Euk_34r (5′-gcatcgccagttctgcttacc-3′). PCRs were performed using REDTaq DNA polymerase (Sigma-Aldrich) and a PCR× enhancer system (Invitrogen) in order to amplify GC-rich haptophyte sequences. The reaction followed denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 68 °C for 2 min. Thirty-five cycles were performed with initial denaturation and final extension steps. PCR products were purified using Qiaquick PCR purification kit (Qiagen) and then sequenced in both directions using a 3100-Avant Genetic Analyzer. All sequences obtained in this study were deposited in GenBank (see table 1 for accession numbers).

Computational Analyses

The four genes, *SSU*, *LSU*, *rbcL*, and *tufA*, were aligned individually with Clustal ver. 1.83 (Thompson et al. 1997). Alignments were visually inspected and edited where necessary with the Genetic Data Environment ver. 2.2 software (Larsen et al. 1993). Two sets of alignments were analyzed: a "complete alignment" and an alignment where ambiguous regions were removed (*LSU*: positions 344–434, 514–612, 725–777, and 1015–1026; *SSU*: positions 1204–1238, 1270–1295, and 2612–2620). Both alignments are available upon request.

Phylogenetic analyses were based on several Bayesian approaches in order to test the robustness of our results to a number of assumptions. First, the four genes were concatenated into one single partition that was analyzed under general time reversible (GTR) + Γ_4 + I, as selected by Modeltest (Posada and Crandall 1998) based on the Akaike information criterion. BEAST (ver. 1.4.8; Drummond and Rambaut 2007), which permits the joint estimation of tree topology and divergence times, was employed. Uncertainty in the mean substitution rate was integrated out along the Markov chain Monte Carlo (MCMC) samplers. Speciation times were assumed to follow a pure birth (Yule) process, and rates of evolution were assumed to follow an uncorrelated lognormal process (Drummond et al. 2006). Calibration constraints (CCs) were set as minimum divergence ages, represented by the offsets of the exponential prior distributions (table 2). These included five fossil dates based on nannofossil biostratigraphy (e.g., Perch-Nielsen 1985; Young 1998; Bown et al. 2004; see details below) and one additional weak constraint from a previous molecular clock estimate, the divergence of the two haptophyte classes (node 47 of supplementary fig. S1, Supplementary Material online, estimated to be well >350 Ma by both Berney and Pawlowski 2006 and Medlin et al. 2008) in order to test whether our estimations were biased by the use of relatively young (<220 Ma) fossil constraints. To assess the robustness of our results with respect to the type and number of constraint, three models were run: with 4, 5, or 6 CCs (table 2). The five-CC analysis excluded the molecular clock-based constraint at node 47, whereas the four-CC analysis also excluded the sole character-based constraint (node 57—see "fossil constraints" below). For each model, four independent MCMC samplers were run. Each sampler was run for 25 million steps with 2,000 steps of thinning. Convergence was checked with Tracer, which was also used to compute marginal probabilities of the data. The initial two million steps were removed as a burn-in, and results from all four runs were merged with an in-house script that removes burn-in periods and uses BEAST's treeannotator to summarize the results. The final results were analyzed with R (http://cran.r-project.org/).

Second, to assess the impact of concatenating genes that evolve at different rates, we performed two sets of analyses: 2) the four genes were concatenated or 2) the data were partitioned according to the four sampled genes. Under this latter partitioning scheme, the two protein-coding genes, *rbcL* and *tufA*, were further partitioned across the three coding positions, so that in total eight partitions were considered. Partitions only shared the tree topology, all the other parameters being independent or "unlinked." Here again, the most appropriate model of evolution was selected with Modeltest (Posada and Crandall 1998) based on the Akaike information criterion. The resulting model, GTR + Γ_4 + I, was used with MrBayes ver. 3.1.2 (Ronquist and Huelsenbeck 2003). Each MCMC sampler was run for five million steps; autocorrelation was decreased by sampling every 1,000 steps (thinning); and mixing was improved by using tempering with three heated chains. Two independent such samplers were run to check convergence under each model of evolution (with or without partition); at stationarity, split frequencies were checked to be <0.015. The first two million steps were discarded as burn-in. Trees were compared with the SH test (Shimodaira and Hasegawa 1999) as implemented in PAML 4 (Yang 2007) and by estimating marginal probabilities as in Suchard et al. (2001) with Tracer (http://tree.bio.ed.ac.uk/software/tracer/). For this partition test, eight partitions were

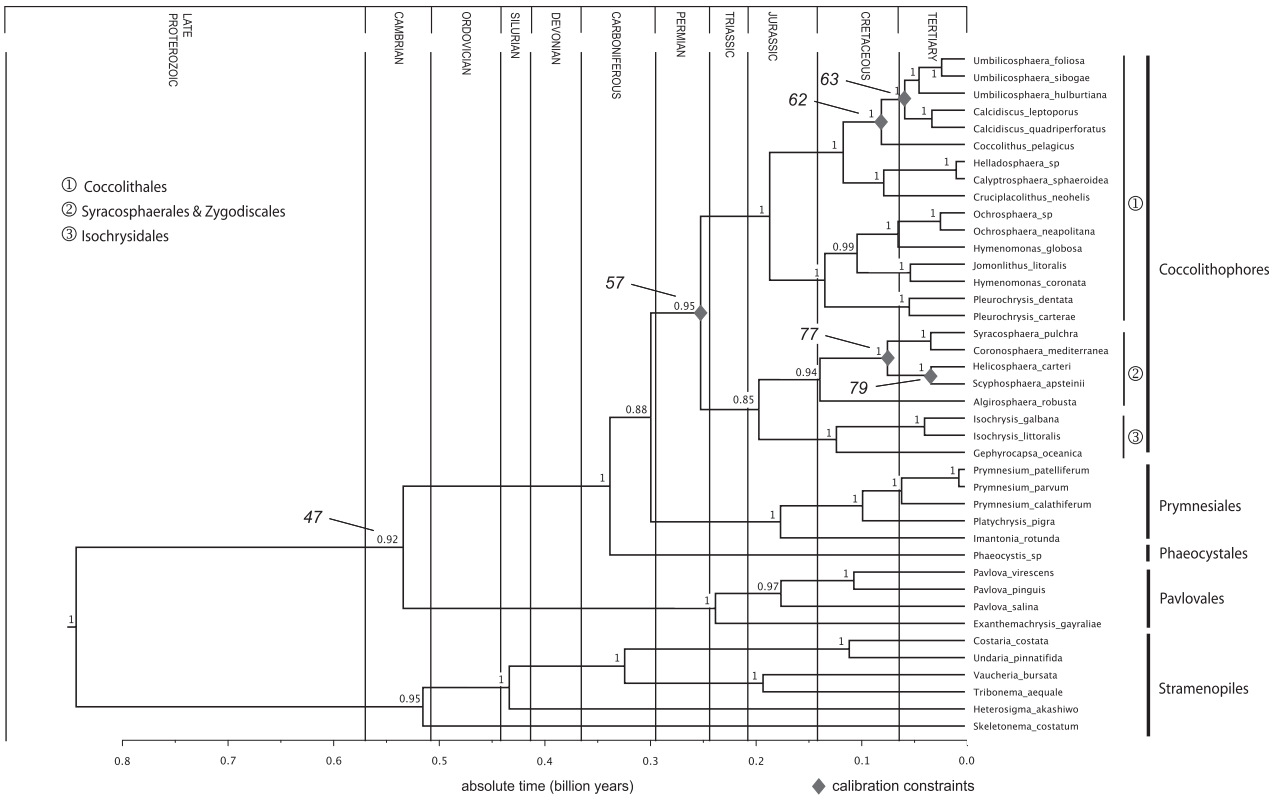

FIG. 1. Phylogeny and divergence times of the Haptophytes (BEAST analysis). The lognormal model of rate change with five CCs was assumed (see text). Placement of CCs on the tree is indicated by red diamonds, labeled as in supplementary figure S1 (Supplementary Material online). Numbers represent clade PP. Times are in billion years.

assumed and the GTR + Γ_5 nucleotide substitution model was used with all parameters unlinked.

zThird, the robustness to LBA artifacts was assessed by successively removing the taxa that showed the longest root-to-tip branch lengths as in Brinkmann et al. (2005) and Hampl et al. (2009) and rerunning the MrBayes and BEAST analyses as described above. The MrBayes analyses can be construed as "unconstrained," in the sense that the time dependency of the evolutionary process is not taken into account; on the other hand, the BEAST analyses directly incorporate the time dependency of the evolutionary process. For all these analyses, we tracked stability in terms of posterior probabilities (PP) of five groups, the stramenopiles (outgroup), the Pavlovales, the Phaeocystales (one species), the Prymnesiales, and the coccolithophores, as a function of the number of taxa removed.

Fourth, we tested if both nuclear and plastid genes reconstructed the same phylogeny, as an analysis of deep divergences based on both nuclear and plastid genes might be affected by endosymbiotic events. For assessing this potential effect, we ran two separate analyses with MrBayes. The first included the two nuclear RNA genes with two unlinked partitions. The second analysis included the two protein-coding plastid genes with six unlinked partitions (three codon positions for each gene). For this comparison of nuclear versus plastid trees, species whose plastid genes were not included in our data were removed from the nuclear tree with Analysis of Phylogenetics and Evolution (APE; Paradis 2006).

Finally, we tested for the potential effect of saturation due, on the one hand, to multiple substitutions at highly exchangeable nucleotides and, on the other hand, to variation of the rate of evolution in time. These two substitution processes can be responsible for incorrect phylogenetic reconstructions (Lartillot et al. 2007) due to the LBA artifact (Felsenstein 1978). The CAT-GTR model (Lartillot and Philippe 2004), abbreviated as CAT here for CATegory and implemented in PhyloBayes (ver. 2.3c), accounts for spatial variation of substitution rates (across sites). It was used here to assess the potential impact of across-site rate variation on the reconstructed phylogenetic trees. The CAT-break point (CAT-BP) model (Blanquart and Lartillot 2008) implemented in nh_PhyloBayes accounts for both the spatial (across sites) and the temporal variation of substitution rates (across lineages). It was used here to test for the effect of rate variation in time (BP model) or in space and time (CAT-BP model). The four genes were concatenated into one single partition. Two independent MCMC samplers were run for 10^5 steps under each model, and split frequencies were checked to be <0.015 at stationarity.

Ancestral characters and paleoenvironments were reconstructed by maximum likelihood with the R package APE (Pagel et al. 2004; Paradis 2006). All analyses are based on the consensus tree estimated under the model described above and implemented in BEAST (see fig. 1 for support values). The outgroup species (stramenopiles) were removed from the BEAST consensus tree with APE.

Characters and environmental features were assumed to follow a model where all rates of change are different.

Fossil Constraints

Two approaches can be used to place temporal constraints on the internal nodes of a tree using either "character-based" constraints or "divergence-based" constraints (Medlin et al. 2008). Character-based constraints refer to the first occurrence (FO) in the fossil record of a shared derived character or synapomorphy; divergence-based constraints refer to the FO of an ancestor from which descendants within a clade evolved. Of the five fossil constraints used in this study, the oldest (node 57 of supplementary fig. S1, Supplementary Material online) was a character-based constraint for the FO of heterococcoliths (i.e., coccoliths consisting of cycles of interlocking crystal units produced during the diploid phase of many coccolithophores). Heterococcolith calcification, a highly distinctive character (Young et al. 1999), is present in the entire coccolithophore clade in our tree (with a secondary loss in the Isochrysidales). In the fossil record, the first heterococcoliths occur in the Norian stage of the Late Triassic, ca. 220 Ma (Bown 1998).

The four other constraints employed were divergence based. From the fossil record alone, a number of uncertainties persist as to the phylogenetic relationships between the Syracosphaeraceae (represented in our tree by *Syracosphaera pulchra* and *Coronosphaera mediterranea*) and other members of the order Syracosphaerales and between this order and the Zygodiscales (Pontosphaeraceae and Helicsphaeraceae; see Bown 2005). Molecular phylogenies tend to indicate a more recent link between the Syracosphaeraceae and the Zygodiscales (node 77 of supplementary fig. S1, Supplementary Material online) than can be confidently inferred from stratigraphic studies. *S. pulchra* is used as a default identification for larger fossil *Syracosphaera* coccoliths, so that we adopted a conservatively young date for this node by setting it at ca. 55 Ma.

We followed Sáez et al. (2003) in dating the divergence of *Umbilicosphaera* and *Calcidiscus* (node 63 of supplementary fig. S1, Supplementary Material online) at 24 Ma and set the divergence between *Coccolithus pelagicus* and *Umbilicosphaera hulburtiana* (node 62 of supplementary fig. S1, Supplementary Material online) to 65 Ma. However, Medlin et al. (2008) suggested to use 65 Ma for the divergence of *Coccolithus* and *Cruciplacolithus* (node 61 of supplementary fig. S1, Supplementary Material online). We therefore ran a second set of analyses setting node 61, instead of node 62, to 65 Ma.

Coccoliths assigned to the Pontosphaeraceae (including *Scyphosphaera*) occur down to the late Paleocene, ca. 55 Ma (Bown 2005). Medlin et al. (2008) dated the divergence of the Helicosphaeraceae from the Pontosphaeraceae (node 79 of supplementary fig. S1, Supplementary Material online) at 50 Ma on the basis of interpreting the fossil record of *Helicosphaera* as being continuous down to this date in the early Eocene. However, Aubry et al. (in preparation) postulated that the morphological similarity of coccoliths of *Helicosphaera carteri*, which has a FO ca. 25 Ma, with older species assigned to the Helicosphaeraceae is a result of convergent evolution. In light of this uncertainty, we adopted the younger FO of *H. carteri* (25 Ma) as the CC of this divergence. To assess the impact of this choice on date estimates, we also run an additional set of analyses constraining node 79 to 50 Ma, the older FO of the Helicosphaeraceae (50 Ma).

Results

Times Are Robust to Alignments, CCs, and Data Partitions

The divergence times of the haptophytes were estimated assuming one single partition under the time homogeneous GTR $+ \Gamma_4 +$ I substitution model. Note that with our approach, implemented in BEAST, the phylogeny and the divergence times are jointly estimated.

The resulting phylogeny estimated from the complete alignment is shown in figure 1. Most of the nodes are highly supported, with almost all clade PP $\geq$0.80 and the vast majority $\geq$0.95. The long branches around the root indicate that early divergences have likely been lost to extinction or not sampled. All order-level groups of taxa according to current taxonomy were resolved in this phylogeny, with the early divergence within the Calcihaptophycideae of the orders Isochrysidales and Syracosphaerales receiving the weakest support (PP = 0.85). Of the cases where two species of the same genus were included in the analysis, only *Hymenomonas* proved to be paraphyletic. All nodes used for calibrating the tree with dates from the fossil record were highly supported. The analysis of the alternative alignment without the ambiguous regions resulted in a topology where the only difference was the position of Isochrysidales, which branched with a very low support (PP = 0.63) at a basal position right after the divergence of the Pavlovales. As the Isochrysidales belong to the coccolithophores, the placement of this clade is likely the result of a LBA artifact with the alternative alignment (see section below). In spite of this topological discrepancy between the complete and the alternative alignment, the estimated divergence times are essentially the same irrespective of the alignment used (supplementary fig. S2A, Supplementary Material online). Indeed, the overlapping 95% credibility intervals with the first diagonal (supplementary fig. S2B, Supplementary Material online) suggest that the differences are not significant, except for node 73 that represents the most basal divergence of the Isochrysidales. Because the dating results are robust to the alignment choice, the complete alignment is used throughout the rest of the text.

The influence of the CCs appears to be minimal on time estimates (supplementary tables S2 and S3, Supplementary Material online) as all three series of mean posterior estimates, with 4, 5, or 6 CCs, are highly correlated ($\rho > 0.997$) and, more significantly, marginal probabilities are all with one log-likelihood unit (supplementary table S2, Supplementary Material online). Besides, marginal log-likelihood

values indicate that, although the model with five CCs was the most likely, the difference with the 4- and 6-CC models is not significant (see supplementary table S2, Supplementary Material online). Note that our use of relatively vague priors (table 2) ensures that our results are robust to the disputed use of calibration of node 79 at 25 Ma instead of 50 Ma (supplementary tables S4 and S5, Supplementary Material online), as well as to the potential misidentification of node 62 for node 61 (supplementary tables S6 and S7, Supplementary Material online).

Because supplementary figure S3C (Supplementary Material online) suggests a potential issue with *Undaria pinnatifida*, whose long branch could indicate a misaligned part of the SSU gene (this sequence is actually corrupt and consists essentially of ITS and LSU sequence), we first removed this taxon from our data set and reran the analyses as in table 2 (4, 5, and 6 CCs); second, we also moved the CC at node 61 to node 62 (again, using a total of 4, 5, and 6 CCs). The results show very little difference between the estimated dates when *U. pinnatifida* is removed from the analysis, be it for the model as in table 2 (supplementary fig. S4C and tables S8 and S9, Supplementary Material online) or when node 62 is misidentified for node 61 (supplementary fig. S4D and tables S10 and S11, Supplementary Material online). In spite of the robustness of our date estimates to the potential misalignment of *U. pinnatifida*, we note that deep divergences would be potentially overestimated if *SSU* and *LSU* were analyzed on their own (supplementary fig. S4E, Supplementary Material online), although the 95% credibility intervals are so large (supplementary fig. S5E, Supplementary Material online) that these differences are rarely significant. Because all the dating results are robust to 1) the CCs employed and 2) the inclusion of *U. pinnatifida*, the results with five CCs (fig. 1) as specified in table 2 are those that are used in the rest of this study.

From this analysis, haptophytes were estimated to have diverged from the other eukaryotes included in the analysis 824 Ma (95% highest probability density: 1,017–640—see supplementary table S3, Supplementary Material online) in the mid-Neoproterozoic Cryogenian period. The divergence of the two extant haptophyte classes is estimated to have occurred 543 Ma (823–328) in the early Cambrian. The divergences between the four taxa within the Pavlovophyceae, including representatives of each of the three extant groups within this class defined by Van Lenning et al. (2003), are all estimated to be relatively ancient, occurring in the Mesozoic between 230 and 103 Ma. Within the Prymnesiophyceae, the two noncalcifying orders are estimated to have diverged prior to the Mesozoic, the Phaeocystales at ca. 329 Ma, and then the Prymnesiales at ca. 291 Ma, both in the Carboniferous period. According to our estimates, the primary divergence within the Calcihaptophycideae occurred at 243 Ma, around the Permian/Triassic (P/Tr) boundary event. Molecular divergence within both the Prymnesiales and Calcihaptophycideae is estimated to have occurred throughout the Jurrassic, Cretaceous, and Tertiary periods, with 10 calcihaptophyte lineages (from the 24 species in this clade as included in our analysis) crossing the K/T boundary, representing a much weaker signal of divergence occurring predominately after the K/T boundary than reported by Medlin et al. (2008).

Because the model used above makes the simplifying assumption that the data can be concatenated, we need to test that our results are not sensitive to the data partitioning scheme or affected by LBA artifacts.

Robustness of the Phylogeny to Data Partitioning

Our first simplifying assumption was that we could combine all four genes into one single partition without affecting the estimated tree. The two partitioning strategies compared were as follows: 1) one single partition where the two RNA genes and the two protein-coding genes were concatenated and 2) one partition for each RNA gene plus one partition for each codon position of each protein-coding gene, which amounts to a parameter-rich model with a total of eight partitions. The substitution model selected by Modeltest with the Akaike information criterion was GTR $+ \Gamma_4 +$ I for all data sets.

Supplementary figure S3A and B (Supplementary Material online) shows the trees obtained under these two partitioning schemes. In both cases, the Pavlovophyceae were resolved and branched off first. Within the Prymnesiophyceae, intermediate branching orders (Phaeocystales/Prymnesiales/Calcihaptophycideae) were identical, even though these were the least well-supported nodes of each tree (PP slightly less than 0.80 or $\in$ [0.80,0.90]). Note also that in both partitioning schemes, the genus *Hymenomonas* is paraphyletic with high PP ($=1$). The main differences between the two trees (supplementary fig. S3A and B, Supplementary Material online) occur within the coccolithophores, notably in resolving the early branching between the Isochrysidales and Syracosphaerales as well as in the exact position of two (out of 24) taxa, *Algirosphaera robusta* and to a lesser extent *Cruciplacolithus neohelis*. Note that this "MrBayes tree" with one single partition is not significantly different from the "BEAST tree" estimated above (SH: $P = 0.208$).

The initial motivation for partitioning was to account for the fact that some of the genes or partitions evolve much faster than others. Indeed, the relative rates of the different partitions, as estimated by maximum likelihood with PAML, are as follows: 1.00 (*LSU*); 0.37 (*SSU*); 0.25 (*rbcL*, codon position 1); 0.07 (*rbcL*, codon position 2); 1.54 (*rbcL*, codon position 3); 0.48 (*tufA*, codon position 1); 0.17 (*tufA*, codon position 2); and 119.44 (*tufA*, codon position 3). The third codon positions are therefore likely to be noisy. In spite of all these differences and potential issues about noise, the two trees are not significantly different at the 1% level (SH: $P = 0.196$), which suggests that the data can be analyzed under the simplest model that contains one single partition without significantly affecting the reconstructed tree. However, the more appropriate computation of marginal probabilities, m, suggests that the more complex model (m[8 partitions] $= -39{,}168$) outperforms the simpler model (m[1 partition] $= -40{,}553$). In spite of the inaccuracy of these m estimates (Lartillot and Philippe

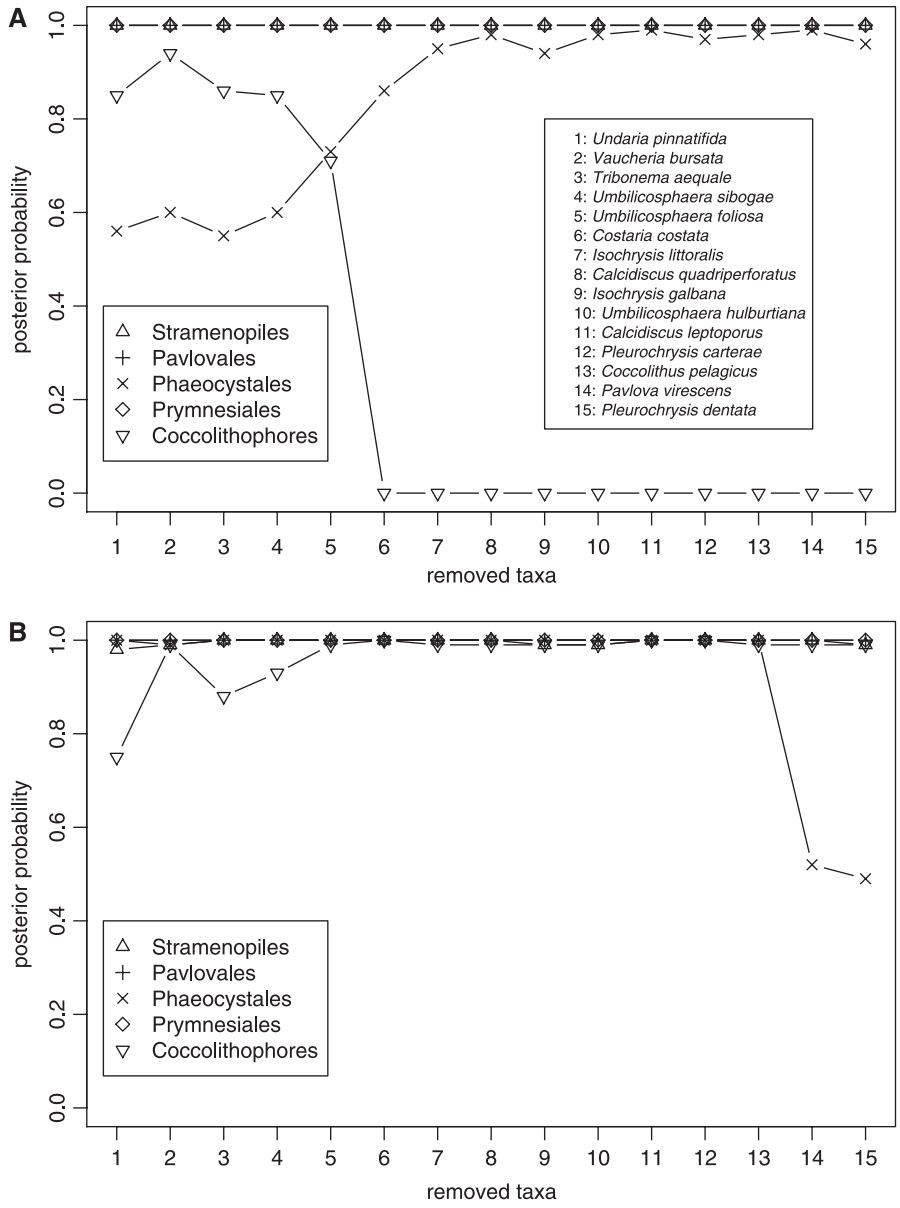

Fig. 2. Long-branch analysis by taxon removal based on the concatenated alignments: PP of monophyly for (*A*) time-independent analyses (MrBayes) and (*B*) time-dependent analyses (BEAST; monophyly not enforced; all six CCs placed and set as in table 2). The insert in (*A*) defines the species identifiers used on the *x* axis to represent the taxa that were sequentially removed. For the Phaeocystales, the plotted values represent the clade PP of *Phaeocystis* sp.

2006), these latter results suggest that a stability analysis of the trees and of the estimated divergence times is required.

Robustness of the Phylogeny to LBA

To assess the stability of the reconstructed tree (fig. 1), in particular with respect to LBA artifacts caused by model misspecification, we successively removed the taxa that showed the longest root-to-tip branch lengths (Brinkmann et al. 2005; Hampl et al. 2009). Two approaches were used. In the first approach, tree topologies were completely unconstrained in the sense that the time dependency of the evolutionary process was not taken into account. With this approach, estimated PPs for the stramenopiles, the Pavlovales, and the Prymnesiales were unaffected and consistently close to 1 (fig. 2*A*), suggesting that LBA is not an issue for these groups. On the other hand, progressive taxon removal changed the position of Phaeocystales from being sister to the Prymnesiales and the coccolithophores to being sister to the Prymnesiales with high PP, although support for the coccolithophores collapsed completely (fig. 2*A*) due to the position of Isochrysidales. This suggests that the position of Phaeocystales as sister to Prymnesiales and coccolithophores, as in figure 1, is probably the result of an LBA artifact. However, the effect of LBA on the nonmonophyly of coccolithophores is quite intriguing, as their monophyly has repeatedly been supported by previous studies (e.g., Edvardsen et al. 2000; Fujiwara et al. 2001; de Vargas et al. 2007; Medlin et al. 2008).

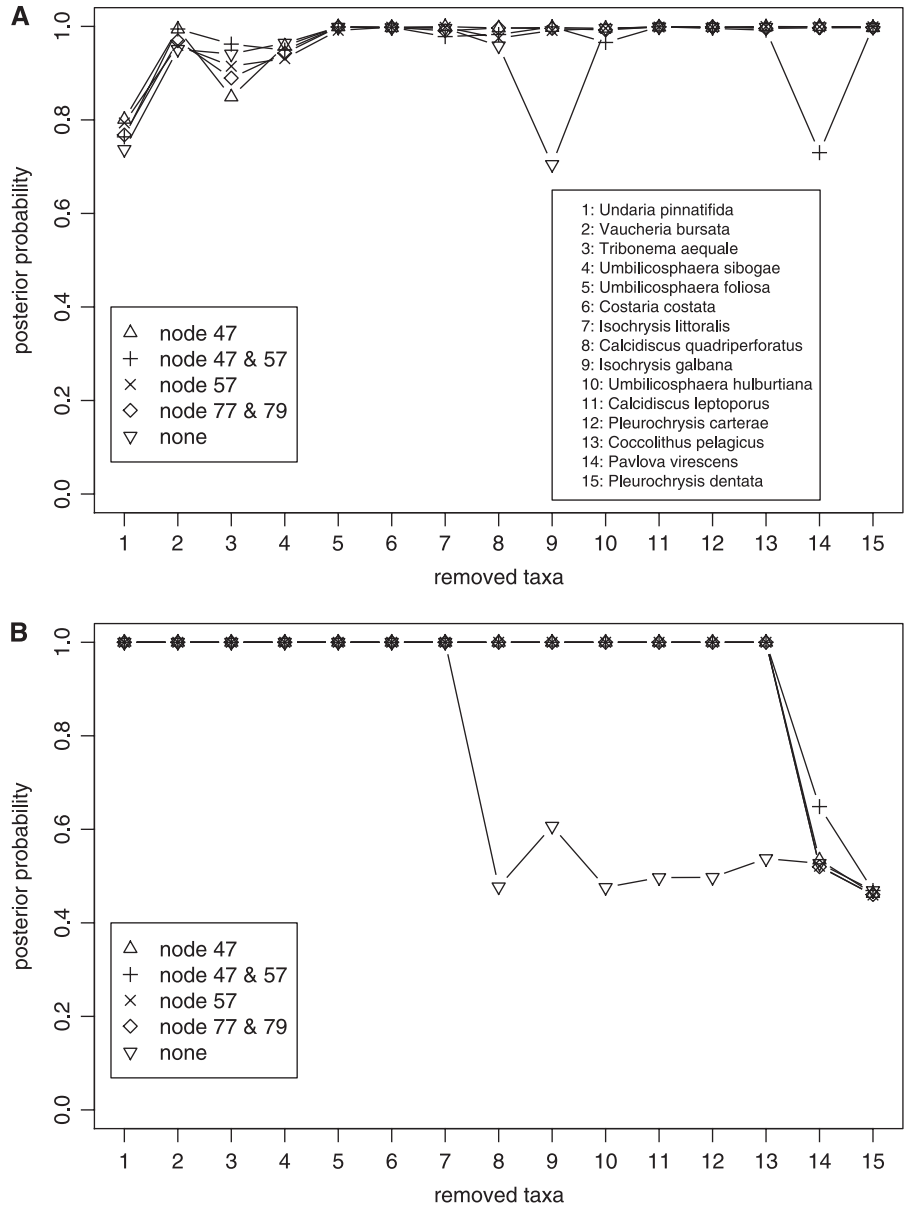

Fig. 3. Effect of the choice of CCs on the LBA analysis by CC removal in time-dependent analyses (BEAST): (*A*) PP of the monophyly of Coccolithophores and (*B*) PP of the Phaeocystales. The insert in (*A*) defines the species identifiers used on the *x* axis to represent the taxa that were sequentially removed.

As noted above, one very general cause of LBA is model misspecification, which is known to impact PP (e.g., Buckley 2002; Lemmon and Moriarty 2004; Yang and Rannala 2005). In particular, unconstrained analyses as performed above with MrBayes implicitly assume that the tree topology provides no information about relative branch lengths. Drummond et al. (2006) suggested that modeling the time dependency of the evolutionary process should improve tree reconstruction. To assess this proposition here, we incorporated time dependency by setting CCs in the taxon removal analysis. BEAST was used as described in the Materials and methods; in particular, the monophyly of the different groups was not enforced. The results (fig. 2*B*) show that all five groups studied here are monophyletic and highly stable to the exception of Phaeocystales that exhibit signs of LBA and tend to become sister to the Prymnesiales only when >13 taxa are removed from the analysis. Therefore, the enforcement of time dependency stabilizes the tree reconstruction process, minimizing the impact of highly divergent taxa, and thereby appears to remove most LBA artifacts from the analysis.

In the face of this result, it is desirable to know whether a particular CC or set of constraints has a major stabilizing effect or if the mitigation of LBA is mainly due to the time dependency structure of the model. To address this question, we reran the taxon removal analyses with select CCs (node 47, nodes 47 and 57, node 57, or nodes 77 and 79) or only with the extremely diffuse prior on the root (see table 2). To simplify the presentation of the results, we focus on the two clades that showed evidence for LBA in the unconstrained analysis: the coccolithophores and the Phaeocystales. Figure 3*A* shows that in the case of

the coccolithophores, the introduction of time dependency into the model is solely responsible for the stabilizing effect, irrespective of the CCs used. On the other hand, PP stabilization for Phaeocystales depends on the CCs included: when no constraints other than the vague root prior are incorporated into the model, LBA apparently disappears when a small number of taxa (8) are removed. Alternatively, in the presence of (internal) CCs, LBA removal requires that more taxa be removed (fig. 3*B*). Therefore, the introduction of time dependency into a model might help mitigate some of the LBA artifacts but is not eliminating them all.

Both Nuclear and Plastid Genes Share the Same Phylogeny

Our fourth intrinsic assumption was that both the nuclear and the plastid genes share the same history. This need not be so as endosymbiotic gene transfers postdating the divergence of haptophytes could have affected the history of these genomes.

Supplementary figure S3C and D (Supplementary Material online) shows the trees obtained for the nuclear and for the plastid genes, respectively. Note that the branch lengths were longer for the plastid tree, reflecting the fact that the protein-coding genes evolve much faster than the RNA genes (see above). Some differences were observed in the relative positions of certain taxa within the coccolithophore clade, notably for *A. robusta*, and PPs were generally lower within this clade in the plastid gene tree. In spite of these differences, the nuclear and the plastid trees were not significantly different at the 1% level (SH: $P = 0.020$). Again, this test might not be the most appropriate in the context of hypotheses derived from Bayesian analyses, but it nonetheless indicates that 1) there is no strong evidence that the phylogenetic signal has been blurred by horizontal gene transfer, endosymbiotic gene transfer (or replacement), or by a "tertiary transfer" from which the haptophytes would have received their plastid (Hackett et al. 2007) and 2) the data can be analyzed under the simplest model that contains one single partition without significantly affecting the reconstructed tree.

Robustness of the Phylogeny to the Evolutionary Process

Our last assumption was that the evolutionary processes assumed here are time homogeneous, that is, do not change in time across the different lineages.

Supplementary figure S6 (Supplementary Material online) shows the trees estimated under a rate-across-site model (CAT, panel A), a rate-across-lineage model (BP, panel B), and a rate-across-site and lineage model (CAT-BP, panel C). Under the GTR + Γ_5 substitution model, the best (maximum likelihood) tree was the one estimated under CAT-BP, and the two other trees were not significantly different from this one at the 1% level (SH test: $P_{CAT} = 0.415$; $P_{BP} = 0.549$). This result suggests that it is safe to ignore spatiotemporal variation of substitution rates and that saturation is not an issue.

Reconstruction of Ancestral Characters and of Paleoenvironments

Because the phylogeny obtained is relatively well supported (fig. 1), a simple maximum likelihood reconstruction of ancestral characters is appropriate and does not require integrating over topological uncertainty. Our reconstruction of the evolution of the ability to calcify is represented in figure 4*a*. Calcification is shared by most coccolithophores and has clearly been secondarily lost in the genus *Isochrysis*. The model predicts that whereas the ability to calcify had not evolved in the earliest haptophyte (with a probability of 0.828), the most recent common ancestor of the Calcihaptophycideae and Prymnesiales (node 52 of supplementary fig. S1, Supplementary Material online) has a 0.815 probability of having possessed the ability to calcify. Intracellular calcification may thus have evolved early before the divergence of the Calcihaptophycideae and Prymnesiales (between 329 and 291 Ma). Calcification was later lost twice along the branches leading to 1) the Prymnesiales (between 291 and 171 Ma) and 2) the Isochrysidaceae (between 119 and 37 Ma).

The ability to calcify required the presence of organic plate scales, but these scales were probably not a sufficient condition. Figure 4*B* shows that the cenancestor of the haptophytes had a high probability (0.906) of possessing organic plate scales. This suggests that knob scales evolved on the branch leading to the Pavlovales, that is, between 543 and 230 Ma.

Similarly, the model indicates that the cenancestor of the haptophytes inhabited a coastal environment (probability 1.000; fig. 2*C*). The cenancestor of the Prymnesiales may not have left coastal environments ($P = 0.645$), in which case only one transition toward oceanic environments occurred, presumably after the divergence of coccolithophores and Prymnesiales, between 291 and 243 Ma around the time of the P/Tr boundary event (251 Ma). Some taxa then returned to a coastal environment: first with the most recent common ancestor of the Hymenomonadaceae and Pleurochrysidaceae (between 181 and 130 Ma) and a second time, independently, prior to the divergence of the Isochrysidaceae (between 119 and 37 Ma).

Finally, figure 4*D* shows the reconstructed trophic modes and suggests that autotrophy evolved from mixotrophy probably only once ($P = 0.677$) by losing the phagocytic ability around the split Phaeocystales/Prymnesiales between 329 and 291 Ma. Our model suggests that mixotrophy was then regained along the branch leading to the Prymnesiales between 291 and 171 Ma.

Discussion

Molecular clocks represent a potentially powerful tool for generating insights into the major events in the evolutionary history of groups of organisms, provided they are applied and interpreted with appropriate caution. Here we attempt to further develop these insights by combining

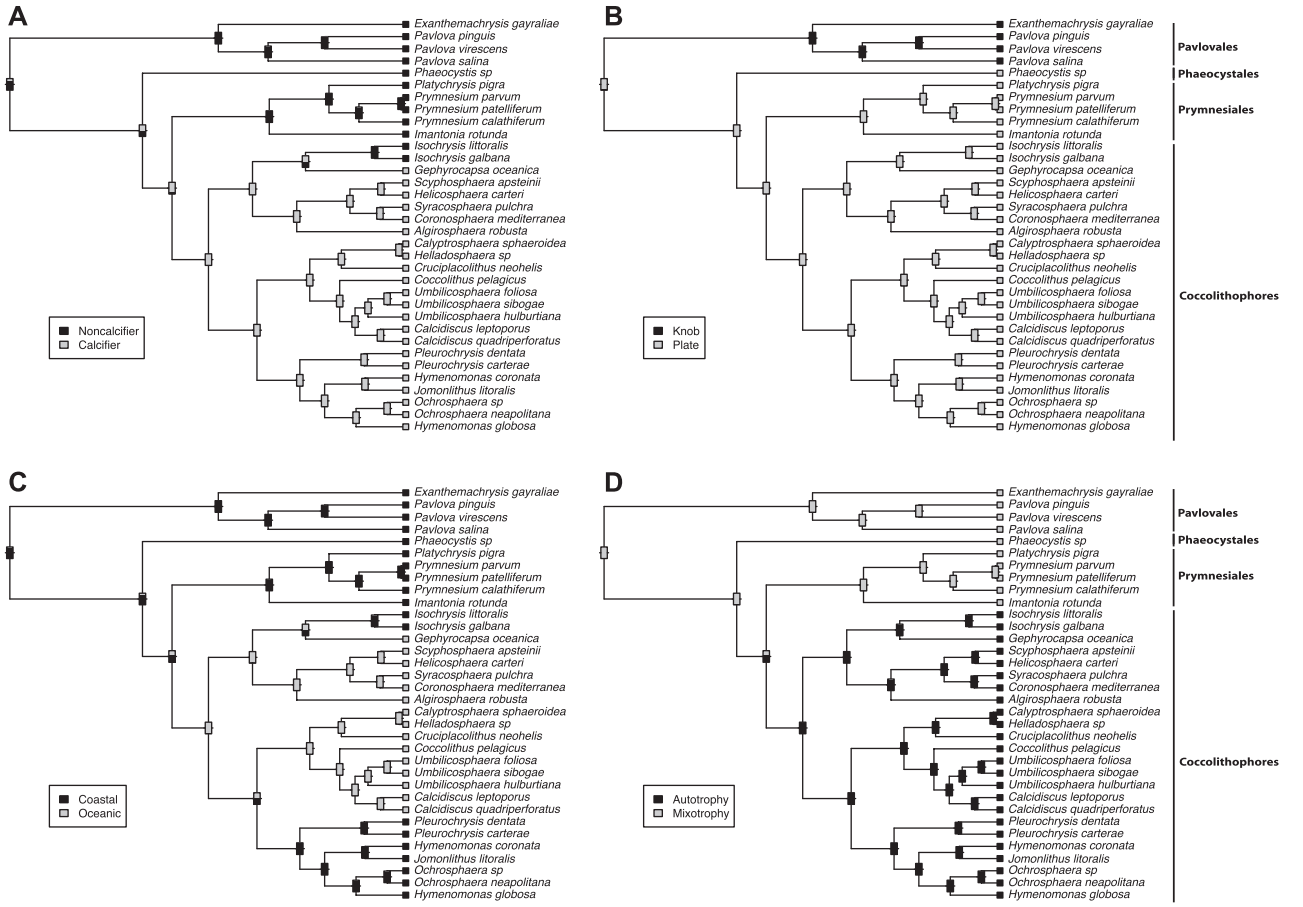

FIG. 4. Maximum likelihood reconstructions of the paleoecology of the haptophytes: (*A*) calcification, (*B*) type of organic plates, (*C*) marine environment, and (*D*) trophic mode. Box shading at leaf nodes indicates the observed states. Box shading at internal nodes represents the relative probabilities of the different pairs of states.

a relaxed molecular clock analysis based on a statistically sound Bayesian phylogenetic reconstruction of the haptophytes, with a maximum likelihood reconstruction of probable ancestral character states. The resulting estimate of the time line of phenotypic and ecological evolution in this important group of photosynthetic protists can then be interpreted in a geological context.

Phylogeny of the Haptophytes

Our extensive analysis of a large four-gene haptophyte data set indicates that the phylogenetic tree shown in figure 1 is robust to 1) data partitioning, 2) the genomic origin of the genes (host vs. endosymbiont), 3) across-site rate variation, and 4) across-lineage rate variation. Overall, this haptophyte tree is highly consistent with existing taxonomic schemes and with previous molecular phylogenies, notably the single gene Haptophyta phylogenies presented by Medlin et al. (2008). Although their analysis included almost twice the number of taxa, the data generated and assembled here have a comparable taxonomic range and underwent a more comprehensive analysis.

The order of the early divergences within the Calcihaptophycideae remains the most contentious part of the molecular phylogeny of the Haptophyta, but our extensive analyses suggest that the results based on models that include time dependency are robust to LBA with the complete alignment. The earliest divergence within the Calcihaptophycideae has most often placed the Isochrysidales as a sister group to all other coccolithophore orders (Edvardsen et al. 2000; de Vargas et al. 2007; *TufA* phylogeny of Medlin et al. 2008). This scenario would imply that holococcolith, a structure that results from a unique calcification process in haploid phase cells and that is not present in extant Isochrysidales, evolved after this divergence along the branch leading to all other coccolithophores (Syracosphaerales/Zygodiscales/Coccolithales). The placement of the divergence of the nannolith-bearing genus *Braarudosphaera* as basal to the entire calcihaptophyte clade by Takano et al. (2006) would not alter this interpretation because this genus is not known to produce a holococcolith-bearing haploid phase. In contrast, our consensus tree gives the earliest divergence within the Calcihaptophycideae by placing the Coccolithales as a sister group to a clade that includes Isochrysidales and Syracosphaerales + Zygodiscales. The *SSU* phylogeny of Medlin et al. (2008) also groups *Coronosphaera*, a genus probably affiliated to the Syracosphaerales, with the Isochrysidales. Our results and those of Medlin et al. (2008) would therefore both support an early

origin of holococcoliths. This is consistent with a number of observations: 1) a number of cytological features of the Isochrysidales, such as the structure of plate scales (when present) or that of flagellar roots, are relatively simple; 2) although the oldest fossil holococcoliths date back to ca. 185 Ma (Bown 1998) instead of the expected 220 Ma under the scenario of an early divergence of Isochrysidales, holococcoliths are more fragile and significantly less well preserved in the fossil record than heterococcoliths, so that an earlier origin cannot be ruled out on the grounds of absence of fossils; 3) secondary loss of holococcoliths is known to have occurred in a clade of coastal Coccolithales species (Young et al. 2005), so that an additional secondary loss in the Isochrysidales is conceivable as this clade possesses a number of other lost features such as the haptonema. Because holococcoliths are formed by a complex calcification process that is quite unlikely to have evolved more than once (Young et al. 1999), our result of an early origin of calcification with a secondary loss of holococcoliths in Isochrysidales (fig. 4*A*) is reasonable.

Timing of the Diversification of the Haptophytes

The reliability of molecular clock estimates is obviously related to the accuracy of fossil CCs. In this context, the fossil record of coccolithophores is unique in providing many well-documented fossil dates. Ongoing work on their biostratigraphy is likely to continually refine the accuracy of fossil dates in the future. In the current context, however, our analysis proved to be robust to the specification of CCs, as the removal or the addition of a constraint to the five that were employed did not affect the estimated dates significantly (supplementary tables S3, S5, and S7, Supplementary Material online).

Because in practice fossil dates are defined with varying levels of accuracy, there is often a trade-off in choosing the number of constraints to be employed in a molecular clock analysis between the degree of coverage (of the phylogeny and the time range) and the degree of confidence in constrained dates. The Bayesian modeling adopted here elegantly accommodates both sides of the trade-off, first with an increased coverage by employing more CCs within the coccolithophore clade than any previous study (de Vargas et al. 2007; Medlin et al. 2008) and second with the placement of relatively vague prior distributions on these CCs. As the taxonomic range of multigene data sets of the coccolithophores increases, notably within the undersampled Syracosphaerales and Zygodiscales clades, a number of additional fossil dates known with relative confidence could be used to calibrate relaxed molecular clocks to extend our work and further test our results.

Reciprocally, the predictive power of our analysis can be assessed by checking congruence between our time estimates and nodes that are unconstrained in our analysis but for which fossil dates do exist. One such example of congruence with an older, character-based constraint is the date for the origin of alkenones. These are distinctive lipids produced exclusively by members of the Isochrysidales and best known for their use as paleoindicators of surface water temperature (Brassell et al. 1986; Prahl and Wakeham 1987; Conte et al. 1998). The known geological record of alkenones extends down to the mid-Cretaceous at ca. 120 Ma (Brassell and Dumitrescu 2004), a FO that is not very well constrained and could conceivably have a much earlier age (Medlin et al. 2008). Because alkenones are produced by all members of the Isochrysidales, they must have evolved some time between the divergence of this order from other coccolithophores and the first divergence within the order, a range that we estimated at 203–119 Ma. Given that molecular divergences always predate morphological divergences, there is quite a remarkable degree of congruence between our time estimates and external fossil dates in this very particular example. On existing evidence, an interpretation would be that alkenones evolved shortly before the crown divergence of species within this order, although in the absence of information relating to the function of alkenones there is no reason to believe that the two events were causally linked.

Despite overall consistency of the reconstructed phylogeny with previous studies and the use of comparable fossil constraints, our results did depart from those of previous studies, in particular with respect to the estimated times of early divergences within the Haptophyta division. First, the divergence between stramenopiles and the haptophytes was here dated ca. 824 Ma (1,031–637), which is significantly more recent than the 1,100 Ma average often estimated (e.g., Medlin et al. 1997, 2008; Yoon et al. 2004). Second, we dated the divergence of the Pavlovales at 543 Ma (823–328), whereas Medlin et al. (2008) estimated it around 800 Ma. Third, the divergence of the Phaeocystales, which Medlin et al. (2008) estimated at ca. 490 Ma, was dated at 329 Ma (428–248) in our analysis. In each of these cases, our divergence time estimates were younger and significantly so in 2 cases out of 3 than previous molecular estimates reported in the literature. Several factors could explain these differences.

First, we used a Bayesian approach that integrates over all uncertainties of the model parameters, including the tree topology. However, because the estimated topology proved to be highly supported, this factor is unlikely to explain the important difference in time estimates compared with previous studies. Second, previous studies were mainly based on the slowly evolving *SSU* RNA gene. The incorporation of all three codon positions of protein-coding genes is often criticized for incorporating noise into phylogenetic analyses, but a theoretical study suggests otherwise (Seo and Kishino 2008). Therefore, it might be important to incorporate these rapidly evolving positions to help discriminate otherwise poorly resolved nodes and dates. Again, because most of the nodes were well resolved, incorporating rapidly evolving genes is unlikely to explain completely why our time estimates are younger than previous molecular studies. Third, we used a relaxed uncorrelated clock model to estimate divergence times. Simulation studies show that this class of models outperforms all other dating methodologies (Aris-Brosou 2007), including the penalized likelihood approaches (Sanderson 2002, 2003) that have

been used in most previous molecular studies (Berney and Pawlowski 2006). Finally, note that relaxed clock models are specifically designed to account for among-lineage rate variation either by means of an autocorrelated process (Sanderson 2002, 2003) or like here with an uncorrelated process (Drummond et al. 2006). Therefore, the inclusion of genes that exhibit "non–clock-like" behavior such as *tufA* (Medlin et al. 2008) is unlikely to affect our analysis, as these models account for these nonlinear effects (Aris-Brosou 2007).

The phylogenetic relationships between haptophytes and other chromalveolate lineages remain unresolved, but there is general consensus that the crown divergence leading to modern lineages occurred early in the history of this supergroup some few hundred My after the origin of the eukaryotes (e.g., Cavalier-Smith 2006). A recent molecular study (Berney and Pawlowski 2006) dated the cenancestral eukaryote at ca. 1,126 (948–1,357) Ma and the origin of haptophytes at slightly later than 900 Ma. Considering that "Bayesian algorithms" can miss rapid rate variation (but see Aris-Brosou 2007), Cavalier-Smith (2006) proposed that eukaryotes originated 900 $\pm$ 100 Ma, with chloroplasts and Plantae evolving between 570 and 850 Ma, and chromalveolates, opisthokonts, Rhizaria, and excavates originating "ca. 570 Ma" when the Proterozoic snowball Earth melted. Our clock estimate for the origin of haptophytes at 824 (1,031–637) Ma is not consistent with this theory that chromalveolates originated shortly before the Cambrian explosion.

Cavalier-Smith (2006) also argues for the sudden origin of many phyla near the Precambrian boundary followed by the origin of novel classes and/or orders in the early Mesozoic and early Cenozoic, presumably by exploiting niches or whole adaptive zones emptied by the greatest mass extinctions. Our estimated date for the divergence of the two extant haptophyte classes puts this event near Precambrian boundary (543 Ma). Our analysis indicates that the primary radiation within the Prymnesiophyceae (the divergence of Phaeocystales from other prymnesiophytes) occurred in the Carboniferous period, a considerable time before the P/Tr boundary event (end of the Paleozoic/start of the Mesozoic). Alternatively, widespread shallow epicontinental seas persisted throughout much of the Carboniferous, a period that was preceded by an important extinction event at the end of the Devonian. Our estimates for the timing of the next two divergences within the Prymnesiophyceae, the divergence of the Prymnesiales and the primary radiation of the Calcihaptophycideae both coincide with important Earth system transitions early in the Permian and the Triassic, respectively.

Timing of the Environmental Adaptations of the Haptophytes

Our reconstructions of four key evolutionary transitions (calcification, nature of organic scales, oceanic environment, and trophic mode) suggest that calcification evolved along the same lineage where the phagocytic ability was lost, just after the divergence of the Phaeocystales (ca. 230 Ma). Although conceivably calcification might hamper predation, a strict dependence or ecological link between these two transitions does not seem to be supported by the situation in *Isochrysis*, a genus that has lost its calcification capacity without regaining mixotrophy.

Likewise, the transition to an oceanic environment took place after the two previous transitions to calcification and full autotrophy, ca. 230–172 Ma. Yet, it is difficult to argue that both calcification and autotrophy were prerequisites for the transition to an oceanic environment. Indeed, the two clades that returned to coastal environments either remained autotrophic calcifiers (*Hymenomonadaceae* and *Pleurochrysidaceae*) or never completely lost the phagocytic ability in the first place and lost the ability to calcify (Prymnesiales). As a result, calcification, trophic mode, and transitions to an oceanic environment seem to be mutually independent transitions. This result is consistent with previous studies that suggested that the interaction between calcification and photosynthesis may not be direct (Brownlee and Taylor 2004).

Explaining the origin of algal plastids continues to be a major challenge in evolutionary biology (Yoon et al. 2002). As the most recent common ancestor of all haptophytes was most probably mixotrophic, photosynthesis must have evolved before the origin of this group, that is, before 824 Ma. This time interval is fully consistent with the 1,072–767 Ma time window estimated by Douzery et al. (2004) with 129 proteins from 34 eukaryotes and supports their claim that this secondary event occurred shortly after the primary endosymbiosis. However, these time windows are extremely wide, so that it is difficult to argue about the exact point in time when a red algal plastid was acquired by secondary endosymbiotic event (*contra* Medlin et al. 1997). Consequently, we cannot rule out the possibility that the haptophytes led an early heterotrophic life before acquiring their plastid. Only the inclusion of an early-branching organism to break up the long branch that leads to haptophytes or the evidence that the haptophyte plastid is shared with another lineage such as the Cryptophytes, as suggested by Rice and Palmer (2006) from the existence of a bacterial gene in both plastids, can help resolve this issue.

Our reconstruction of the emergence of organic scales supports an alternative scenario to what was previously proposed (de Vargas et al. 2007). Indeed, our results show that the proto-haptophytes may have already evolved the ability to produce organic plate scales and that plate scales reverted to a simple and less elaborate knoblike scale on the branch leading to the Pavlovales. Therefore, the ability to control the intracellular precipitation of calcite on the plate scale most likely emerged in the prymnesiophyte ancestor of the coccolithophores.

Conclusions

We have presented the first robust and extensive phylogeny of the haptophytes. Our results are consistent with previous work based on morphology (Young et al. 1999) or on the fossil record (Bown et al. 2004). Although we dated the most recent common ancestor of calcifying haptophytes to

243 Ma, our analyses suggest that the ability to calcify evolved much earlier than this, probably between 329 and 243 Ma, in the Carboniferous/early Triassic. As this innovation was shortly followed by the transition of these organisms to the global ocean in the P/Tr, our results imply that global carbon cycles were probably impacted by the haptophytes much earlier than previously thought (Ridgwell and Zeebe 2005; Fabry 2008).

Supplementary Material

Supplementary tables and figures are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Funding

US National Science Foundation (DEB-0415351); Actions Thématiques et Incitatives sur Programme fellowship from the "Centre National de la Recherche Scientifique" to C.d.V.; Canadian Natural Sciences and Engineering Research Council to S.A.B.; Canada Foundation for Innovation to S.A.B.; Initiation of Research & New Direction program at the University of Ottawa to S.A.B.; "Institut Français de la Biodiversité" via the "Agence Nationale de la Recherche" (ANR-05-BDIV-004 to C.d.V.).

Acknowledgments

We thank Swati Narayan-Yadav for her help in obtaining the sequences of known haptophytes. Comments from our handling editor, Andrew Roger, as well as from two anonymous reviewers substantially improved the quality of the manuscript. It is part of the pluridisciplinary project Biodiversity of Open Ocean Microcalcifiers.

References

Aris-Brosou S. 2007. Dating phylogenies with hybrid local molecular clocks. *PLoS ONE*. 2:e879.

Berney C, Pawlowski J. 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc Biol Sci*. 273:1867–1872.

Blanquart S, Lartillot N. 2008. A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol*. 25:842–858.

Bown P. 2005. Selective calcareous nannoplankton survivorship at the Cretaceous-Tertiary boundary. *Geology*. 33:653–656.

Bown PR. 1998. Calcareous nannofossil biostratigraphy. Chapman & Hall.

Bown PR, Lees JA, Young JR. 2004. Calcareous nannoplankton evolution and diversity through time. In: Thierstein HR, Young JR, editors. Coccolithophores: from molecular processes to global impact. Springer Verlag, Berlin (Germany). p. 427–554.

Brassell SC, Dumitrescu M. 2004. Recognition of alkenones in a lower Aptian porcellanite from the west-central Pacific. *Org Geochem*. 35:181–188.

Brassell SC, Eglinton G, Marlowe IT, Pflaumann U, Sarnthein M. 1986. Molecular stratigraphy: a new tool for climatic assessment. *Nature*. 320:129–133.

Brinkmann H, van der Giezen M, Zhou Y, Poncelin de Raucourt G, Philippe H. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst Biol*. 54:743–757.

Brown CW, Yodar JA. 1994. Coccolithophorid blooms in the global ocean. *J Geophys Res*. 99:7467–7482.

Brownlee C, Taylor A. 2004. Calcification in coccolithophores: a cellular perspective. In: Thierstein HR, Young JR, editors. Coccolithophores: from molecular processes to global impact. Springer-Verlag, Berlin (Germany). p. 31–49.

Buckley TR. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Syst Biol*. 51:509–523.

Cavalier-Smith T. 2004. Chromalveolate diversity and cell megaevolution: interplay of membranes, genomes and cytoskeleton. In: Hirt RP, Horner DS, editors. Organelles, genomes and eukaryotic phylogeny. Taylor and Francis, London. p. 75–108.

Cavalier-Smith T. 2006. Cell evolution and Earth history: stasis and revolution. *Philos Trans R Soc Lond B Biol Sci*. 361:969–1006.

Conte MH, Thompson A, Lesley D, Harris RP. 1998. Genetic and physiological influences on the alkenone/alkenoate versus growth temperature relationship in *Emiliania huxleyi* and *Gephyrocapsa* oceanica. *Geochim Cosmochim Acta*. 62:51–68.

Daugbjerg N, Andersen RA. 1997. Phylogenetic analyses of the *rbcL* sequences from haptophytes and heterokont algae suggest their chloroplasts are unrelated. *Mol Biol Evol*. 14:1242–1251.

de Vargas C, Aubry MP, Probert I, Young J. 2007. The origin and evolution of Coccolithophores: from coastal hunters to oceanic farmers. In: Falkowski PG, Knoll AH, editors. Evolution of primary producers in the sea. Elsevier Academic Press, New York. p. 251–286.

Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci USA*. 101:15386–15391.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 4:e88.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 7:214.

Edvardsen B, Eikrem W, Green JC, Andersen RA, Moon-van der Staay SY, Medlin L. 2000. Phylogenetic reconstructions of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data. *Phycologia*. 39:19–35.

Edvardsen B, Paasche E. 1998. Bloom dynamics and physiology of Prymnesium and Chrysochromulina. In: Anderson DM, Cembella AD, Hallegraeff GM, editors. Physiological ecology of harmful algal blooms. NATO ASI Series G. Springer-Verlag, Heidelberg (Germany). p. 193–208.

Fabry VJ. 2008. Ocean science: marine calcifiers in a high-CO_2 ocean. *Science*. 320:1020–1022.

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

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*. 281:237–240.

Fujiwara S, Sawada M, Someya J, Minaka N, Kawachi M, Inouye I. 1994. Molecular phylogenetic analysis of *rbcL* in the Prymnesiophyta. *J Phycol*. 30:863–871.

Fujiwara S, Tsuzuki M, Kawachi M, Minaka N, Inouye I. 2001. Molecular phylogeny of the Haptophyta based on the *rbcL* gene and sequence variation in the spacer region of the RUBISCO operon. *J Phycol*. 37:121–129.

Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D. 2007. Phylogenomic analysis supports the monophyly of Cryptophytes and Haptophytes and the association of Rhizaria with Chromalveolates. *Mol Biol Evol*. 24:1702–1713.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009. Phylogenomic analyses support the monophyly

of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci USA*. 106:3859–3864.
Harper JT, Waanders E, Keeling PJ. 2005. On the monophyly of Chromalveolates using a six-protein phylogeny of eukaryotes. *Int J Syst Evol Microbiol*. 55:487–496.
Inouye I. 1997. Systematics of haptophyte algae in Asia-Pacific waters. *Algae (Kor J Phycol)*. 12:247–261.
Jordan RW, Cros L, Young JR. 2004. A revised classification scheme for living haptophytes. *Micropaleontology*. 50:55–79.
Kawachi M, Inouye I, Maeda O, Chihara M. 1991. The haptonema as a food-capturing device: observations on *Chrysochromulina hirta* (Prymnesiophyceae). *Phycologia*. 30:563–573.
Lancelot C, Keller MD, Rousseau V, Smith WO. 1998. Autecology of the marine haptophyte *Phaeocystis sp*. In: Anderson DM, Cembella AD, Hallegraeff GM, editors. Physiological ecology of harmful algal blooms. NATO-ASI series 41. Springer, Berlin (Germany). p. 209–224.
Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol Evol*. 23:268–275.
Larsen N, Olsen GJ, Maidak BL, McCaughey MJ, Overbeek R, Macke TJ, Marsh TL, Woese CR. 1993. The ribosomal database project. *Nucleic Acids Res*. 21:3021–3023.
Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol*. 7:S4.
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 21:1095–1109.
Lartillot N, Philippe H. 2006. Computing Bayes factors using thermodynamic integration. *Syst Biol*. 55:195–207.
Lemmon AR, Moriarty EC. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Syst Biol*. 53:265–277.
Medlin LK, Kooistra WHCF, Potter D, Saanders G, Wandersen RA. 1997. Phylogenetic relationships of the 'golden algae' (haptophytes, heterokont chromophytes) and their plastids. Bhattacharya D, editor. The origin of the algae and their plastids (plant systematics and evolution supplementa). New York: Springer. p. 187–219.
Medlin LK, Sáez AG, Young JR. 2008. A molecular clock for coccolithophores and implications for selectivity of phytoplankton extinctions across the K/T boundary. *Mar Micropaleontol*. 67:69–86.
Milliman JD. 1993. Production and accumulation of calcium carbonate in the ocean: budget of a nonsteady state. *Global Biogeochem Cycles*. 7:927–957.
Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst Biol*. 53:673–684.
Paradis E. 2006. Analysis of phylogenetics and evolution with R. Springer, New York.
Patron NJ, Inagaki Y, Keeling PJ. 2007. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol*. 17:887–891.
Perch-Nielsen K. 1985. Cenozoic calcareous nannofossils. In: Bolli HM, Saunders JB, Perch-Nielsen K, editors. Plankton stratigraphy. Cambridge University Press, Cambridge (UK). p. 427–555.
Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics*. 14:817–818.
Prahl FG, Wakeham SG. 1987. Calibration of unsaturation patterns in long-chain ketone compositions for palaeotemperature assessment. *Nature*. 330:367–369.
Probert I, Houdan A. 2004. The laboratory culture of coccolithophores. In: Thierstein HR, Young JR, editors. Coccolithophores: from molecular processes to global impact. Springer Verlag, Berlin (Germany). p. 217–250.
Rice DW, Palmer JD. 2006. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol*. 4:31.
Ridgwell A, Zeebe RE. 2005. The role of the global carbonate cycle in the regulation and evolution of the Earth system. *Earth Planet Sci Lett*. 234:299–315.
Robertson JE, Robinson C, Turner DR, Holligan P, Watson AJ, Boyd P, Fernandez E, Finch M. 1994. The impact of a coccolithophore bloom on oceanic carbon uptake in the northeast Atlantic during summer 1991. *Deep Sea Res Part I*. 41:297–314.
Ronquist F, Huelsenbeck J. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–1574.
Sáez AG, Probert I, Quinn P, Young JR, Geisen M, Medlin LK. 2003. Pseudocryptic speciation in coccolithophores. *Proc Natl Acad Sci USA*. 100:7163–7168.
Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol*. 19:101–109.
Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. 19:301–302.
Seo T-K, Kishino H. 2008. Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. *Syst Biol*. 57:367–377.
Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol*. 16:1114–1116.
Simon N, Brenner J, Edvardsen B, Medlin LK. 1997. The identification of Chrysochromulina and Prymnesium species (Haptophyta, Prymnesiophyceae) using fluorescent or chemiluminescent oligonucleotide probes: a means for improving studies on toxic algae. *Eur J Phycol*. 32:393–401.
Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough TG. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proc Natl Acad Sci USA*. 99:4430–4435.
Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol*. 18:1001–1013.
Takano Y, Hagino K, Tanaka Y, Horiguchi T, Okada H. 2006. Phylogenetic affinities of an enigmatic nannoplankton, *Braarudosphaera bigelowii* based on the *SSU* rDNA sequences. *Mar Micropaleontol*. 60:145–156.
Thompson J, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 25:4876–4882.
Thomsen HA, Buck KR, Chavez FP. 1994. Haptophytes as components of marine phytoplankton. In: Green JC, Leadbeater BSC, editors. The haptophyte algae. Clarendon Press, Oxford. p. 187–208.
Tyrrel T, Merico A. 2004. Emiliania huxleyi: bloom observations and the conditions that induce them. In: Thierstein HR, Young JR, editors. Coccolithophores: from the molecular processes to global impact. Springer Verlag, Berlin (Germany). p. 75–97.
Van Lenning K, Latasa M, Estrada M, Sáez AG, Medlin L, Probert I, Véron B, Young JR. 2003. Pigment signatures and phylogenetic relationships of the Pavlovophyceae (Haptophyta). *J Phycol*. 39:379–389.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
Yang Z, Rannala B. 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst Biol*. 54:455–470.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol*. 21:809–818.

Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. *Proc Natl Acad Sci USA*. 99:15507–15512.

Young JR. 1998. Neogene. In: Bown PR, editor. Calcareous nannofossil biostratigraphy. p. 225–265 British micropalaeontology society series. Dordrecht (the Netherlands): Kluwer Academic.

Young JR, Davis SA, Bown PR, Mann S. 1999. Coccolith ultrastructure and biomineralisation. *J Struct Biol*. 126: 195–215.

Young JR, Geisen M, Cros L, Kleijne A, Sprengel C, Probert I. 2003. A guide to extant calcareous nannoplankton taxonomy. *J Nannoplankton Res*. 1:1–125.

Young JR, Geisen M, Probert I. 2005. A review of selected aspects of coccolithophore biology with implications for paleobiodiversity estimation. *Micropaleontology*. 51:267–288.