Protist, Vol. 163, 217–231, March 2012
http://www.elsevier.de/protis

Protist

ORIGINAL PAPER

Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles

Eun Chan Yang[a,1], **Ga Hun Boo**[b,1], **Hee Jeong Kim**[a], **Sung Mi Cho**[b], **Sung Min Boo**[b,2], **Robert A. Andersen**[c,2], and **Hwan Su Yoon**[a,d,2]

[a]Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME 04575, USA
[b]Department of Biology, Chungnam National University, Daejeon 305-764, Korea
[c]Friday Harbor Laboratories, University of Washington, Friday Harbor, WA 98250 USA
[d]Department of Biological Sciences, Sungkyunkwan University, Suwon 440-746, Korea

Submitted January 24, 2011; Accepted August 7, 2011
Monitoring Editor: Hervé Philippe

Molecular data had consistently recovered monophyletic classes for the heterokont algae, however, the relationships among the classes had remained only partially resolved. Furthermore, earlier studies did not include representatives from all taxonomic classes. We used a five-gene (nuclear encoded SSU rRNA; plastid encoded *rbc*L, *psa*A, *psb*A, *psb*C) analysis with a subset of 89 taxa representing all 16 heterokont classes to infer a phylogenetic tree. There were three major clades. The Aurearenophyceae, Chrysomerophyceae, Phaeophyceae, Phaeothamniophyceae, Raphidophyceae, Schizocladiophyceae and Xanthophyceae formed the SI clade. The Chrysophyceae, Eustigmatophyceae, Pinguiophyceae, Synchromophyceae and Synurophyceae formed the SII clade. The Bacillariophyceae, Bolidophyceae, Dictyochophyceae and Pelagophyceae formed the SIII clade. These three clades were also found in a ten-gene analysis. The approximately unbiased test rejected alternative hypotheses that forced each class into either of the other two clades. Morphological and biochemical data were not available for all 89 taxa, however, existing data were consistent with the molecular phylogenetic tree, especially for the SIII clade.

Key words: Heterokont algae; molecular phylogenetic analysis; multigene phylogeny; reduced flagellar clade; stramenochromes.

Introduction

The photosynthetic stramenopiles (= heterokont algae or stramenochromes) are one of the most actively studied groups of protists, and 14 new taxonomic classes have been described since 1972. Yet despite this effort, the relationships among the classes remain only partially resolved (e.g., Riisberg et al. 2009). One major reason is that stramenopiles morphology is exceptionally diverse and it is often difficult or impossible to establish homologous characters. For example, the cellulose cell wall of brown algae, the siliceous frustule of diatoms, and the naked cells of chrysophytes cannot be directly compared. Furthermore, morphological features such as the silica frustules of diatoms, the silica scales of synurophytes and

---

[1]These authors contributed equally to this work.
[2]Corresponding authors; fax +1 207 633 9641.
e-mail smboo@cnu.ac.kr (S.M. Boo), raa48@uw.edu (R.A. Andersen), hsyoon@bigelow.org (H.S. Yoon).

doi:10.1016/j.protis.2011.08.001

the silica skeletons of silicoflagellates suggest that these three groups may be closely related. However, molecular phylogenetic analyses contradict this similarity (e.g., Riisberg et al. 2009). Even the primary synapomorphy for the stramenopiles, the tripartite flagellar hair, has exceptions. For example, *Pelagomonas* has bipartite flagellar hairs (Andersen et al. 1993) and two genera within the Pinguiophyceae lack flagellar hairs entirely (Kawachi et al. 2002). Finally, biochemical markers such as chloroplast pigmentation (Bjornland and Liaaen-Jensen 1989) have not resolved evolutionary relationships.

Molecular clock estimations place the origin of photosynthetic stramenopiles at 719 - 414 Ma (million years ago) based upon nuclear SSU rRNA (Brown and Sorhannus 2010) and the origin of all extant stramenopiles at 1077 – 1025 Ma based upon multi-plastid gene data (Yoon et al. 2004). Most molecular phylogenetic analyses show that the heterokont algae are a monophyletic group that is either derived from, or sister to, a clade of entirely nonphotosynthetic stramenopiles (e.g., Leipe et al. 1994; Moriya et al. 2002). Even though molecular studies have contributed significantly, most studies have been based on one or two genes (e.g., nuclear encoded SSU rRNA, plastid encoded *rbc*L), and the limited number of genes probably results in the lack of phylogenetic resolution for the deep branches. One notable study was based on seven-gene data (nuclear encoded LSU rRNA, SSU rRNA, actin, beta-tubulin, *hsp*90; mitochondrial encoded *cox*1; plastid encoded *rbc*L) from 35 taxa representing 10 of the 16 heterokont algal classes (Riisberg et al. 2009). Their study presented two strongly supported monophyletic groups [(Phaeophyceae + Xanthophyceae + Phaeothamniophyceae) and (Pelagophyceae + Dictyochophyceae), however, many relationships, such as the positions of the Chrysophyceae, Eustigmatophyceae, Pinguiophyceae and Raphidophyceae, remained unresolved. Using largely published sequences, Riisberg and colleagues found it necessary to combine sequences from different species, genera, families, and even from different orders (i.e., subclass level). Furthermore, these mixed taxonomic combinations had many missing data, up to 34% in one case. The sequences of one strongly supported monophyletic group (Pelagophyceae + Dictyochophyceae) lacked 13 – 28% of the total positions per taxon, particularly in the protein coding genes (i.e., 1 – 4 genes missing for the nuclear encoded actin, beta-tubulin, *cox*1, *hsp*90). These omissions may have impacted the monophyly of the two-class clade because there is strong phylogenetic signal from the nuclear rRNA nucleotides (LSU and SSU) for uniting the two classes (e.g., Ben Ali et al. 2002; Leipe et al. 1994).

In this paper, we focus on photosynthetic heterokonts (stramenochromes) so that we may include chloroplast genes. We provide results from a five-gene dataset (nuclear encoded SSU rRNA, plastid encoded *psa*A, *psb*A, *psb*C, *rbc*L) using 89 taxa representing all 16 currently recognized classes. For the first time, we present reasonably supported relationships for the photosynthetic heterokont algae in our unrooted phylogenies that exclude heterotrophic heterokonts and outgroups. We arbitrarily root the tree between two major groups, and although the term “convex group” (Estabrook 1978) or “clan”(Wilkinson et al. 2007) have been proposed for a group in unrooted trees, we use the term “clade” in this paper. Secondly, in the Supplementary Material, we used 10 genes (nuclear encoded LSU and SSU rRNA, actin, beta-tubulin, *hsp*90; plastid encoded *psa*A, *psb*A, *psb*C, *rbc*L; mitochondrial encoded *cox*1) with some combined taxa, and these results support our findings from the five-gene data.

Results

Two different datasets were used to determine relationships among the 16 heterokont algal classes. For the five-gene dataset, we generated 241 new sequences of nuclear SSU rRNA and plastid-encoded *psa*A, *psb*A, *psb*C, and *rbc*L from a subset of 89 heterokont taxa (Table 1). The same culture strain was used for determining all five genes in every case but five species where we combined publicly available sequences from different strains of the same species (see Fig. 1 and Table 1). All new sequences have been deposited in GenBank (accession numbers HQ710550–HQ710794). For the ten-gene dataset, the mitochondrial *cox*1 and the nuclear LSU rRNA, actin, beta-tubulin, and *hsp*90 genes were added to the five-gene dataset, and these sequences were obtained from GenBank following Riisberg et al. (2009). All sequences were combined under a higher taxonomic rank (e.g., genus, family, order), when ten-gene sequences from same strain or species were not available (Supplementary Table S1).

A maximum likelihood phylogeny using a concatenated five-gene dataset (1548 amino acids + 1362 rRNA nucleotides) recovered three strongly supported clades: SI, SII and SIII