

Heterotachy Processes in Rhodophyte-Derived Secondhand Plastid Genes: Implications for Addressing the Origin and Evolution of Dinoflagellate Plastids

Kamran Shalchian-Tabrizi,* Marianne Skånseng,* Fredrik Ronquist,†¹ Dag Klaveness,‡
Tsvetan R. Bachvaroff,§² Charles F. Delwiche,§ Andreas Botnen,|| Torstein Tengs,*³ and
Kjetill S. Jakobsen*

*Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, Oslo, Norway;

†Evolutionary Biology Centre, Department of Systematic Zoology, Uppsala University, Uppsala, Sweden; ‡Program for Plankton Biology, Department of Biology, University of Oslo, Oslo, Norway; §Cell Biology and Molecular Genetics, University of Maryland, College Park; and ||Scientific Computer Group, Center for Information Technology Services, University of Oslo, Oslo, Norway

Serial transfer of plastids from one eukaryotic host to another is the key process involved in evolution of secondhand plastids. Such transfers drastically change the environment of the plastids and hence the selection regimes, presumably leading to changes over time in the characteristics of plastid gene evolution and to misleading phylogenetic inferences. About half of the dinoflagellate protists species are photosynthetic and unique in harboring a diversity of plastids acquired from a wide range of eukaryotic algae. They are therefore ideal for studying evolutionary processes of plastids gained through secondary and tertiary endosymbioses. In the light of these processes, we have evaluated the origin of 2 types of dinoflagellate plastids, containing the peridinin or 19'-hexanoyloxyfucoxanthin (19'-HNOF) pigments, by inferring the phylogeny using "covarion" evolutionary models allowing the pattern of among-site rate variation to change over time. Our investigations of genes from secondary and tertiary plastids derived from the rhodophyte plastid lineage clearly reveal "heterotachy" processes characterized as stationary covarion substitution patterns and changes in proportion of variable sites across sequences. Failure to accommodate covarion-like substitution patterns can have strong effects on the plastid tree topology. Importantly, multigene analyses performed with probabilistic methods using among-site rate and covarion models of evolution conflict with proposed single origin of the peridinin- and 19'-HNOF-containing plastids, suggesting that analysis of secondhand plastids can be hampered by convergence in the evolutionary signature of the plastid DNA sequences. Another type of sequence convergence was detected at protein level involving the *psaA* gene. Excluding the *psaA* sequence from a concatenated protein alignment grouped the peridinin plastid with haptophytes, congruent with all DNA trees. Altogether, taking account of complex processes involved in the evolution of dinoflagellate plastid sequences (both at the DNA and amino acid level), we demonstrate the difficulty of excluding independent, tertiary origin for both the peridinin and 19'-HNOF plastids involving engulfment of haptophyte-like algae. In addition, the refined topologies suggest the red algal order, Porphyridales, as the endosymbiont ancestor of the secondary plastids in cryptophytes, haptophytes, and heterokonts.

Introduction

It is widely accepted that rhodophytes, glaucophytes, and viridiplantae obtained their plastids from an association with endosymbiotic cyanobacteria (Goksøyr 1967). The plastids of these algae are therefore termed primary. All other algae have probably obtained their plastids secondarily by engulfing eukaryotic algae (Delwiche 1999; Palmer 2003). In this manner, rhodophyte plastids have become incorporated into cryptophytes, heterokonts, haptophytes (defined as the chromists; Cavalier-Smith 1999) as well as into dinoflagellates and apicomplexa (Cavalier-Smith 1999; Fast et al. 2001; Patron et al. 2004). About half of the dinoflagellates are photosynthetic, harboring endosymbionts and plastids from several eukaryotic phyla (Watanabe et al. 1990; Chesnick et al. 1997; Tengs et al. 2000; Takishita et al. 2002). The vast majority of the dinoflagellate plastids use chlorophyll *c* and the pigment peri-

dinin, usually regarded as the ancestral plastid form (Hoek et al. 1995; Saunders et al. 1997; Saldarriaga et al. 2001). The other small groups of dinoflagellates with aberrant pigmentation have until recently been regarded as representatives of newer plastid lineages that replaced the peridinin-containing plastid during the radiation of dinoflagellates (Saldarriaga et al. 2001; Cavalier-Smith 2003; Patron et al. 2006). One of these replacement events involved a tertiary endosymbiosis of a haptophyte, resulting in dinoflagellates with 19'-hexanoyloxyfucoxanthin, the characteristic carotenoid of haptophytes (hereafter 19'-HNOF; Tengs et al. 2000). This evolutionary scenario for the plastids among dinoflagellates would imply several independent plastid acquisitions, but phylogenetic inference by Yoon et al. (2002) indicated a single origin for the peridinin- and 19'-HNOF-containing plastids involving engulfment of a haptophyte alga. In contrast to this study, subsequent analysis of protein characters applying 5 plastid-encoding genes divided the 2 dinoflagellate plastid groups (Yoon et al. 2005) and placed the peridinin plastids weakly within the heterokonts. Using 9 (and 10) plastid genes, Bachvaroff et al. (2005) showed the peridinin plastid and the haptophytes as sister groups with moderate to weak support. Thus, the inferences of the dinoflagellate plastids have so far generated incongruent results.

When plastids are transferred from one algal host to another, one might expect significant effects on the nature

¹ Florida State University, Department of Biological Science.

² Center of Marine Biotechnology, Baltimore, Maryland.

³ National Veterinary Institute, Department of Microbiology, Ullevålsveien 68, Oslo, Norway.

Key words: chromalveolates, chromists, covarion, dinoflagellates, heterotachy, plastid evolution.

E-mail: k.s.jakobsen@bio.uio.no.

Mol. Biol. Evol. 23(8):1504–1515. 2006

doi:10.1093/molbev/msl011

Advance Access publication May 15, 2006

of the evolutionary process. If such changes are not accommodated in the model of sequence evolution, and if taxa independently have come to evolve under similar conditions (i.e., being endosymbionts in similar host cells), this could lead to erroneous phylogenetic inferences. Three important aspects of molecular evolution are reflected in the standard nucleotide substitution models: differences in substitution rates, differences in nucleotide frequencies, and rate variation across sites. Of these, the rate variation component usually has a strong impact on the model fit (Yang 1996). Therefore, it is logical to explore evolutionary heterogeneity by focusing on changes in the pattern of among-site rate variation. To some extent, heterogeneous spatial substitution processes can be accommodated under a stationary reversible model, such as by a rates across sites (RASS) model (Yang 1996) or a covarion models (Tuffley and Steel 1998; Galtier 2001; Huelsenbeck 2002). These latter models are somewhat different from the original covarion model of Fitch and Markowitz (1970), but they do allow sites that are variable in some taxa to be invariable in other taxa. Like the model of Fitch and Markowitz (1970), these models all have the restrictive assumption that the proportion of variable sites (Pvar) must be the same in all lineages. A general concept that encompasses these models is that of heterotachy which ascribes shifts in the sequence evolution leading to heterogeneous composition of variable sites between lineages (Philippe and Germot 2000; Lopez et al. 2002; Lockhart and Steel 2005; Lockhart et al. 2005). Covarion or heterotachy substitution patterns have recently been uncovered in several examined genes, including genes from cyanobacteria and primary plastids, and have been suggested to cause inconsistency if not accounted for (Lockhart et al. 1998; Galtier 2001; Huelsenbeck 2002, Inagaki, Susko et al. 2004; Ané et al. 2005; Lockhart and Steel 2005; Lockhart et al. 2005). Covarion models have been implemented in maximum likelihood (ML) and Bayesian software, but heterogeneous processes leading to different Pvars across sequences are not yet properly modeled (Lockhart and Steel 2005). It is commonly recognized that reconstruction of phylogenies for plastid genes can be misled by unequal rates of change in different branches (Tengs et al. 2000; Zhang et al. 2000; Yoon et al. 2002; Holder and Lewis 2003; Bachvaroff et al. 2005). However, it is unclear whether the cause of this is unequal rates as studied by Felsenstein (1978) or because of lineage-specific differences in proportions of variable sites (Lockhart and Steel 2005). In either case, applying a gamma distribution to sequences appears to be somewhat helpful when reconstructing phylogenies as it sometimes provides improved estimates of genetic divergence. Yoon et al. (2002) implemented a gamma distribution in their study of dinoflagellate plastid sequences with minimum evolution (ME). However, only under this tree selection criterion and not Bayesian inference (BI) did they find support for monophyly of dinoflagellates. Other authors have emphasized that where there are violations of model assumptions, probabilistic models that analyze site patterns are more robust than distance methods (Huelsenbeck 1995a, 1995b; Holder and Lewis 2003; Lockhart et al. 2005). Lineage-specific differences of Pvar may also have an effect on the dominance of compositional biases if deviating substitution patterns

occur in the majority of sites that are free to vary. The characteristic codon usage in the fast-evolving dinoflagellate plastid genes could explain the conflicting DNA and protein phylogenies of these plastids (Yoon et al. 2002; Inagaki, Simpson et al. 2004; Yoon et al. 2005).

Here, we investigate the discrepancies between DNA and protein tree topologies, and reevaluate recent DNA and protein trees taking into account covarion-like substitution patterns, which has been shown for the green algal/plant plastid lineage but to a lesser extent been investigated for the rhodophyte-derived plastid lineages (Lockhart et al. 1998; Ané et al. 2005, Lockhart et al. 2005). Using new and previously published plastid genes (*psaA*, *psbB*, and *psbA*) and the nuclear encoded, plastid-targeted *psbO* (altogether encoding proteins of photosystem I and II), we reveal substitution patterns that are better accommodated by a stationary covarion model than a noncovarion model. We also make observations on substitution patterns suggesting nonstationary covarion-like processes (lineage-specific changes in proportions of variable sites). Our results suggest that failure to accommodate covarion-like processes has a profound effect on dinoflagellate plastid phylogenies. Most importantly, our DNA trees are consistent with a polyphyletic origin of the 19'-HNOF and peridinin plastids, previously only been shown in protein trees. In addition, detailed examination of character state changes suggests that previous incongruity between DNA and protein trees (Inagaki, Simpson et al. 2004; Yoon et al. 2005) is likely caused by substitution pattern at the protein level corresponding to the functional domain regions of the *psaA* gene. Excluding the *psaA* sequences, our analysis of protein sequences became consistent with trees inferred from DNA and recoded DNA sequences by grouping the peridinin and haptophyte plastids as a monophyletic group. In conclusion, by improved modeling of heterogeneous substitution processes, the inferred phylogeny make it difficult to rule out the possibility that the peridinin plastid was acquired from haptophytes by tertiary endosymbiosis.

Material and Methods

Cultures, Polymerase Chain Reaction, and Sequencing

All algal cultures were obtained from the Department of Marine Botany, University of Oslo and the Scandinavian Culture Center for Algae and Protozoa in Copenhagen. DNA was isolated from centrifuged log phase cultures using either magnetic beads or with a hexadecyltrimethylammonium bromide (CTAB) protocol as previously described (Doyle JJ and Doyle JL 1987; Rudi et al. 1997). *PsbA* sequences from 19'-HNOF dinoflagellates and haptophytes as well as *psbB* genes from haptophytes were provided using polymerase chain reaction (PCR) and combinations of degenerate and matching primers (Supplementary Table 1, Supplementary Material online).

Model Testing and Phylogenetic Analyses

All *psaA*, *psbA*, *psbB*, and *psbO* nucleotide sequences used in this work (Supplementary Table 2, Supplementary Material online) were aligned using ClustalX (Thompson

et al. 1997) and subsequently edited in accordance to the reading frames. To infer the phylogeny of 19'-HNOF and peridinin plastids, we analyzed the following DNA alignments: *psbA* alone (including our new haptophyte and 19'-HNOF dinoflagellate sequences), *psaA* and *psbA* sequences concatenated (identical to the data used by Yoon et al. 2002), and an alignment of *psaA*, *psbA*, *psbB*, and *psbO* concatenated (including new *psbB* sequences from haptophytes). Protein trees were inferred from the latter alignment and an additional data set composed of *psbA*, *psbB*, and *psbO* amino acid sequences (excluding *psaA*).

For each data set, the most appropriate homogenous process model was determined using hierarchical likelihood ratio tests (LRTs) with the Modeltest (for DNA sequences) and ProtTest (for protein sequences) programs (Posada and Crandall 1998, Abascal et al. 2005). These models were then applied in Bayesian inferences (BI) and compared with corresponding covarion models by calculating the Bayes factor defined as the ratio of the posterior probabilities of the hypotheses given that the prior probabilities of the hypotheses are equal. The marginal likelihood (predictive probability) of each hypothesis was estimated using the harmonic mean of the likelihood values from the stationary phase of Markov chain Monte Carlo (MCMC) runs, as suggested by Newton and Raftery (1994). We interpreted the Bayes factor according to the guidelines provided by Kass and Raftery (1995).

The evolutionary model used in the ML analysis of DNA sequences included a general time-reversible model (GTR), gamma distribution of site rates with 4 rate categories (Γ) and proportion of invariable sites (I); hereafter only $\Gamma + I$. ME trees were generated with LogDet distances from all codon sites and with exclusion of the third codon positions. Proportion of invariable sites was estimated in PAUP* (Swofford 1998) from a neighbor-joining Kimura 2-parameter tree, and invariable characters were removed according to base frequencies estimated from the constant sites only. ME and ML trees were determined by 10 heuristic searches, random addition of sequences and branch swapping with Tree Bisection-Reconnection and nearest neighbor interchange algorithms, respectively. The robustness of the tree topologies was tested with ME and ML analyses of 100 pseudoreplicates generated using the program CodonBootstrap v3.0b4 (Jonathan P. Bollback, http://www.binf.ku.dk/~bollback/bollback_nonie.html) and tree searches similar to the initial inferences.

BIs were performed using the program MrBayes v2.01 and v3.0 (Huelsenbeck and Ronquist 2001) by employing 4 different evolutionary models, including 2 homogenous and 2 heterogeneous process models with the following composition of rate parameters: 1) gamma distribution of site rates and proportion of invariant sites ($\Gamma + I$) and 2) site-specific rates (SS) for codon tripartitions. The corresponding covarion models were used by 3) including Γ and replacing the constant proportion of invariable site parameter (I) with dynamic covarion parameters (cov; resulting in $\Gamma + \text{cov}$) and 4) by adding the covarion parameters to the SS (SS + cov). Covarion estimation involves use of 2 parameters allowing sites to convert between invariable and variable modus (i.e., invariable > variable and variable > invariable). Analyses of proteins were done

with the $\Gamma + I$ and $\Gamma + \text{cov}$ models only. In addition, the cpREV substitution model (i.e., the best fitting models in ProtTest) was used in analyses of protein sequences. Four discrete rate categories were used to approximate the gamma distribution of RASs (8 rate categories were also used for some data sets without changing the results significantly), whereas priors for other model parameters were used with default values. Metropolis coupling was used with 3 incrementally heated (temperature parameter 0.2) and 1 cold chain, and swapping of 2 randomly selected chains attempted every generation. Two separate runs with randomly generated starting trees were carried out for 1 000 000–6 000 000 generations each (10 000 000 generations were also used for *psaA + psbA + psbB + psbO* DNA data without any changes in posterior probability values and likelihood scores). Sampling of trees was done every 100 generations. Burn-in of trees (i.e., sampled before the MCMC chains reached convergence) was usually set to 3000 trees based on assessment of the likelihood plots. Consensus of the remaining trees was used to calculate the posterior probabilities of the clades. In all analyses, the clade probabilities were essentially identical in the 2 independent runs starting from different, random topologies, supporting our conclusion that the chains produced a reasonable sample from the posterior distribution of interest with the chosen burn-in period. Here, we only show the results from the first of each pair of independent runs. Furthermore, BI of nonparametric bootstrap confidence for the DNA *psaA + psbA + psbB + psbO* sequences was performed as follows: hundred pseudoreplicates were generated using CodonBootstrap and analyzed with MrBayes with the same settings as before. Stationarity of the MCMC chains was assessed by calculating the standard deviation (SD) of the likelihood values obtained from MCMC chains in each of the 100 inferences. Visual inspection of the SD plots revealed stationarity at relatively early stages for all the runs (Supplementary Fig. 1, Supplementary Material online). The last 1000 trees from each analysis (i.e., in total 100 000 trees) were used for calculation of the majority-rule consensus tree. To further test the convergence of the MCMC chains, all 100 pseudoreplicates were analyzed in 2 independent runs. Because the posterior probability bootstrap values in the 2 resulting consensus trees were nearly identical, we regarded it likely that the chains had converged. All phylogenetic analyses were performed at the freely available Bioportal computer resources (<http://www.bioportal.uio.no/>).

Estimation of Lineage-Specific Patterns of Substitution

Sequences were analyzed to determine nonhomogeneous patterns of substitution. First, we made observations on lineage-specific patterns that could not be explained by more commonly assumed substitution models. We took particular note of sites that were unvaried in some lineages but varied in others. Second, we examined codon preferences in different lineages. Codon bias was estimated using codon adaptation index (CAI) and codon bias index (CAB) with the CodonW program (Peden 1997). DNA sequences for Leu, Ser, and Arg were recoded to TTN, TCN, and CGN, respectively to detect the potential impact

Table 1
Marginal Likelihood Values for Evolutionary Models Estimated from Bayesian MCMC Inferences

| Data set | DNA Models | | | | Protein Models | |
|---|------------|--------------------|-----------------------|----------------------|----------------------|------------------------|
| | GTR + SS | GTR + Γ + I | GTR + SS + cov | GTR + Γ + cov | cpREV + Γ + I | cpREV + Γ + cov |
| <i>psbA</i> | -10 060 | -9 948 | -9 863 (170) | -9 934 | | |
| <i>psaA</i> | -33 216 | -31 946 | -32 010 | -31 601 (690) | | |
| <i>psaA</i> + <i>psbA</i> | -49 548 | -47 750 | -47 527 | -47 283 (934) | | |
| <i>psaA</i> + <i>psbA</i> ^a | -21 605 | -21 356 | -21 130 (452) | -21 181 | | |
| <i>psaA</i> + <i>psbB</i> | -30 346 | -30 350 | -29 853 (986) | -30 217 | | |
| <i>psaA</i> + <i>psbB</i> ^b | -14 640 | -14 100 | -13 938 (324) | -13 986 | | |
| <i>psaA</i> + <i>psbB</i> + <i>psbO</i> | -39 232 | -39 080 | -38 539 (1082) | -38 922 | | |
| <i>psaA</i> + <i>psbB</i> + <i>psbO</i> ^b | -17 257 | -16 673 | -16 526 (294) | -16 576 | | |
| <i>psaA</i> + <i>psbA</i> + <i>psbB</i> + <i>psbO</i> | -43 733 | -43 566 | -43 054 (1024) | -43 395 | | |
| <i>psbA</i> + <i>psbB</i> + <i>psbO</i> | | | | | -9464 (12) | -9470 |
| <i>psaA</i> + <i>psbA</i> + <i>psbB</i> + <i>psbO</i> | | | | | -15 301 | -15 252 (98) |

NOTE.—Values in bold denote the best evolutionary model for the various data set, and the Bayes factor supporting the model compared with the best models without/with the covarion parameters are indicated in parentheses.

^a The likelihood values were estimated exclusively from rhodophyte sequences.

^b Sequences without third codon positions.

of lineage-specific codon preferences on tree building (Inagaki, Simpson et al. 2004). Third, we assessed compositional heterogeneity at all codon positions in recoded DNA data after deletion of invariant sites. Finally, we traced apparent synapomorphic amino acid changes in the *psaA* + *psbA* + *psbB* + *psbO* alignment using MacClade (Maddison D and Madisson W 2000). These characters were subsequently correlated with protein domain structures information from the cyanobacteria *Synechococcus elongatus* (accession number: 1JB0_A) downloaded from the National Center for Biotechnology Information structure database.

Results

Comparison of Evolutionary Models and Empirical Estimation of Heterotachy

Hierarchical LRTs of homogeneous models for DNA and protein data suggested the GTR + Γ + I and cpREV + Γ + I models to be significantly more likely than other models (results not shown), respectively. In contrast, Bayesian model tests of the marginal likelihoods revealed significant covarion-like structure in the evolution of all DNA data sets and one of the concatenated protein sequences (table 1). The best DNA covarion model was from 85 (Bayes factor = 170) to more than 500 (Bayes factor = 1000) log-likelihood units better than the best model without covarion structure. The signal is ambiguous concerning the best model for accommodating the basic among-site rate variation in evolutionary rates; the Γ + I model was usually better than the SS model when covarion structure was not taken into account, but the opposite was usually true for nucleotide data sets (SS + cov better than Γ + cov) when the covarion model component was added. This result is probably due to the effect of the proportion of invariant sites parameter; at least partly compensating for the lack of accommodating covarion-like structure in the Γ + I model but not in the SS model. Further investigation of the DNA alignments uncovered heterotachy in all genes and for all plastid groups but were most apparent in the dinoflagellates. Among all conserved sites, the dinoflagellate peridinin and 19'-HNOF sequences contained 432 and 108 unique

changes, respectively. Another 119 sites were variable in both plastid lineages, probably representing convergent substitution pattern. Hence, in total, 659 sites showed lineage-specific changes in the dinoflagellate sequences, but only 77 sites showed unique changes within the chromists (table 2).

PsbA DNA Phylogeny

The *psbA* tree topology, which was inferred from 29 taxa and 807 nt characters, was first reconstructed by using LogDet distances and ML (Γ + I) methods (LogDet analyses of whole codon and 1 + 2 positions generated similar results; fig. 1a). In this tree, both groups of dinoflagellate plastids were placed together with haptophytes (ME bootstrap value = 100, ML = 74), either as a sister group to the haptophytes or embedded among the haptophytes, respectively. Further, the monophyletic clustering of the dinoflagellate plastids was weakly supported in both analyses. Because accelerated evolutionary rates in these plastids (indicated by long branches) could have misled the inferences to cluster these 2 plastids, each of the plastid groups with

Table 2
Number of Variable Sites Unique for Each Plastid Type and for Defined Groups of Plastids Estimated from the Concatenated *psaA* + *psbA* + *psbB* + *psbO* DNA Alignment

| Plastid Lineage/Groups | Number of Sites |
|--|------------------|
| Dinoflagellate-peridinin | 432 |
| Dinoflagellate-19'-HNOF | 108 |
| Heterokonts | 30 |
| Haptophytes | 31 |
| Cryptophytes | 7 |
| Glaucophytes | 6 |
| Rhodophytes | 48 |
| Viridiplantae | 98 |
| Dinoflagellate peridinin + 19'-HNOF group | 659 ^a |
| Chromists (Hetero, Hapto, Crypto) group | 77 |
| Primary plastids (Rhodo, Glauco, Viride) group | 208 |

^a Of these 659 sites, 432 sites are variable only in the peridinin lineage, whereas 108 are variable in only the 19'-HNOF lineage. Thus, 119 sites have convergently become variable in both these groups.

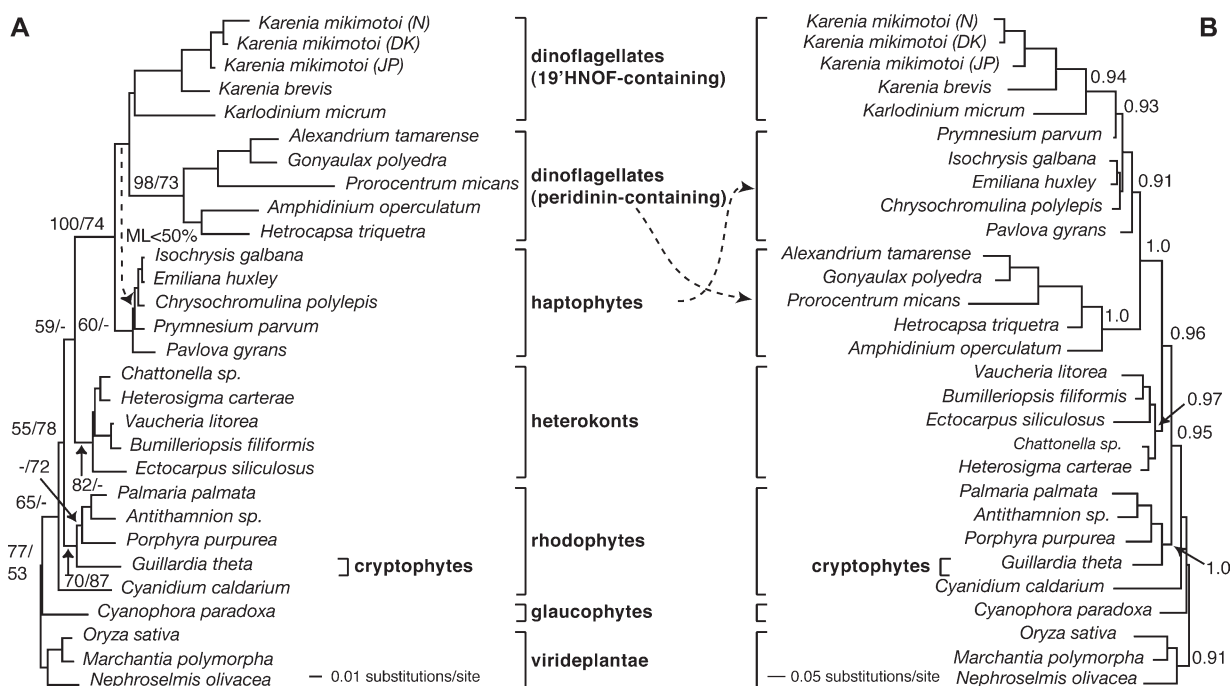


FIG. 1.—*PsbA* nucleotide trees. (A) Minimal evolution tree with LogDet distances. Bootstrap values >50% received from ME and ML analyses, respectively, are indicated at the nodes. (B) Bayesian tree generated with SS + cov model, which was determined as the most likely evolutionary model for the *psbA* gene. Posterior probability values >0.9 are indicated on the internal branches.

19'-HNOF and peridinin pigments were excluded in 2 subsequent analyses. However, both plastids remained together with haptophytes in LogDet and ML bootstrap analyses (results not shown). BI with $\Gamma + I$ and SS models recovered topologies similar to the ML tree, with 19'-HNOF- and peridinin-containing plastids as a monophyletic group (Supplementary Fig. 2, Supplementary Material online). Thus, our phylogenetic analyses of the *psbA* gene with homogenous evolutionary process models resulted in topologies highly congruent with *psbA* and *psaA + psbA* trees in Yoon et al. (2002). In contrast, Bayesian analysis under the favored covarion model (SS + cov) generated several new clades with moderate or high posterior probabilities (fig. 1b). Most importantly, the 19'-HNOF clade was no longer placed as the most basal haptophyte plastid branch (close to the plastids of the class Pavlovophyceae), but rather formed a branch among the plastids of the class Prymnesiophyceae with $P = 0.93$, whereas the peridinin plastid was separated from the 19'-HNOF and placed as a sister group to the haptophytes.

PsaA + psbA DNA Phylogeny

Model testing based on the *psaA* and *psaA + psbA* data (47 taxa and 2352 nt characters) suggested $\Gamma + cov$ as the optimal model. The *psaA + psbA* tree differed from the *psbA* tree at some important points (fig. 2A). The 2 rhodophyte plastid branches were highly supported (posterior probability, $P = 1.0$), whereas the cryptophyte plastids were placed in the same lineage as haptophytes and heterokonts with $P = 0.98$. In addition, the dinoflagellate plastids were clustered within the haptophytes as 2 separate groups; the 19'-HNOF together with Prymnesiophyceae and the peridinin with the Pavlovophyceae, congruent with

the analysis of the *psaA* data (Supplementary Fig. 3, Supplementary Material online). The same overall tree topology was generated when implementing the less favored $\Gamma + I$ model in ML analysis, although the bootstrap values only moderately or weakly supported the haptophyte affinity and the division of the dinoflagellates. However, the internal nodes of the rhodophyte plastid branches received considerably higher posterior probabilities with the SS + cov model and placed the single-celled *Flintiella* and *Porphyridium* genera at the base close to the divergence of the secondary plastids (fig. 2B; $P = 0.97$ and $P = 1.0$ for not being related to the other rhodophytes). Analysis of solely the rhodophyte species confirmed that the SS + cov model was significantly better for this particular algal lineage than the $\Gamma + cov$ model (table 1).

PsaA + psbA + psbB + psbO DNA Phylogeny

Further investigation of the plastid phylogeny was carried out with addition of *psbB* and *psbO* sequences (the alignment now comprising *psaA + psbA + psbB + psbO*, i.e., 18 taxa and 4428 nt characters; fig. 3). Bayesian analysis based on the highly favored SS + cov model placed the peridinin plastids together with the plastids of the basal haptophyte class Pavlovophyceae and the 19'-HNOF to the Prymnesiophyceae (both $P = 1.0$) as in the *psaA + psbA* tree (fig. 2). In addition, the majority of algal plastids were clustered as before with high posterior probability, except for the rhodophyte plastids that were not divided in 2 groups ($P = 0.92$) leading to a sister relationship to all chromist and dinoflagellate plastids. It should be noted that the cryptophytes emerged basal to the lineage with heterokonts, haptophytes, and dinoflagellates ($P = 1.0$). Bayesian analysis of bootstrapped *psaA + psbA + psbB + psbO* sequences

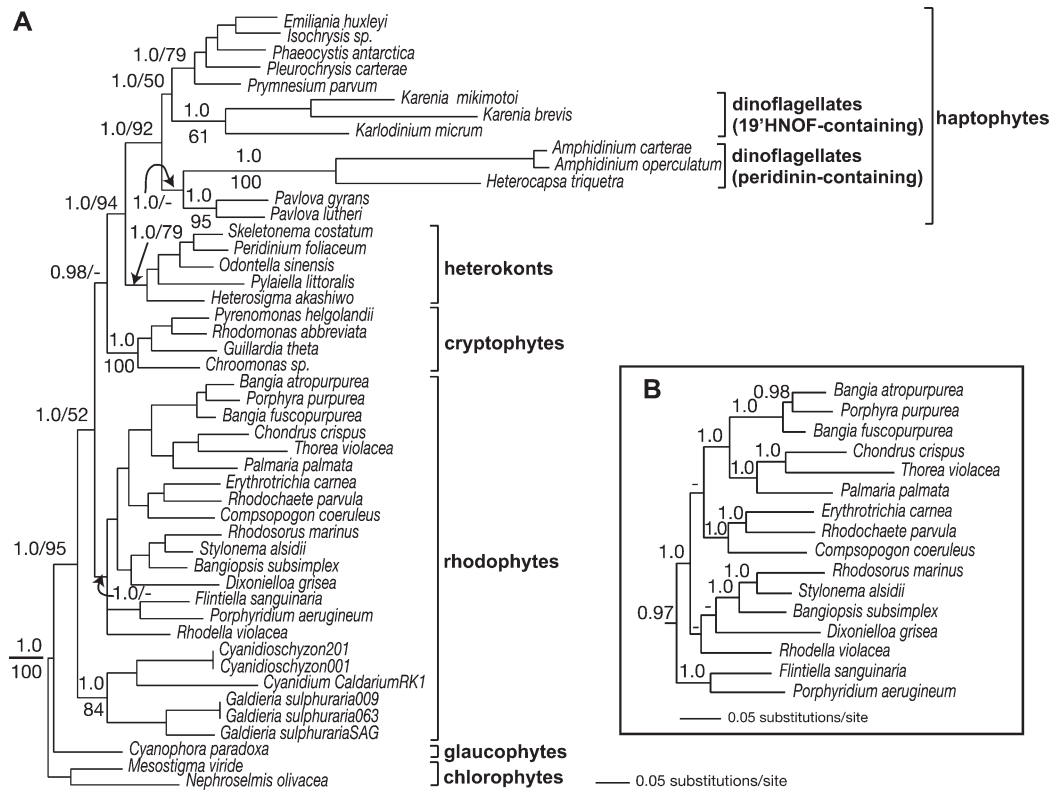


FIG. 2.—*PsaA* + *psbA* nucleotide phylogenetic trees reconstructed with Bayesian inferences. Posterior probability >0.9 and bootstrap values >50% from ML analysis using the Γ + I model are depicted at nodes. (A) Tree topology obtained using the favored Γ + cov model (B) A section of a tree obtained under the SS + cov model that was clearly favored for the rhodophytes (see table 1), receiving higher posterior probability values for the basal branches in rhodophytes.

gave similar results as the inferences of the original data set, placing the peridinin group together with the Pavlovophyceae and 19'-HNOF together with Prymnesiophyceae lineages with moderate support (69% and 75%, respectively). The ML analysis with the Γ + I model reconstructed a phylogeny that resembled the overall phylogeny inferred with BI, although the bootstrap support for dividing the dinoflagellates were lower than with BI (60% support for 19'-HNOF + Prymnesiophyceae, 54% for peridinin + Pavlovophyceae). Importantly, all the retrieved trees from the MCMC chains placed the dinoflagellates in 2 distinct clades, implying an extremely low posterior probability for alternative topologies under the given covariotide substitution model (the main topological differences were related to the position of the rhodophyte species, either as monophyletic or polyphyletic). In 2 additional analyses, sequences were excluded to test whether putative codon bias (in *psbA*) and putative cryptic endosymbionts (i.e., the *psbO* sequences) would affect the specific placement of the 2 dinoflagellate plastids, but all trees were congruent (Supplementary Figs. 4 and 5, Supplementary Material online).

Phylogeny of Recoded *PsaA* + *psbA* + *psbB* + *psbO* DNA Sequences

Estimating the codon preferences, the dinoflagellate 19'-HNOF plastids and some of the peridinin plastids received lower CAI and CBI indexes for *psbA* than the vast

majority of the other plastid lineages (Supplementary Table 3, Supplementary Material online). This bias could potentially create erroneous attraction between these 2 plastid forms, but exclusion of the *psbA* from concatenated sequences did not significantly change the topology. Separate analyses of first + second codon position and only the third position resulted in congruent tree topologies, suggesting that all sites—even the third codon positions—contain congruent phylogenetic information (Supplementary Figs. 6–8, Supplementary Material online). The CAI and CAB indexes showed no overall codon bias between the plastids in dinoflagellates and the other algal groups. However, as recently shown for the *psbA* gene (Inagaki, Simpson et al. 2004), the haptophytes and some of the peridinin showed similar tendencies of overlapping codon usage for the 3 amino acids Leu, Ser, and Arg in the other genes applied here (Supplementary Fig. 9, Supplementary Material online). Because these biases could mislead the placement of the peridinin plastid, we recoded the codons for Leu, Ser, and Arg in the concatenated *psaA* + *psbA* + *psbB* + *psbO* data and excluded the 19'-HNOF sequences to avoid long-branch attraction problems between the 2 dinoflagellate plastid groups, as done previously (Inagaki, Simpson et al. 2004). However, the resulting tree still supported the monophyly of the peridinin and haptophytes, and the grouping of the peridinin and Pavlovophyceae plastids, although the latter topology was only weakly supported (fig. 4). The haptophyte-peridinin branch received posterior

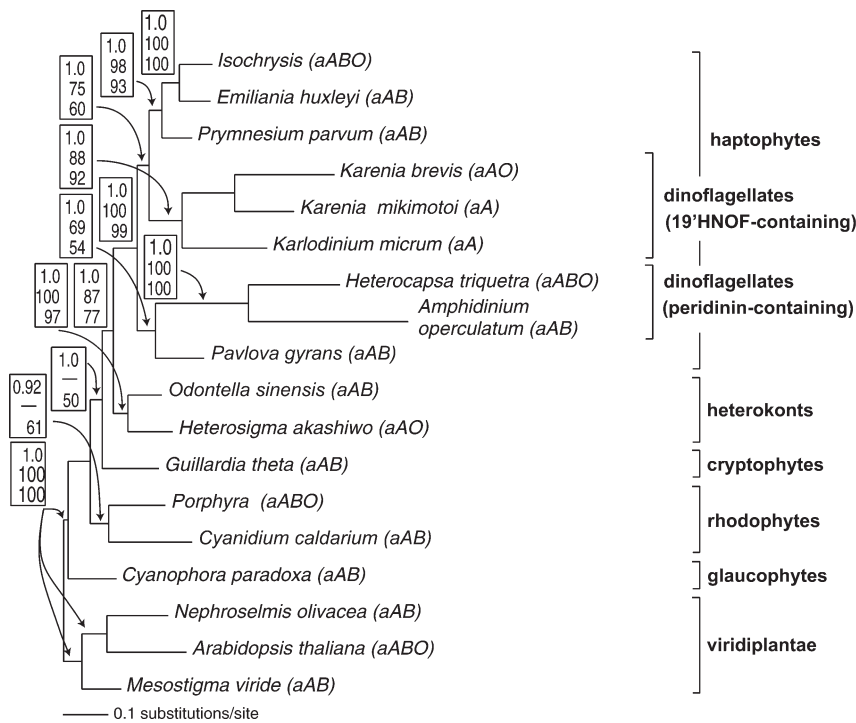


FIG. 3.—Phylogeny of the *psaA* + *psbA* + *psbB* + *psbO* DNA sequences inferred with Bayesian MCMC analysis (BI) under the SS + cov evolution model. Support values at internal nodes were received from 3 approaches in the following order: BI, nonparametric bootstrap analyses with BI and ML methods. Posterior probability values >0.9 and bootstrap values >50% are indicated. The concatenated *Porphyra* sequence is composed of *Porphyra purpurea* and *Porphyra yezoensis* genes, whereas *Isochrysis* is composed of *Isochrysis galbana* and *Isochrysis* sp. sequences. a, *psaA*; A, *psbA*; B, *psbB*; O, *psbO*.

probability of 1.00, which means that very few other topologies for these 2 groups were retrieved from the MCMC chains. Nucleotide composition bias is another factor that may erroneously cluster sequences, but using recoded DNA data and chi-square test in PAUP*, only the third codon position showed base frequencies deviating significantly from the null hypothesis of a homogeneous distribution of base frequencies (results not shown). Trees

inferred from the recoded DNA sequences using first + second and third positions resulted in similar tree topology, grouping the peridinin and haptophyte plastids (Supplementary Figs. 10 and 11, Supplementary Material online), suggesting that the nucleotide composition bias is not important for the inferences using the entire codon. It also suggests that the third codon position contain substantial phylogenetic information even after the recoding of the data.

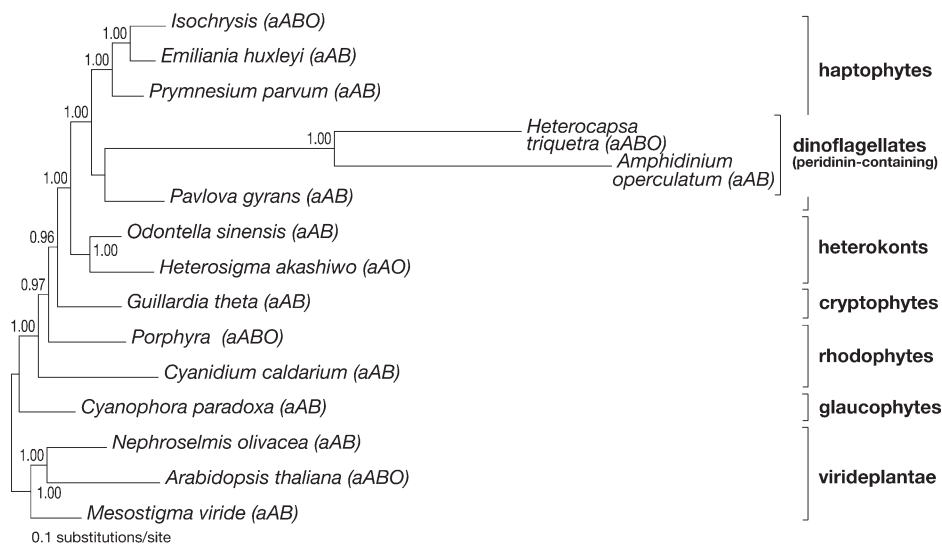


FIG. 4.—Phylogeny of recoded DNA *psaA* + *psbA* + *psbB* + *psbO* sequences inferred with Bayesian MCMC analysis under the SS + cov evolution model. Posterior probability values >0.9 are indicated at the nodes.

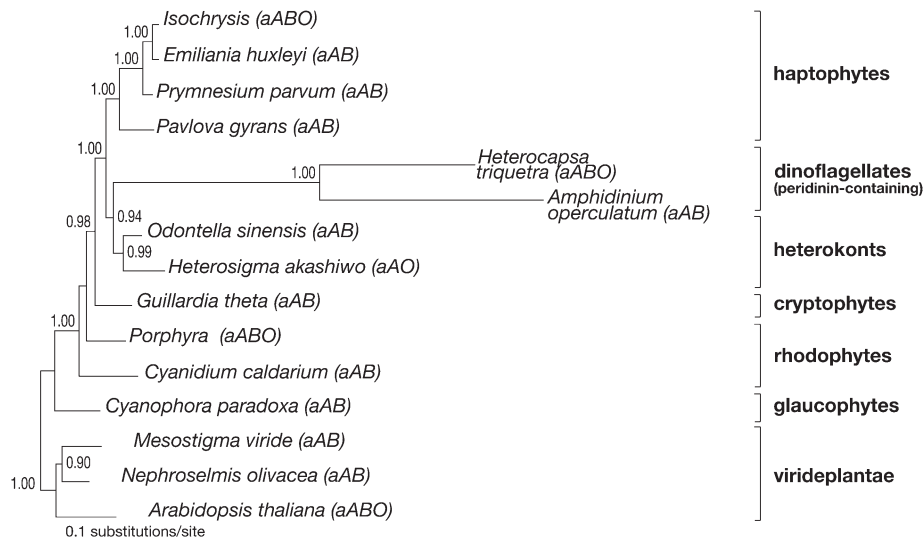


FIG. 5.—Protein *psaA* + *psbA* + *psbB* + *psbO* tree reconstructed with Bayesian inference applying cpREV + Γ + cov model. Posterior probability values >0.9 are indicated at the branches.

PsaA + *psbA* + *psbB* + *psbO* Protein Phylogeny and Convergence in the *psaA* Gene

Model comparison using ProtTest suggested the cpREV + Γ + I evolutionary model to be the best fitting model for the *psaA* + *psbA* + *psbB* + *psbO* protein data (table 1). Applying this model, the inferred protein tree resulted in high congruence with the DNA tree for almost all groups except the peridinin plastid, which was placed as a sister group to the heterokonts (fig. 5) with posterior probability = 0.94. Altogether, therefore, using the same genes, the DNA (including the recoded DNA) and protein characters resulted in 2 different but almost equally supported placement of the peridinin plastid. Tracing the amino acid changes in the *psaA* + *psbA* + *psbB* + *psbO* alignment on the DNA and protein trees revealed striking patterns in the

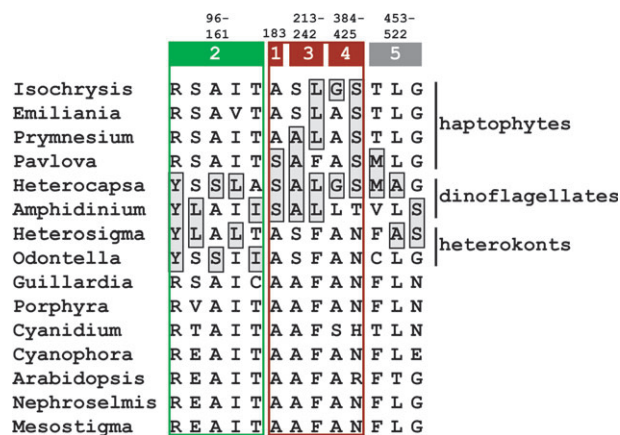


FIG. 6.—Distribution of synapomorphic characters in the *psaA* gene aligned according to the domain regions in the *psaA* from *Synechococcus elongatus*. Boxes colored in green show domain region 2 and synapomorphic character states grouping the heterokont and peridinin plastid, whereas red boxes mark domain regions 1, 3, and 4, containing sites clustering the haptophyte and peridinin plastid. Domain 5 shown in light gray contains synapomorphic information for both type of placement of the peridinin plastid.

psaA gene (fig. 6): synapomorphic characters supporting the alternative placements of the peridinin plastid were only found in defined domain structures of photosystem I. Compared with the region structure in the *psaA* gene in *S. elongatus*, 3 domains (1, 3, and 4) were composed entirely of synapomorphic characters clustering the peridinin and haptophyte plastids, whereas one of the domains (number 2) showed only synapomorphic characters for grouping the peridinin and heterokont plastids (domain 5 contains synapomorphic characters for both topologies). Because *psaA* displays inconsistent phylogenetic information that is highly related to functional regions of the molecule, we excluded the *psaA* sequence from the concatenated protein sequences. When analyzing the *psbA* + *psbB* + *psbO* sequences, the best fitting model switched from a covarion model to a noncovarion cpREV + Γ + I model. The tree inferred by the latter evolutionary model resulted in a different protein tree (fig. 7), clustering the peridinin plastid with the haptophytes with posterior probability = 0.83. This branching pattern is congruent with all inferred DNA trees. Additional inferences were done deleting other genes to uncover change in the placement of the peridinin plastid. Although the support values in general changed for the peridinin and heterokont clade, none of the data sets showed similar change in topology and support values as when excluding the *psaA* sequences. However, among the other genes, the *psbA* gene may have largest impact on the trees (results not shown). Altogether, these inferences suggest that the phylogenetic information in *psaA* is incompatible with the information in the other genes.

Discussion

Covarion and Heterotachy Processes in Evolution of Plastid Genes

Tests of various evolutionary models for the plastid DNA sequences in a Bayesian framework consistently favor the covarion evolution models (SS + cov or Γ + cov) over GTR + Γ + I models for single and concatenated

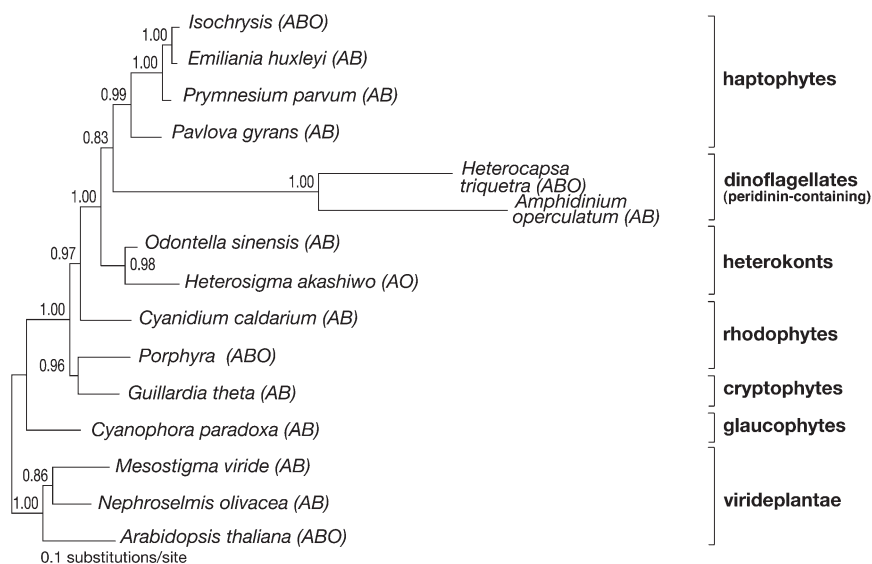


FIG. 7.—Protein tree inferred using the same sequences as in figure 5, but with exclusion of the *psaA* gene (i.e., *psbA* + *psbB* + *psbO* genes). Bayesian posterior probability generated with cpREV + Γ + I model >0.8 are indicated at the branches.

sequences by Bayes factors larger than 170. It strongly supports the use of covarion models on these data, bearing in mind that a difference of 10 is considered significant evidence for a better model (Kass and Raftery 1995). This is also in agreement with detailed examination of site patterns, which suggested lineage-specific difference in evolutionary properties of sequences for all plastids and in particular for the dinoflagellate peridinin and 19'-HNOF plastids. It clearly reveals the importance of heterotachy in both primary and secondhand plastid lineages (Lockhart et al. 1998; Ané et al. 2005; Lockhart et al. 2005).

Using the various evolutionary models in analyses of *psbA* DNA data, only the favored covarion evolutionary model divided the peridinin- and 19'-HNOF-containing plastids into polyphyletic groups, in agreement with the protein sequences (Inagaki, Simpson et al. 2004). The placement of these 2 plastid groups occurred despite relatively high codon bias in these plastids (revealed by CAI and CBI indexes) but was also congruent with trees inferred when excluding one of the dinoflagellate plastid lineages. These observations suggest that codon bias is not causing substantial false attraction between the peridinin and 19'-HNOF plastids when applying the covarion substitution model. Altogether, the relatively low CAI and CBI indexes, clear covarion structure, and apparent differences in Pvar suggest relaxed selectional constraints in the *psbA* gene in peridinin and 19'-HNOF plastids. Similar processes have been observed in the *psbA* and RNA polymerase subunits from green plastid lineage (Morton and Levin 1997; Lockhart et al. 2005).

Increasing the number of nucleotide characters by adding *psaA*, *psbB*, and *psbO* sequences, the effects of covarion substitutions seem to be of less importance because both the among-site rate (used in ML analyses) and covarion models (used in BI analyses) reconstructed congruent trees. All analyses of concatenated sequences received high posterior probability values for all critical nodes dividing the 2 dinoflagellate plastid types and for embedding these lineages in

the haptophyte clade. The Bayesian and ML analyses on bootstrapped data also resulted in similar topology with slightly higher support values from the BIs than for the ML analysis, indicating the appropriateness of using the covarion evolution model on these sequences.

Peridinin and 19'-HNOF Plastids Acquired from 2 Separate Haptophyte Lineages

Altogether, the congruent topologies received from different DNA data sets using covarion and among-site rate models with probabilistic methods clearly suggest a polyphyletic origin of the 19'-HNOF and peridinin plastids. This is inconsistent with a single origin inferred from DNA sequences using distance methods (Yoon et al. 2002). Further, if as we suggest, dinoflagellate plastid genes are convergent as a result of heterotachy, then explanation for the discordance between trees built from distance and site pattern methods may also reflect the inefficiency of distance methods under conditions of substitution model misspecification. Similar problems have been reported elsewhere (Swofford et al. 1996; Felsenstein 2001; Holder and Lewis 2003; Susko et al. 2004; Lockhart et al. 2005). The obtained split of dinoflagellate plastids in covarion analyses of DNA data, so far only seen in protein trees, indicate that the implemented covarion model takes into account much of the heterotachy processes at DNA level. However, the covarion model implemented in these analyses does not account for nonstationary substitution processes, such as when Pvar changes between lineages. From our observations, it seems likely that this property of the data may be causing some instability in phylogenetic reconstruction (Lockhart et al. 2005; Lockhart and Steel 2005).

The specific relationship between the peridinin and haptophyte plastids inferred from photosynthetic genes applying DNA characters was recently suggested to be erroneously caused by convergent codon preferences in

psbA for 3 amino acids, the Leu, Ser, and Arg (Inagaki, Simpson et al. 2004). We therefore excluded this gene from the concatenated DNA sequences, but the peridinin plastid still remained as sister group to the Pavlovophyceae. Examination of the other genes used here showed some tendencies of being composed of similar biased codon pattern, although not as apparent as for the *psbA* gene. Thus, to eliminate this problem, we recoded all triplets encoding the Leu, Ser, and Arg in the *psaA* + *psbA* + *psbB* + *psbO* nucleotide alignment as done by Inagaki, Simpson et al. (2004). Using the covarion model on the recoded data, however, the tree still showed a monophyletic peridinin-haptophyte phylogeny with high posterior probability. In contrast, applying amino acid characters, the inferences of the *psaA* + *psbA* + *psbB* + *psbO* amino acid sequences grouped together the peridinin plastids and heterokonts, also supported with high posterior probability. Altogether, therefore, phylogenetic inferences of all DNA data (both original and recoded) and protein sequences are generating different, but supported, results. There may be several reasons why the DNA and protein characters could produce different topologies (Swofford et al. 1996; Simmons et al. 2004), but because codon bias has been accounted for by recoding of the data, and the first + second and third codon positions have been shown to contain congruent phylogenetic information, we investigated whether the discrepancy was caused by convergent evolution of the protein sequences. Mapping synapomorphic characters for clustering peridinin plastids together with either haptophytes or heterokonts showed that in the *psaA* gene, most of the synapomorphic characters supporting the heterokont-peridinin or the haptophyte-peridinin topologies were exclusively found in distinct, but separate, domain structures of the molecule. Because the domains often represent functional units each interacting with specific components of photosystem I, the character distribution pattern may be shaped by uneven selection forces, to become either the heterokont- or haptophyte-like domains. Thus, changes in the nature of protein-protein interactions are likely to have implications for phylogenetic reconstruction of plastid genes (Lockhart et al. 2005). Excluding the *psaA* sequences from the analysis of the concatenated data changed the best fitting model to a noncovarion model and placed the peridinin plastid together with haptophytes congruent with all DNA trees (see fig. 7). This opens up the possibility that previously identified codon biases in peridinin and haptophyte plastids (Inagaki, Simpson et al. 2004) may actually represent valid information reflecting phylogenetic branching of the plastids, which is removed by recoding the data (Lockhart et al. 1998; Simmons et al. 2004). Altogether, therefore, applying multiple genes and better fitting DNA and protein substitution models than previously used, our results demonstrate the difficulty of excluding the hypothesis that the peridinin plastid was acquired from haptophytes through tertiary endosymbiosis. If the peridinin plastid was acquired from a haptophyte, our data are consistent with the 19'-HNOF and peridinin plastid originated from 2 different haptophyte lineages, the Prymnesiophyceae and Pavlovophyceae, respectively. Together with the broad phylogenetic distribution of the peridinin plastid among genera in the current classification of dinoflagellates

and in nuclear rRNA and Hsp90 gene trees, this plastid phylogeny suggests that the peridinin plastid was present in the earliest lineages of phototrophic dinoflagellates, and in a single dinoflagellate lineage, replaced by the 19'-HNOF plastid at a later stage in evolution (Saldarriaga et al. 2001; Patron et al. 2006; Shalchian-Tabrizi et al. 2006).

Origin of the Chromist Plastids

The trees obtained from the largest taxon sampling (included *psaA* + *psbA*), confirm the red algal origin of all chromist and dinoflagellate plastids (Zhang et al. 1999; Fast et al. 2001; Yoon et al. 2002). However, the rhodophyte plastids are apparently evolving under a different process than other plastid lineages because model tests of the *psaA* + *psbA* sequences favor different models for rhodophytes (SS + cov) than the whole data set (Γ + cov). The SS + cov model more precisely resolves the possible ancestor of the secondary chromist and tertiary dinoflagellate plastids by embedding all these lineages in the single-celled Porphyridiales, supporting the idea that the precursor of all chlorophyll *c*-containing plastids originated from this particular rhodophyte order (Seckbach 1994). One of these rhodophyte lineages is composed of *Cyanidium* and *Galdieria* genera living in hostile acidic and hot environments, rarely frequented by other eukaryotes, suggesting that plastids in alveolates and chromists were more likely obtained from species related to *Flintiella* or *Porphyridium* (Yoon et al. 2005).

How Reticulate is the Evolution of Secondary and Tertiary Plastids?

Phylogenetic investigations of nuclear-encoded plastid genes and concatenated plastid sequences have recently provided arguments for a single origin of both the plastids and host lineages of chromist and alveolates (i.e. chromalveolates; Cavalier-Smith 1999; Fast et al. 2001; Cavalier-Smith 2002; Patron et al. 2004). Serial transfers are less parsimonious than the chromalveolate hypothesis, implying more endosymbiotic events and parallel transfer of genes from the endosymbiont to the host nucleus. However, both the peridinin- and 19'-HNOF-containing dinoflagellate species have unusual ability to transfer genes from their enslaved alga to their own nucleus (Yoon et al. 2005, Bachvaroff et al. 2004; Hackett et al. 2004; Patron et al. 2006). Some of these genes were likely transferred directly from the nucleus of the endosymbiont rather from the organelles, such as the *psbO* and *GAPDH* genes (Ishida and Green 2002; Takishita et al. 2004), indicating that the establishment of nucleus-sited plastid-targeted genes in dinoflagellates could be considerably simplified by recycling genes and modifying plastid targeting systems already associated with the genes (Ishida and Green 2002; Takishita et al. 2004; Patron et al. 2006).

Based on previous studies (Harper and Keeling 2003, Bachvaroff et al. 2005, Yoon et al. 2005) and results presented here, there are at least 3 alternative scenarios for the origin and evolution of the peridinin and chromist plastids: 1) a common, single ancestor, likely a single rhodophyte, in accordance with the original chromalveolate hypothesis; 2) independent origins, where the peridinin plastid was

obtained tertiary from a haptophyte (i.e., rhodophyte derived); and 3) a single, rhodophyte-derived plastid was established in the chromalveolate ancestor, followed by a replacement event early in the radiation of the dinoflagellates (or alveolates; Leander and Keeling 2003), involving loss of the original rhodophyte-derived chromalveolate plastid and gain of a new plastid from haptophytes. In addition, our data do not exclude a more extensive serial endosymbiosis scenario among the chromists and dinoflagellates (Bachvaroff et al. 2005). This work demonstrates that relatively subtle, but nonetheless important, modifications of probabilistic methods may have substantial impact on the tree topology (Lockhart et al. 2005) and thus on our understanding of central questions in the evolution of plastids and their host phylogenies. Therefore, further empirical studies are needed to elucidate the importance of nonstationary covarion and heterogeneous processes in phylogenetic reconstruction of plastid evolution.

Supplementary Material

Supplementary tables 1–3 and Figs. 1–11 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We would like to thank W. Eikrem, J. Nylander, Y. Inagaki, and T. Bjørnland for discussions and V. Knutsen, V. Eide, and J. F. Myklebust at the High Performance Computing network in Norway for computer assistance. We are grateful for algal cultures kindly provided by W. Eikrem and we are indebted to D. Bhattacharya for providing the *psaA* + *psbA* alignment and to B. R. Green and K. Ishida for giving us the *Heterosigma akashiwo psbO* sequence prior to publication. This work was supported by research grants from the Norwegian Research Council (NRC) to K.S.J. (grant no 118894/431 and to D.K. (grant no 121187/720) and from the NRC Program for Supercomputing.

Literature Cited

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–5.
- Ané C, Burleigh JG, McMahon MM, Sanderson MJ. 2005. Covarion structure in plastid genome evolution: a new statistical test. *Mol Biol Evol* 22:914–24.
- Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM, Delwiche CF. 2004. Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* 155:65–78.
- Bachvaroff TR, Puerta MVS, Delwiche CF. 2005. Chlorophyll *c* containing plastid relationships based on analyses of a multi-gene data set with all four chromalveolate lineages. *Mol Biol Evol* 22:1772–82.
- Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* 46:347–66.
- Cavalier-Smith T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa. *Int J Syst Evol Microbiol* 52:297–354.
- Cavalier-Smith T. 2003. Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Philos Trans R Soc Lond B Biol Sci* 358:109–34.
- Chesnick JM, Kooistra WH, Wellbrock U, Medlin LK. 1997. Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *J Eukaryot Microbiol* 44:314–20.
- Delwiche CF. 1999. Tracing the thread of plastid diversity through the tapestry of life. *Am Nat* 154(suppl):S164–77.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–5.
- Fast NM, Kissinger JC, Roos DS, Keeling PJ. 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* 18:418–26.
- Fast NM, Xue L, Bingham S, Keeling PJ. 2002. Re-examining alveolate evolution using multiple protein molecular phylogenies. *J Eukaryot Microbiol* 49:30–7.
- Felsenstein J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst Zool* 27:401–10.
- Felsenstein J. 2001. Taking variation of evolutionary rates between sites into account in inferring phylogenies. *J Mol Evol* 53:447–55.
- Fitch WM, Markovitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixations of mutations in evolution. *Genetics* 4:579–93.
- Galtier N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol Biol Evol* 18:866–73.
- Goksøyr J. 1967. Evolution of eucaryotic cells. *Nature* 214:1161.
- Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavert TL, Scheetz TE, Nosenko T, Bhattacharya D. 2004. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr Biol* 14:213–8.
- Harper TH, Keeling PJ. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol* 20:1730–5.
- Hoek Cvd, Mann DG, Jahns HM. 1995. *Algae: an introduction to phycology*. Cambridge, UK: Cambridge University Press.
- Holder M, Lewis PO. 2003. Phylogeny estimation: traditional and Bayes approaches. *Nat Rev Genet* 4:275–84.
- Huelsenbeck JP. 1995a. The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of maximum likelihood over neighbor joining. *Mol Biol Evol* 12:843–9.
- Huelsenbeck JP. 1995b. Performance of phylogenetic methods in simulation. *Syst Biol* 44:17–48.
- Huelsenbeck JP. 2002. Testing a covarion model of DNA substitution. *Mol Biol Evol* 19:698–707.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–5.
- Inagaki Y, Simpson A, Dacks J, Roger A. 2004. Phylogenetic artifacts can be caused by leucine, serine, and arginine codon usage heterogeneity: dinoflagellate plastid origins as a case study. *Syst Biol* 53:582–93.
- Inagaki Y, Susko E, Fast NM, Roger AJ. 2004. Covarion shifts cause a long-branch attraction artifact that unites microsporidia and archaeobacteria in EF-1 α phylogenies. *Mol Biol Evol* 21:1340–9.
- Ishida K, Green BR. 2002. Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc Natl Acad Sci USA* 99:9294–9.
- Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc* 90:773–95.
- Leander BS, Keeling PJ. 2003. Morphostasis in alveolate evolution. *Trends Ecol Evol* 18:395–402.

- Lockhart P, Novis P, Milligan BG, Riden J, Rambaut A, Larkum T. 2005. Heterotachy and tree building: a case study with plastids and eubacteria. *Mol Biol Evol* 32:40–5.
- Lockhart P, Steel M. 2005. A tale of two processes. *Syst Biol* 54:948–51.
- Lockhart PJ, Steel MA, Barbrook AC, Huson DH, Charleston MA, Howe CJ. 1998. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol Biol Evol* 15:1183–8.
- Lopez P, Casane D, Philippe H. 2002. Heterotachy, an important process of protein evolution. *Mol Biol Evol* 19:1–7.
- Maddison D, Maddison W. 2000. *MacClade 4: analysis of phylogeny and character evolution*. Sunderland, MA: Sinauer Associates Inc. Publishers.
- Morton BR, Levin JA. 1997. The atypical codon usage of the plant *psbA* gene may be the remnant of an ancestral bias. *Proc Natl Acad Sci USA* 54:948–51.
- Newton MA, Raftery AE. 1994. Approximate Bayes inference by the weighted likelihood bootstrap (with discussion). *J R Stat Soc* 56:3–48.
- Palmer JD. 2003. The symbiotic birth and spread of plastids: how many times and whodunit? *J Phycol* 39:4–11.
- Patron NJ, Matthew BR, Keeling PJ. 2004. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot Cell* 3:1169–75.
- Patron NJ, Waller RF, Keeling PJ. 2006. A tertiary plastid uses genes from two endosymbionts. *J Mol Biol* 305:1373–82.
- Peden J. 1997. CodonW. Available from: <http://www.molbiol.ox.ac.uk/cu/codonW.html>.
- Philippe H, Germot A. 2000. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. *Mol Biol Evol* 17:830–4.
- Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–8.
- Rudi K, Kroken M, Dahlberg OJ, Deggerdal A, Jakobsen KS, Larsen F. 1997. Rapid universal method to isolate PCR-ready DNA using magnetic beads. *Biotechniques* 22:506–11.
- Saldarriaga JF, Taylor FJ, Keeling PJ, Cavalier-Smith T. 2001. Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *J Mol Evol* 53:204–13.
- Saunders GW, Hill DRA, Sexton JP, Andersen RA. 1997. Small-subunit ribosomal RNA sequences from selected dinoflagellates: testing classical evolutionary hypotheses with molecular systematic methods. *Plant Syst Evol* 11(Suppl):237–59.
- Seckbach J. 1994. *Evolutionary pathways and enigmatic algae: Cyanidium caldarium (Rhodophyta) and related cells*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Shalchian-Tabrizi K, Minge MA, Cavalier-Smith T, Nedreklepp JM, Klaveness D, Jakobsen KS. 2006. Combined Hsp90 and rRNA sequence phylogeny supports multiple replacements of dinoflagellate plastids. *J Eukaryot Microbiol* 53:217–24.
- Simmons MP, Carr TG, O'Neill K. 2004. Relative character-state space, amount of potential phylogenetic information, and heterogeneity of nucleotide and amino acid characters. *Mol Phylogenet Evol* 32:913–26.
- Susko E, Inagaki Y, Roger AJ. 2004. On inconsistency of the neighbor-joining, least squares, and minimum evolution estimation when substitution processes are incorrectly modelled. *Mol Biol Evol* 21:1629–42.
- Swofford DL. 1998. *PAUP*: phylogenetic analysis using parsimony (*and other methods)*. Sunderland, MA: Sinauer Associates Inc. Publishers.
- Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. *Phylogenetic inference*. In: Hillis MH, Moritz C, Mable B, editors. *Molecular systematics*. Sunderland, MA: Sinauer Associates Inc. Publishers.
- Takishita K, Ishida K, Tadashi M. 2004. Phylogeny of nuclear-encoded plastid-targeted GAPDH gene supports separate origins for the peridinin- and the fucoxanthin derivative-containing plastids of dinoflagellates. *Protist* 155:447–58.
- Takishita K, Koike K, Maruyama T, Ogata T. 2002. Molecular evidence for plastid robbery (Kleptoplastidy) in *Dinophysis*, a dinoflagellate causing diarrhetic shellfish poisoning. *Protist* 153:293–302.
- Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, Jakobsen KS. 2000. Phylogenetic analyses indicate that the 19' Hexanoxyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol* 17:718–29.
- Thompson JD, 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* 25:4876–82.
- Tuffley C, Steel M. 1998. Modeling the covarion hypothesis of nucleotide substitution. *Math Biosci* 147:63–91.
- Watanabe MM, Suda S, Inouye I, Sawaguchi T, Chihara M. 1990. *Lepidodinium viride*, new genus new species (Gymnodiniales, Dinophyta), a green dinoflagellate with a chlorophyll A-containing and chlorophyll B-containing endosymbiont. *J Phycol* 26:741–51.
- Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol* 11:367–72.
- Yoon HS, Hackett JD, Bhattacharya D. 2002. A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc Natl Acad Sci USA* 99:11724–9.
- Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D. 2005. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol Biol Evol* 22:1299–308.
- Zhang Z, Green BR, Cavalier-Smith T. 1999. Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400:155–9.
- Zhang Z, Green BR, Cavalier-Smith T. 2000. Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. *J Mol Evol* 51:26–40.

Peter Lockhart, Associate Editor

Accepted May 4, 2006