Protist, Vol. xx, xxx–xxx, xx 2011
http://www.elsevier.de/protis
Published online date 11th October 2011

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

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), 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

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001

Table 1. List of taxa and GenBank accession numbers. GenBank accessions for new sequences are in bold. Underlined species and culture strains indicate that the data analysis combined sequences from different strains of the same species or from different species.

Taxa	Culture strain	ID	SSU	*rbc*L	*psa*A	*psb*A	*psb*C
Aurearenophyceae							
Aurearena cruciata Kai et al.	NIES-1863	Aur01	AB365192	AB365193	**HQ710635**	**HQ710692**	-
	NIES-1864	Aur02	AB365194	AB365195	**HQ710636**	**HQ710693**	**HQ710747**
Aurearena sp.	CCMP 1538	Aur03	**HQ710550**	**HQ710590**	**HQ710637**	**HQ710694**	**HQ710748**
	CCMP 1618	Aur04	**HQ710551**	**HQ710591**	**HQ710638**	**HQ710695**	**HQ710749**
Bacillariophyceae							
Asterionellopsis glacialis Castracane	KMCC-B-296	Bac01	**HQ710552**	**HQ710592**	**HQ710639**	**HQ710696**	**HQ710750**
Bacterosira bathyomphala (Gran) Syvertsen & Hasle	NB04-B6	Bac02	DQ514894	DQ514816	-	-	DQ514734
Chaetoceros didymus Ehrenberg	KMCC-B-170	Bac03	**HQ710553**	**HQ710593**	**HQ710640**	**HQ710697**	**HQ710751**
Phaeodactylum tricornutum Bohlin	CCAP 1055/1	Bac04	-	EF067920	EF067920	EF067920	EF067920
	CS-29	Bac05	EF140622	-	-	-	-
Skeletonema marinoi Sarno & Zingone	SZN B 211	Bac06	DQ396524	-	-	-	-
	CCMP 1332	Bac07	-	AF015569	AY119725	AY119761	AY876221
Thalassiosira conferta Hasle	KMCC-B-020	Bac08	**HQ710554**	**HQ710594**	**HQ710641**	**HQ710698**	**HQ710752**
Thalassiosira pseudonana Hasle & Heimdal	CCMP 1335	Bac09	AY485452	NC_008589	NC_008589	NC_008589	NC_008589
Bolidophyceae							
Bolidomonas mediterranea Guillou & Chrétiennot-Dinet	CCMP 1867	Bol01	**HQ710555**	AF333977	**HQ710642**	**HQ710699**	-
Bolidomonas pacifica Guillou & Chrétiennot-Dinet	CCMP 1866	Bol02	AF123595	AF372696	**HQ710643**	**HQ710700**	**HQ710753**
Chrysomerophyceae							
Chrysowaernella hieroglyphica (Waern) Gayral & Billard	K 0368	Chm01	**HQ710556**	**HQ710595**	**HQ710644**	-	**HQ710754**
Giraudyopsis sp.	CCMP 1666	Chm02	**HQ710557**	**HQ710596**	**HQ710645**	**HQ710701**	**HQ710755**
Giraudyopsis stellifera Dangeard	CCMP 1308	Chm03	U78034	**HQ710597**	**HQ710646**	**HQ710702**	-
Chrysophyceae							
Chromulina nebulosa Cienkowski	CCMP 262	Chr01	-	-	**HQ710647**	**HQ710703**	**HQ710756**
	CCMP 263	Chr02	AF123285	AF155876	-	-	-
Hibberdia magna (Belcher) Andersen	CCMP 453	Chr03	**HQ710558**	AF015572	**HQ710648**	**HQ710704**	**HQ710757**
Ochromonas distigma nom. nud.	AC 25	Chr04	EF165136	EF165177	-	EU851959	-
Ochromonas tuberculata Hibberd	CCMP 1861	Chr05	AF123293	**HQ710598**	**HQ710649**	**HQ710705**	-
Dictyochophyceae							
Apedinella spinifera Throndsen	CCMP 1767	Dic01	**HQ710559**	**HQ710599**	**HQ710650**	**HQ710706**	**HQ710758**
Dictyocha speculum Ehrenberg	CCMP 1381	Dic02	U14385	AY043280	**HQ710651**	**HQ710707**	**HQ710759**
Pseudopedinella elastica Skuja	CCMP 716	Dic03	**HQ710560**	**HQ710600**	**HQ710652**	**HQ710708**	**HQ710760**
Pseudopedinella sp.	CCMP 1476	Dic04	**HQ710561**	**HQ710601**	**HQ710653**	**HQ710709**	

Table 1 (*Continued*)

Taxa	Culture strain	ID	SSU	*rbc*L	*psa*A	*psb*A	*psb*C
	CCMP 3052	Dic05	**HQ710562**	**HQ710602**	**HQ710654**	-	
Rhizochromulina marina Hibberd & Chretiennot-Dinet	CCAP 950/1	Dic06	**HQ710563**	**HQ710603**	**HQ710655**	**HQ710710**	**HQ710761**
Rhizochromulina sp.	CCMP 237	Dic07	U14388	**HQ710604**	**HQ710656**	**HQ710711**	-
	CCMP 1243	Dic08	-	**HQ710605**	**HQ710657**	**HQ710712**	-
Eustigmatophyceae							
Eustigmatos magnus (Petersen) Hibberd	CCMP 387	Eus01	U41051	AF015575	-	-	-
Microtalis aquatica J.C. Bailey	CCMP 3153	Eus02	**HQ710564**	**HQ710606**	**HQ710658**	**HQ710713**	**HQ710762**
Microtalis reticulata J.C. Bailey	CCMP 3154	Eus03	-	**HQ710607**	**HQ710659**	-	-
Monodus unipapilla Reisigl	SAG 8.83	Eus04	**HQ710565**	**HQ710608**	-	**HQ710714**	**HQ710763**
Nannochloropsis oculata (Droop) Hibberd	CCMP 525	Eus05	**HQ710566**	**HQ710609**	**HQ710660**	**HQ710715**	**HQ710764**
Nannochloropsis salina Hibberd	EUS-001	Eus06	**HQ710567**	**HQ710610**	**HQ710661**	**HQ710716**	-
Nannochloropsis sp.	EC-009	Eus07	**HQ710568**	**HQ710611**	**HQ710662**	**HQ710717**	-
Vischeria helvetica (Vischer & Pascher) Hibberd	UTEX 49	Eus08	**HQ710569**	**HQ710612**	**HQ710663**	**HQ710718**	-
Vischeria stellata (Chodat) Pascher	SAG 33.83	Eus09	**HQ710570**	**HQ710613**	**HQ710664**	**HQ710719**	-
Pelagophyceae							
Ankylochrysis lutea (van der Veer) Billard	RCC-286	Pel01	**HQ710571**	**HQ710614**	**HQ710665**	**HQ710720**	**HQ710765**
Aureococcus anophagefferens Hargraves & Sieburth	CCMP 1984	Pel02	**HQ710572**	**HQ710615**	**HQ710666**	**HQ710721**	**HQ710766**
Aureoumbra lagunensis Stockwell et al.	CCMP 1507	Pel03	**HQ710573**	NC_012903	NC_012903	NC_012903	NC_012903
	CCMP 1510	Pel04	**HQ710574**	**HQ710616**	**HQ710667**	**HQ710722**	**HQ710767**
Chrysoreinhardia giraudii (Derbès & Solier) C. Billard	CCMP 2349	Pel05	**HQ710575**	**HQ710617**	**HQ710668**	**HQ710723**	**HQ710768**
Pelagococcus subviridis Norris	CCMP 1429	Pel06	U14386	**HQ710618**	-	**HQ710724**	**HQ710769**
Pelagomonas calceolata Andersen & Saunders	CCMP 1214	Pel07	U14389	**HQ710619**	**HQ710669**	**HQ710725**	**HQ710770**
Sarcinochrysis sp.	A11,551	Pel08	**HQ710576**	**HQ710620**	**HQ710670**	**HQ710726**	**HQ710771**
	A11,864	Pel09	**HQ710577**	**HQ710621**	**HQ710671**	-	**HQ710772**
	CCMP 770	Pel10	U78033	**HQ710622**	**HQ710672**	**HQ710727**	**HQ710773**
	CCMP 1664	Pel11	-	**HQ710623**	**HQ710673**	**HQ710728**	**HQ710774**
Phaeophyceae							
Choristocarpus tenellus (Kützing) Zanardini	-	Pha01	AB252658	AJ287862	**HQ710674**	**HQ710729**	**HQ710775**
Desmarestia viridis (Müller) Lamouroux	CNUK PD005	Pha02	AJ295828	**HQ710624**	**HQ710675**	**HQ710730**	**HQ710776**
Fucus vesiculosus Linnaeus	CNUK PF002	Pha03	**HQ710578**	AF195515	AY372960	AY528845	**HQ710777**
Ishige okamurae Yendo	CNUK PE006	Pha04	AY232602	AY372974	AY372944	AY528829	**HQ710778**
Laminaria digitata (Hudson) Lamouroux	CNUK PL003	Pha05	AF091286	AY372984	AY372964	AY528849	**HQ710779**
Sphacelaria divaricata Montagne	Jejudo: Korea	Pha06	-	AY372985	AY372970	AY528855	**HQ710780**
Sphacelaria sp.	UTEX LB800	Pha07	AY307401	-	-	-	-

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001

Phaeothamniophyceae							
Phaeoschizochlamys mucosa Lemmermann	CCMP 635	Phm01	**HQ710579**	AF064747	**HQ710676**	**HQ710731**	-
Phaeothamnion confervicola Lagerheim	CCMP 637	Phm02	AF044846	AF064746	**HQ710677**	**HQ710732**	-
Pinguiophyceae							
Glossomastix chrysoplasta O'Kelly	CCMP 1537	Pin01	AF438325	AF438318	**HQ710678**	**HQ710733**	**HQ710781**
Phaeomonas parva Honda & Inouye	CCMP 2877	Pin02	AB042204	AF438321	**HQ710679**	**HQ710734**	**HQ710782**
Pinguiochrysis pyriformis Kawachi	PP 301	Pin03	**HQ710580**	**HQ710625**	**HQ710680**	**HQ710735**	**HQ710783**
Pinguiococcus pyrenoidosus Andersen et al.	CCMP 1144	Pin04	AF438324	AF438319	**HQ710681**	**HQ710736**	**HQ710784**
Polypodochrysis teissieri Magne	A9606	Pin05	**HQ710581**	**HQ710626**	**HQ710682**	**HQ710737**	**HQ710785**
Raphidophyceae							
Chattonella marina (Subrahmanyan) Hara & Chihara	CCMP 2049	Rap01	**HQ710582**	**HQ710627**	**HQ710683**	**HQ710738**	**HQ710786**
Chattonella subsalsa Biecheler	CCMP 217	Rap02	**HQ710583**	**HQ710628**	**HQ710684**	**HQ710739**	-
Haramonas dimorpha Horiguchi	NIES-716	Rap03	AB365025	AB280608	AB367952	-	-
Heterosigma akashiwo (Hada) Hada	CCMP 452	Rap04	**HQ710584**	**HQ710629**	**HQ710685**	**HQ710740**	**HQ710787**
	CCMP 1595	Rap05	**HQ710585**	**HQ710630**	**HQ710686**	**HQ710741**	**HQ710788**
	NIES-293	Rap06	DQ470658	NC_010772	NC_010772	NC_010772	NC_010772
Schizocladiophyceae							
Schizocladia ischiensis Henry et al.	CCMP 2287	Sch01	AB085614	AB085615	AY528863	AY528859	**HQ710789**
Synchromophyceae							
Synchroma grande Schnetter	Spain	Syc01	DQ788730	DQ788731	-	-	-
Synurophyceae							
Mallomonas caudata Ivanov	AKC	Syn01	EF469638	**HQ710631**	**HQ710687**	**HQ710742**	**HQ710790**
Synura petersenii Korshikov	KNU 01	Syn02	**HQ710586**	**HQ710632**	**HQ710688**	**HQ710743**	**HQ710791**
Xanthophyceae							
Botrydiopsis callosa Trenkwalder	SAG 30.83	Xan01	AJ579340	AJ579569	AM421008	-	-
Botrydiopsis constricta Broady	Antarctica	Xan02	AJ579339	AJ579566	AM421009	-	-
Botrydium granulatum (Linnaeus) Greville	NIES-622	Xan03	**HQ710587**	**HQ710633**	**HQ710689**	**HQ710744**	**HQ710792**
Bumilleriopsis filiformis Vischer	SAG 809-2	Xan04	AF083398	U89900	AJ877110	X79223	-
Chlorellidium tetrabotrys Pascher & Vischer	SAG 5.90	Xan05	AJ580949	AJ580947	AJ878067	-	-
Heterococcus caespitosus Vischer	SAG 835-2a	Xan06	AM490820	AM421002	AM421013	-	-
Heterococcus chodatii Vischer	SAG 835-3	Xan07	AM490822	AM421003	AM421014	-	-
Heterococcus fuornensis Vischer	SAG 835-5	Xan08	AM490821	AY682397	AM422008	-	-
Mischococcus sphaerocephalus Vischer	UTEX B150	Xan10	AF083400	AF064744	AM422009	-	-
Pseudopleurochloris antarctica Andreoli et al.	SAG 39.98	Xan11	AF109728	AJ580924	AM422010	-	-

Table 1 (*Continued*)

Taxa	Culture strain	ID	SSU	*rbc*L	*psa*A	*psb*A	*psb*C
Tribonema aequale Pascher	UTEX B50	Xan12	**HQ710588**	AF084611	AY372972	AY528860	**HQ710793**
Vaucheria bursata (Müller) C. Agardh	CCMP 1084	Xan13	-	-	**HQ710690**	**HQ710745**	**HQ710794**
	UTEX 761	Xan14	-	AF476940	-	-	-
Vaucheria litorea C. Agardh	-	Xan15	-	NC_011600	NC_011600	NC_011600	NC_011600

Abbreviations for public institutions are defined as follows: AC = Algobank Caen (France); CCAP = Culture Collection of Algae and Protozoa (UK); CCMP = Provasoli-Guillard National Center for Culture of Marine Phytoplankton (USA); CNUK = herbarium, Chungnam National University (Korea); CS = Australia National Algae Culture Collection (Australia); K = Scandanavian Culture Center for Algae and Protozoa (Denmark); KMCC = Korean Microbial Culture Collection (Korea); NIES = Microbial Culture Collection at the National Institute for Environmental Studies (Japan); RCC = Roscoff Culture Collection (France); SAG = Sammlung von Algenkulturen Göttingen (Germany); SZN = Stazione Zoologica Anton Dohrn of Naples (Italy); UTEX = Culture Collection of Algae at the University of Texas (USA). Private abbreviations are defined as follows: A = Robert A. Andersen strain number; PAB = Paul A. Broady strain number; PP = Masanobu Kawachi strain number. EC, EUS, N, ACK, NB04 are unspecified abbreviations.

clades (Fig. 1). The tree was arbitrarily rooted between the SI/SII and SIII clades. The SI clade was comprised of the PX clade sensu Kai et al. (2008) (i.e., Aurearenophyceae, Chrysomerophyceae, Phaeophyceae, Phaeothamniophyceae, Schizocladiophyceae, Xanthophyceae) as well as the Raphidophyceae. The SII clade included a monophyletic group of three classes (the Chrysophyceae, Synurophyceae, Synchromophyceae; C/S/S clade), plus the Eustigmatophyceae and Pinguiophyceae. The SIII clade was comprised of the Bacillariophyceae and Bolidophyceae (B/B clade) as well as the Dictyochophyceae and Pelagophyceae.

The SI clade was supported by 82% ML bootstrap value (MLB) and 1.0 Bayesian Posterior -Probability (BPP). Within the PX clade, three sub-clades emerged: the Phaeophyceae + Schizocladiophyceae (MLB 99%, BPP 1.0), Phaeothamniophyceae + Aurearenophyceae (MLB 54%, BPP 0.99), and Xanthophyceae + *Chrysowaernella* (MLB 88%, BPP 1.0). However, the two Chrysomerophyceae taxa, *Chrysowaernella* and *Giraudyopsis,* did not form a monophyletic group.

Within the SII clade (MLB 85%, BPP 1.0), the Pinguiophyceae diverged first followed by the monophyletic clade of the C/S/S clade and the Eustigmatophyceae (MLB 71%, BPP 1.0). Within the C/S/S clade, the Synurophyceae and Chrysophyceae were grouped together (MLB 100%, BPP 1.0), however internal relationships were unresolved. *Synchroma* (Synchromophyceae) was grouped together with the Chrysophyceae/Synurophyceae clade (MLB 98%, BPP 1.0).

The ML tree showed a monophyletic SIII clade (95% MLB, 1.0 BPP) including the monophyly of the Bacillariophyceae and Bolidophyceae (MLB 100%, BPP 1.0), whereas the monophyly of the Dictyochophyceae and Pelagophyceae is not statistically supported (MLB 38%, BPP 0.95). The two genera *Pseudopedinella* (Dictyochophyceae) and *Thalassiosira* (Bacillariophyceae) showed paraphyletic relationships.

Analysis of the ten-gene dataset (2847 amino acids + 3893 nuclear LSU and SSU rRNA nucleotides) provided a topology congruent with the five-gene phylogeny (Supplementary Fig. S1). The SI, SII and SIII clades were recovered as in the five-gene phylogeny (MLB 78%, 67%, 91%, respectively). Within the SI clade, there was significant improvement in ML bootstrap supports for the monophyly of Aurearenophyceae and Phaeothamniophyceae (from 54% to 78%). Also, the

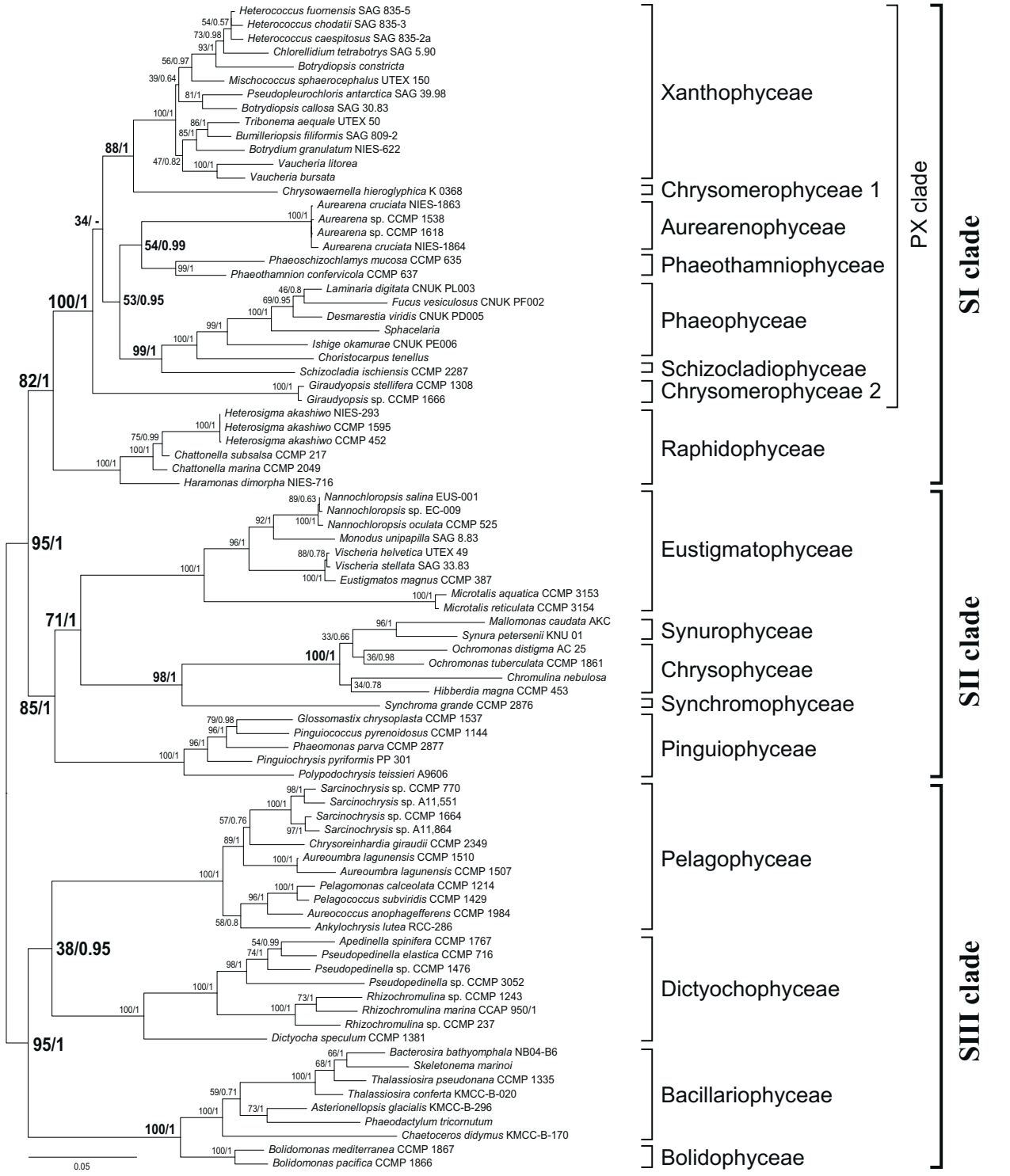


Figure 1. Maximum likelihood tree of the concatenated five-gene dataset of nuclear SSU rRNA, *psa*A, *psb*A, *psb*C, and *rbc*L inferred using RAxML. Separate models for each partition (GTR + G for nuclear SSU rRNA, LG + G for *rbc*L, MtArt + F + G for *psa*A, LG + F + G for *psb*A, and MtArt + F + G for *psb*C protein partitions) were used for the best fitting evolutionary models under the weighted Akaike information criterion (AICc). Ambiguously aligned regions were treated as missing for nuclear rRNA and treated as gaps for protein sequences. Bootstrap support values using RAxML and Bayesian posterior probability are shown near the nodes.

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001

Phaeophyceae + Schizocladophyceae clade was sister to the rest PX-clade (MLB 100%, BPP 1.0), whereas its position was previously unresolved. Within the SII clade, the Chrysophyceae became weakly monophyletic (MLB 53%, BPP 1.0) and showed a sister relationship to the Synurophyceae. Within the SIII clade, the sister relationship of Dictyochophyceae to Pelagophyceae had more robust bootstrap support (up to MLB 85%).

Single gene trees generally supported the monophyly of each class except for the Chrysomerophyceae and Chrysophyceae (Supplementary Figs S2-S6). With the exception of plastid encoded *rbc*L (Supplementary Fig. S3), single gene trees usually failed to recover SI, SII and SIII clades. For example, the nuclear SSU rRNA tree supported the SI and SIII clades, whereas the SII clade was not recovered (Supplementary Fig. S2). Some partially combined datasets recovered the three major clades, e.g., five-gene data minus the nuclear SSU rRNA (only *rbc*L+*psa*A+*psb*A+*psb*C), five-gene data minus *rbc*L (only nuclear SSU rRNA+*psa*A+*psb*A+*psb*C), and five-gene data minus *rbc*L and nuclear SSU rRNA (only *psa*A+*psb*A+*psb*C) supported the grouping of the SI, SII and SIII clades (Supplementary Figs S7, S8, S9, respectively). The tree formed with the ten-gene database minus the nuclear LSU/SSU rRNA data (Supplementary Fig. S10) was basically congruent with the five-gene tree (Fig. 1). However, *Synchroma* (Synchromophyceae) was grouped together with the Eustigmatophyceae (MLB 59%) and the clade of chrysophytes/synurophytes was grouped with the Pinguiophyceae (MLB 50%). When the data included only the four nuclear and mitochondrial protein data (actin+beta-tubulin+*hsp*90+*cox*1; Supplementary Fig. S11) or only the six nuclear and mitochondrial genes (LSU rRNA+SSU rRNA+actin+beta-tubulin+*hsp*90+*cox*1; Supplementary Fig. S12), then the three clades were not recovered except for the SIII clade in six-gene analysis (MLB 28%).

To test alternative hypotheses of internal relationships among the major lineages, we used the approximately unbiased (AU) test (Shimodaira 2002). Eight major lineages (1 = PX clade, 2 = Raphidophyceae, 3 = Eustigmatophyceae, 4 = C/S/S clade, 5 = Pinguiophyceae, 6 = Pelagophyceae, 7 = Dictyochophyceae, and 8 = B/B clade) were clearly recognized as strongly supported monophyletic groups (>98%). Because the relationships among these eight groups were of primary interest for us, we forced each lineage into all possible alternative positions in the five-gene ML tree (i.e., alternative hypotheses), and we tested the statistical difference of the null hypothesis (i.e., best ML tree) to an alternative hypothesis (Fig. 2a-h). For example, 10 alternative hypotheses were generated each time a test-clade (in black boxes in Fig. 2) was moved into the six tip nodes and four internodes. Some alternative hypotheses corresponded to previously reported topologies. For example, the movement of Eustigmatophyceae from SII clade to the sister relationship of PX clade (SI clade) corresponded to Figure 3 in Riisberg et al. 2009 (see (1) in Fig. 2c). Similarly, placing the C/S/S clade as a sister to the Pelagophyceae agreed with Grant et al. 2008 (see (2) in Fig. 2d), and placing it as sister to SI + SII clades agreed with our topology based on the nuclear SSU rRNA phylogeny (see (4) in Fig. 2d). Finally, forcing the Pinguiophyceae to form a monophyletic group with the Raphidophyceae reflected Figure 18 in Kai et al. 2008 (see (3) in Fig. 2e), and moving the Pelagophyceae to be the sister group of B/B clade reflected our *rbc*L phylogeny (see (5) in Fig. 2f). The AU test results were indicated on every branch in Figure 2. Asterisk(s) were used to indicate a significant rejection of the alternative hypothesis, and P-values (above the node) were used when the alternative hypotheses were not significantly different (alpha = 0.05) to the null hypothesis. It was noteworthy that AU test significantly rejected most alternative hypotheses where one lineage was moved to the nodes of the other two major clades. For example, all trees were significantly rejected when the PX or Raphidophyceae of SI clade was moved to the nodes of SII or SIII (P value < 0.05, Fig. 2a-b). Similar results were obtained for the lineages of SII clade (Fig. 2c-e) and SIII clade (Fig. 2f-h), except when the Pinguiophyceae was placed as the sister group of the Raphidophyceae in SI clade (Fig. 2e, P = 0.122). In contrast, reshuffling the Pinguiophyceae within the SII clade could not be rejected.

To test topologies from individual genes, we conducted another set of AU tests (Supplementary Fig. S13). Single gene topologies from *psa*A (P = 4e-04), *psb*A (P = 1e-05), and *psb*C (P = 3e-04) were significantly rejected, but the *rbc*L (P = 0.382) and nuclear SSU rRNA (P = 0.083) trees were not rejected.

Discussion

Phylogeny of the Heterokont Algae

For the first time, we provide a resolved phylogeny with high statistical branch supports (MLB, BPP)

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001

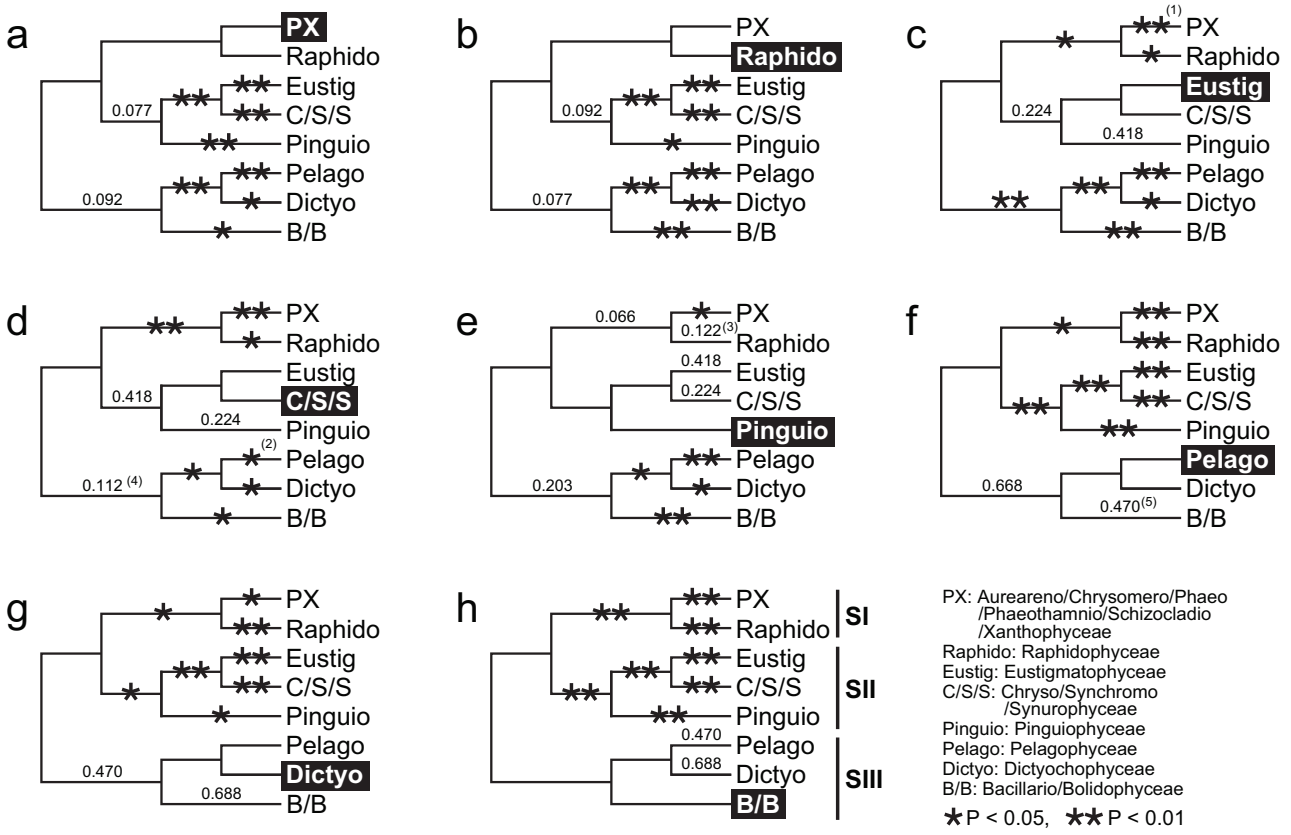
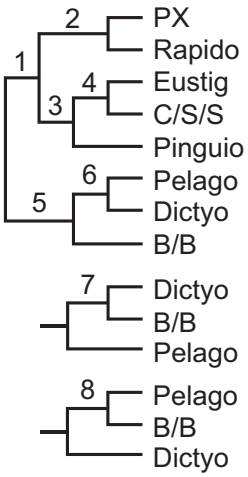


Figure 2. Alternative hypotheses that were tested using the approximately unbiased (AU) test. Based on the five-gene tree, a total of 80 alternative tree topologies were made by moving the target clade (in black boxes) to all possible tip nodes and internodes (a–h). The AU test results are indicated above the nodes. An asterisk indicates a statistically significant rejection of the alternative hypothesis; one asterisk 95% confidence level and two asterisks 99% confidence level. A number (P value) above the node indicates that the alternative hypothesis was not significantly different (< 95% level) from the null hypothesis. The numbers in parentheses (1-5) label alternative topologies that corresponded to previous studies using nuclear SSU rRNA and rbcL gene trees: (1) Riisberg et al. 2009, (2) Grant et al. 2009, (3) Kai et al. 2008, (4) our nuclear SSU rRNA tree, and (5) our *rbc*L tree. See the text for details.

that show the deep relationships (clades SI, SII, SIII) among the photosynthetic heterokont algae. The AU test shows that alternative topologies are unlikely, i.e., most alternative topologies (Fig. 2) were significantly rejected when one lineage of a major clade was moved to another major clade. These results provide evidence for the monophyly of the SI, SII and SIII clades. The SIII clade was weakly recovered in an earlier 10-class phylogeny (Riisberg et al. 2009) and it was also recovered in our ten-gene tree (Supplementary Fig. S1). Furthermore, recently published plastid genome data have led to mega-gene data analyses (137 to 148 genes). Using plastid genomes of *Fucus vesiculosus* and *Ectocarpus siliculosus*, Le Corguille et al. (2009) suggested a sister relationship for these phaeophycean taxa and *Heterosigma akashiwo* (Raphidophyceae) (SI clade) that separated them from diatom species (SIII clade). Using two published pelagophycean plastid genomes from *Aureococcus anophagefferens* and *Aureoumbra lagunensis* (Ong et al. 2010), Janouskovec et al. (2010) presented two well supported monophyletic groups of the Xanthophyceae + Phaeophyceae + Raphidophyceae (RAxML 97%, SI clade) and the Pelagophyceae + Bacillariophyceae (RAxML 76%, SIII clade). Although they used a rich dataset of 34 conserved plastid genes, their taxon sampling was very limited (nine species from five classes). Nevertheless, both plastid genome data analyses are consistent with our result that the SIII clade contains four classes and was separated from the remaining 12 classes (SI and SII clades).

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001



node	5 gene data (- minus)				10 gene data			single gene	
	All	-SSU	-*rbc*L	-SSU&*rbc*L	All	-rRNAs	nu/mt	SSU	*rbc*L
1	95	90	91	75	91	85	28	37	42
2	82	64	81	49	78	53	-	42	45
3	85	67	55	48	67	56	-	-	41
4	71	46	57	37	58	-	-	-	73
5	95	90	91	75	91	85	28	37	42
6	38	-	62	53	85	49	100	46	-
7	-	41	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	77
Fig.	1	S7	S8	S9	S1	S10	S12	S2	S3

Figure 3. Summary of bootstrap supports for major deep nodes from different datasets. Major deep relationships were indicated as nodes 1 – 7 in the five-gene tree and in trees from sub-dataset. Bootstrap support values were compiled from each analysis when the rRNA and/or *rbc*L genes were excluded from the five-gene and ten-gene datasets, as well as from six nuclear and mitochondrial genes excluding plastid genes from ten-gene dataset. Bootstrap values from the two single gene datasets (SSU rRNA, *rbc*L) were included for comparison. The SIII clade and its internal relationships (nodes 5 – 7) are highlighted in grey.

In our study, the multigene analysis, the increased taxon sampling, and the nearly complete restriction of sampling to a single strain of a species are probably important reasons why we have recovered a robust framework for the photosynthetic heterokont algae. There are still some mis-identified cultures/sequences and there are cryptic species complexes that have different genetic divergences (e.g., *Synura petersenii*, see details in Boo et al. 2010). We expect that future studies will provide additional genes/genomes as well as improved taxon sampling, and these should increase support for the broad framework we have established. This pattern has been evident throughout the history of molecular phylogenetic studies, beginning with the early reports that placed the Phaeophyceae and Xanthophyceae together (Ariztia et al. 1991) to recent papers that identified the PX clade (Kai et al. 2008; Riisberg et al. 2009).

Conflict in Datasets

There have been reports where nuclear rRNA trees place the Bacillariophyceae as a sister group to all other photosynthetic heterokont algal phylogenies (e.g., Ben Ali et al. 2002; Leipe et al. 1994; Riisberg et al. 2009) whereas the *rbc*L trees place the Dictyochophyceae at the root (e.g., Horn et al. 2007). Because of these conflicting hypotheses for the early divergence of SIII classes in outgroup-rooted heterokont phylogenies, it becomes an interesting evolutionary question. We suggest that this inconsistency may be caused by conflict in some genes. In our multigene analyses, three clades (SI, SII, and SIII) were clearly identified (see nodes 2, 3, 5, respectively in Fig. 3). However, despite the identification of the SIII clade, the internal relationships for the SIII are not conclusive (nodes 6, 7, 8 in Fig. 3). For example, our unrooted nuclear SSU rRNA tree (Supplementary Fig. S2) shows a monophyletic relationship for the Pelagophyceae and Dictyochophyceae that in turn grouped together with the B/B clade with low boothstrap support (46% and 37%, respectively). Alternatively, in our *rbc*L tree (Supplementary Fig. S3), the Dictyochophyceae (MLB 42%) was sister to the B/B + Pelagophyceae lineage (MLB 77%). In contrast, *psa*A, *psb*A, and *psb*C genes did not have strong support for any deep branches (Supplementary Figs S4-S6). Therefore, we tested the impact of the nuclear SSU rRNA and the *rbc*L gene by excluding each gene from the five-gene dataset (see Fig. 3). When the *rbc*L sequence was excluded, the tree topology was consistent with the five-gene tree except for the position of *Synchroma*, which had only the SSU rRNA and *rbc*L sequences in the original dataset (Supplementary Fig. S7). A similar topology was also recovered when both the nuclear SSU rRNA and the *rbc*L gene were excluded, although there was lower bootstrap support, which may be caused by the reduced data size (Supplementary Fig. S9). On the other hand, a four plastid gene dataset, without the nuclear SSU rRNA but with the *rbc*L gene,

supported the Pelagophyceae diverged first and monophyly of the Dictyochophyceae and B/B clade within the SIII clade (Supplementary Fig. S7). These analyses suggest that the *rbc*L data contains a strong signal to support the "Dictyochophyceae - early" hypothesis, whereas the nuclear SSU rRNA and core photosystem genes (*psa*A, *psb*A, *psb*C) supports the "Bacillariophyceae + Bolidophyceae clade-early" hypothesis for heterokont algal evolution. After these analyses, we next excluded the nuclear SSU and LSU rRNA from the ten-gene data (Supplementary Fig. S10). Interestingly, this eight-gene tree (without both rRNA sequences) was basically congruent with the five-gene tree (Fig. 1) and three-gene tree (*psa*A, *psb*A, *psb*C, Supplementary Fig. S9). Finally, we excluded all plastid genes in addition to the nuclear SSU and LSU rRNA genes, and therefore used only the actin, beta-tubulin, *hsp*90 and *cox*1 sequences to test for plastid gene bias (Supplementary Fig. S11). The tree based on four nuclear and mitochondrial genes, however, did not provide resolution among the classes. The PX clade was weakly supported (MLB 45%), but the B/B and C/S/S clades were not recovered as monophyletic groups; even class-level monophyly was not supported (e.g., Eustigmatophyceae, Raphidophyceae, Xanthophyceae). This was probably caused by insufficient phylogenetic information and limited taxon sampling. Similar results were recovered when all plastid genes were excluded (Supplementary Fig. S12); the SI and SII clades were not recovered and the SIII clade had low bootstrap support (MLB 28%).

Problems associated with the analysis of incongruent genes have been known for some time, and there are methods designed to address these problems (e.g., Chung and Ané 2011; Cranston et al. 2009; Page and Charleston 1997; Sang and Zhong 2000). Implementation of these methods is beyond the scope of this manuscript, but our thorough examination of single genes and combinations of genes suggests that incongruent genes are not a widespread problem within the heterokont algae. Furthermore, there is no suspicion for lateral gene transfer within the heterkont algae, and as far as is known, the incongruence is due to differential evolutionary rates for one or two genes.

We conclude that it is highly likely that the monophyletic groups of the SI, SII and SIII clades, as well as the clade of the combined SI + SII, are true affiliations. All multigene datasets supported these monophyletic relationships, either with or without nuclear rRNA sequences. Secondly, the internal deep branching within the SIII clade is the least supported of our results, apparently because of the conflict in the *rbc*L data. Although the three-gene dataset (without *rbc*L and nuclear rRNA sequences) supported the monophyly of the Pelagophyceae/Dictyochophyceae clade as sister to the B/B clade, the statistical support was relatively low (53%). Additional data may strengthen this relationship, e.g., genomic data (e.g., plastid genomes, transcriptomes) would be applicable for this purpose.

Evolutionary Scenarios of the Stramenopiles

There is morphological evidence to support the SIII clade. For example, the Bacillariophyceae, Dictyochophyceae and Pelagophyceae were previously identified as the reduced flagellar apparatus clade based on a combined molecular and morphological dataset (Saunders et al. 1995). In that study, all included taxa lacked flagellar microtubular roots and the swimming cells had only one emergent flagellum. The basal bodies were attached directly to an anteriorly displaced nucleus, which anchored the flagellum to the cell body. Two years later, a molecular phylogenetic analysis showed that the Sarcinochrysidales belonged within the Pelagophyceae rather than within the Chrysophyceae (Saunders et al. 1997). *Sarcinochrysis* has four microtubular roots, much like taxa in the SI and SII clades (Andersen 1991; O'Kelly 1989). Also following the Saunders et al. (1995) work, numerous phylogenetic analyses, based largely on rRNA data, did not robustly support the reduced flagellar clade. Therefore, the reduced flagellar group has not been widely accepted.

Cavalier-Smith tried to reconcile the rRNA data (with the B/B clade at the base of the heterokont algae) with some reduced flagellar features when he coined the name Khakista (organisms with girdle lamellae and annular chloroplast DNA but without transitional helix and flagellar roots) (Cavalier-Smith 2000; Cavalier-Smith and Chao 2006). However, this classification separated the B/B clade from the Dictyochophyceae and Pelagophyceae; he placed the latter two classes in the Hypogyrista, which in turn was placed in a distinctly different super group, the Phaeista. The distinction between the Khakista and Hypogyrista was based largely upon the occurrence of a proximal transitional helix in the flagellar/basal body complex of the Hypogyrista.

We now reconsider the morphological data because our analysis recovered the SIII clade, which includes both the Khakista and Hypogyrista, and because the SIII clade is equal to the reduced

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001

flagellar clade plus the Sarcinochrysidales (e.g., Potter et al. 1997; Saunders et al. 1995, 1997). We hypothesize that the ancestor of the SIII clade had four microtubular roots because this is the common arrangement not only in heterokont algae, but also nonphotosynthetic stramenopiles, dinoflagellates, haptophytes, and other groups (Andersen 1991; Moestrup 2000; Moriya et al. 2002). Once the microtubular roots were lost and the flagellum became anchored on the nuclear envelope (B/B clade, Dictyochophyceae, Pelagomonadales of the Pelagophyceae), we assume that there was not an evolutionary reversal back to the four microtubular root system, agreeing with Boddi et al. (1999). This leads us to hypothesize that the Sarcinochrysidales (Pelagophyceae) are the early-diverged extant group of the SIII clade, even though this position is not supported by our five-gene trees (but see Fig. 1; MLB 38% between the Pelagophyceae and Dictyochophyceae). It is noteworthy that the AU test alternative hypothesis that repositioned the Pelagophyceae to the base of the SIII clade was not rejected (P = 0.668), and a four-gene analysis without nuclear SSU rRNA and *rbc*L phylogenies support the scenario of an early diverging Pelagophyceae. Our molecular analyses support Boddi et al. (1999) in that *Ankylochrysis* is the earliest diverging taxon of the Sarcinochysidales. Therefore, we hypothesize an evolutionary reduction of the flagellar apparatus in the SIII clade. The Bacillariophyceae, which has more species than any other algal class, continued the flagellar reduction and there is now no evidence of a flagellum in the pennate forms. We also suggest that additional taxon sampling and more molecular data will increasingly provide support for a well-developed flagellar apparatus in the early-diverged taxa of the SIII clade.

The carotenoid composition of the heterokont algae may also provide a glimmer of past evolutionary changes. Many of the heterokont algae contain the diatoxanthin-diadinoxanthin cycle (D-D cycle), which is common in other chlorophyll *c* containing groups such as the haptophytes and dinoflagellates (Bjornland and Liaaen-Jensen 1989). Assuming that the chromalveolate hypothesis is correct, then one also assumes that the ancestral heterokont algae had the D-D cycle. Taxa in the SIII clade have only this carotenoid cycle, as do some taxa in the SI clade. The SII clade is defined by taxa with only the violaxanthin-antheraxanthin cycle (V-A cycle). Interestingly, in the SI clade, we find some taxa with two carotenoid cycles. *Chrysowaernella* (Chrysomerophyceae; pers. comm., Robert Bidigare) and *Aurearena* (Aurearenophyceae, Kai et al. 2008) have the D-D and V-A cycles. The Phaeophyceae have only the V-A cycle, but the Xanthophyceae have the D-D cycle as well as the heteroxanthin-vaucheriaxanthin cycle (H-V cycle). The Phaeothamniophyceae have the D-D cycle as well as heteroxanthin (but not vaucheriaxanthin). Finally, the Raphidophyceae have an even more complex pattern, with the D-D and H-V cycles in the freshwater taxa but the V-A cycle in the marine taxa (Mostaert et al. 1998). Thus unlike the SII and SIII clades, it would appear that during the evolutionary history of the SI clade, new carotenoid cycles arose and dominated in different classes. Our new phylogeny might serve as a framework to revisit the morphological and biochemical characteristics that once formed the basis for heterokont systematics.

Methods

Cell cultures: Organisms were obtained from public culture collections (see Table 1) and generally grown according to the recommendations provided by the culture collections; strains not obtained from public culture collections are available upon request.

DNA extraction, amplification and sequencing: Genomic DNA was extracted from each culture strain using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), according to the manufacturers' instructions. PCR and sequencing were performed with published and newly designed specific primer sets for each gene (Supplementary Table S2); various combinations of newly designed primers were used for amplifying and sequencing for SSU and *rbc*L (Bailey et al. 1998; Daugbjerg and Andersen 1997), psaA130F-psaA1110R, psaA870F-psaA1600R and psaA971F-psaA1760R for *psa*A (Yang and Boo 2004; Yoon et al. 2002), psbAF-psbAR2 for *psb*A (Yoon et al. 2002), and psbC31F-1160R or psbC560F-psbC1160R for *psb*C. PCR amplification was performed on a total volume of 25 uL, containing 0.02 unit of Phusion^{TM} High-Fidelity DNA polymerase (Finnzymes OY, Espoo, Finland), 5 uL of the 5X Phusion^{TM} HF Buffer (contain 1.5 mM $MgCl_2$), 200 uM of each dNTPs, 10 uM of each primer and 1-20 ng of template DNA. PCR was carried out with an initial denaturation at 98 °C for 30 sec, followed by 30 main amplification cycles of denaturation at 94 °C for 10 sec, annealing at 50-55 °C for 30 sec and extension at 72 °C for 1 min, and a final extension at 72 °C for 7 min. Amplified DNA was purified with the QIAquick PCR Purification Kit (Qiagen) and send to a commercial sequencing company. Electropherogram outputs for each specimen were edited using the program Chromas v.1.45 (http://www.technelysium.com.au/chromas.html). Newly determined sequences were deposited in the GenBank databases (http://www.ncbi.nlm.nih.gov) under the accession numbers HQ710550 - HQ710794.

Phylogenetic analyses: Published sequences (e.g. Riisberg et al. 2009) were obtained from GenBank and EMBL-Align database (ftp://ftp.ebi.ac.uk/pub/databaseds/embl/align) and aligned using programs SeaView version 4.2.5 (Gouy et al. 2010) and Se-Al version 2.0a11 (http://tree.bio.ed.ac.uk/software/seal/). For the ten-gene dataset, newly determined nuclear SSU rRNA sequences were added to the Riisberg et al. (2009) alignments and realigned

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001

using ClustalW implemented in SeaView before manual editing; protein coding genes were manually refined based on the inferred amino acid sequences. In order to reduce tree construction artifacts, only the unambiguous regions of the nuclear LSU and SSU rRNA and the translated protein sequences were used. The nuclear LSU rRNA positions (reference sequence *Laminaria digitata* AF331153) that were used in the analyses are: 4-70, 145-178, 218-383, 540-620, 622-646, 671-691, 707-908, 915-1278, 1293-1484, 1512-1628, 1642-1740, 1752-1903, 2007-2399, 2409-2461, 2515-3047. The nuclear SSU rRNA positions (reference sequence *Sphacelaria* sp. UTEX LB 800 AY307401) that were used in the analyses are: 48-73, 88-116, 134-162, 175-531, 643-729, 734-961, 977-1255, 1286-1398, 1408-1595. Any ambiguous position (e.g., N) was treated as missing during subsequent analyses. All alignments are available on the TreeBase (http://purl.org/phylo/treebase/phylows/study/TB2:S11417).

In most cases for the five-gene dataset, the same strain was used when determining all gene sequences. However, we combined publicly available sequences from five species (Table 1). This data set was used to minimize the effect of missing data of the concatenated alignment on phylogeny. For the ten-gene dataset, sequences from different strains were concatenated at the species, genus, families, and order level based on Riisberg et al. (2009). For example, *Mallomonas caudata*, *M. papillosa*, *M. rasilis*, *M. splendens*, and *M. tonsurata* combined as *Mallomonas* in the Synurophyceae. In the Bacillariophyceae, Chrysophyceae and Phaeophyceae, however, sequences data of different genera were combined at the higher level. For example, *Ochromonas* spp. and *Spumella uniguttata* were combined as the Ochromonadaceae.

The evolutionary model for individual gene analysis was chosen by using the weighted Akaike information criterion (AICc) implemented in ModelGenerator version 0.85 (Keane et al. 2006). The selected best fitting models were the general time reversible (GTR) substitution with proportion of invariable site (I) and the gamma distributed rate heterogeneity (G) for the nuclear rRNA data; the LG substitution (Le and Gascuel 2008) with empirical amino acid frequencies (F) and G (LG + F + G model) for the *psb*A and *cox*1 amino acid sequences and without F (LG + G model) for *rbc*L, actin, beta-tubulin and *hsp*90 genes; the MtART (Abascal et al. 2007) + F + G model for the *psa*A and *psb*C genes. We used an independent model for each partition of the concatenated data.

Maximum likelihood (ML) analyses were performed using the RAxML version 7.2.8 (Stamatakis 2006). Tree likelihoods were estimated using 200 independent replications, each with a random starting point. The separate site-specific model was used for partitioned data and the automatically optimized SPR branch rearrangements were used during the rapid hill climbing tree search for each replication. Bootstrap analyses (MLB) were conducted using 1000 replications with same evolution model setting as the best topology search.

Bayesian analysis was conducted with MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003) using the best fitting model for each data set. The best fitting unfixed GTR + G model was used for the nuclear rRNA data partition as in ML search. However, the fixed CPREV substitution matrix (Adachi et al. 2000) was used as a fixed model with rate heterogeneity with F + G for the protein data because MrBayes does not support the LG substitution model. Three million Meteropolis-coupled Markov Chain Monte Carlo (MC^3) simulations were completed with the following parameters: two independent runs with different random start points, one cold chain and three heated chains for each run, and tree sampling at every 100th generation. The burn-in point of the simulation was identified by the average standard deviation of split frequencies (< 0.01) between the two independent runs. Sampled trees after the burn-in point were only used to calculate the Bayesian posterior probability (BPP) of monophyly. Some phylogenetic analyses were carried out on the Bioportal cluster on University of Oslo (http://www.bioportal.uio.no).

Alternative tree topology test: Alternative relationships at the class level were evaluated using the approximately unbiased (AU) test (Shimodaira 2002) implemented in CONSEL version 0.1k (Shimodaira and Hasegawa 2001). The five gene combined data was used for the paired-sites test with best phylogeny and possible alternative phylogenies, which included previously published data (Grant et al. 2009; Kai et al. 2008; Riisberg et al. 2009). The test was performed with 100,000 bootstrap replicates using the same evolutionary models and partitions as estimated in best ML tree search.

Acknowledgements

We thank Dr. Robert Bidigare, University of Hawaii, for unpublished *Chrysowaernella* carotenoid data and for assisting us in preparing the text concerning carotenoids. This research was supported by a grant from the National Science Foundation awarded to RAA (EF 04-3117), RAA and HSY (EF 08-27023), and HSY (DEB 09-37975), and grants of Korean Research Foundation (MEST; 2009-0077972) and MarineBio 21 program of Ministry of Land, Transport and Maritime Affairs, Korea to SMB, and a grant from the Next-Generation BioGreen 21 Program (SSAC; 2011-PJ008177), Rural Development Administration, Korea to HSY.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.protis.2011.08.001.

References

Abascal F, Posada D, Zardoya R (2007) MtArt: a new model of amino acid replacement for Arthropoda. Mol Biol Evol **24**:1–5

Adachi J, Waddell P, Martin W, Hasegawa M (2000) Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. J Mol Evol **50**:348–358

Andersen RA (1991) The cytoskeleton of chromophyte algae. Protoplasma **164**:143–159

Andersen RA, Saunders GW, Paskind MP, Sexton JP (1993) Ultrastructure and 18S rRNA gene sequence for *Pelagomonas calceolata* gen. et sp. nov. and the description of a new algal class, the Pelagophyceae classis nov. J Phycol **29**:701–715

Ariztia EV, Andersen RA, Sogin ML (1991) A new phylogeny for chromophyte algae using 16S-like rRNA sequences from *Mallomonas papillosa* (Synurophyceae) and *Tribonema aequale* (Xanthophyceae). J Phycol **27**:428–436

Bailey JC, Bidigare RR, Christensen SJ, Andersen RA (1998) Phaeothamniophyceae classis nova: a new lineage of chromophytes based upon photosynthetic pigments, *rbc*L sequence analysis and ultrastructure. Protist **149**:245–263

Ben Ali A, De Baere R, De Wachter R, Van de Peer Y (2002) Evolutionary relationships among heterokont algae (the autotrophic stramenopiles) based on combined analyses of small and large subunit ribosomal RNA. Protist **153**:123–132

Bjørnland T, Liaaen-Jensen S (1989) Distribution Patterns of Carotenoids in Relation to Chromophyte Phylogeny and Systematics. In Green JC, Leadbeater BSC, Diver WL (eds) The Chromophyte Algae, Problems and Perspectives. Systematics Association Special Volume No. 38, Clarendon Press, Oxford, pp 37–60

Boddi S, Bigazzi M, Sartoni G (1999) Ultrastructure of vegetative and motile cells, and zoosporogensis in *Chrysonephos lewisii* (Taylor) Taylor (Sarcinochrysidales, Pelagophyceae) in relation to taxonomy. Eur J Phycol **34**:297–306

Boo SM, Kim HS, Shin W, Boo GH, Cho SM, Jo BY, Kim J-H, Kim JH, Yang EC, Siver PA, Wolfe AP, Bhattacharya D, Andersen RA, Yoon HS (2010) Complex phylogeographic patterns in the freshwater alga *Synura* provide new insights into ubiquity vs. endemism in microbial eukaryotes. Mol Ecol **19**:4328–4338

Brown JW, Sorhannus U (2010) A Molecular genetic timescale for the diversification of autotrophic Stramenopiles (Ochrophyta): Substantive underestimation of putative fossil ages. PLoS ONE **5**:e12759

Cavalier-Smith T (2000) Flagellate Megaevolution: The Basis for Eukayote Diversification. In Green JC, Leadbeater BSC (eds) The Flagellates. Taylor and Francis, London, pp 361–390

Cavalier-Smith T, Chao EE (2006) Phylogeny and megasystematics of phagotrophic heterokonts (Kingdom Chromista). J Mol Evol **62**:388–420

Chung Y, Ané C (2011) Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. Syst Biol **60**:261–275

Cranston KA, Hurwitz B, Ware D, Stein L, Wing RA (2009) Species trees from highly incongruent gene trees in rice. Syst Biol **58**:489–500

Estabrook GF (1978) Some concepts for the estimation of evolutionary relationships in systematic botany. Syst Bot **3**:146–158

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

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogentic tree building. Mol Biol Evol **27**:221–224

Grant J, Tekle YI, Anderson OR, Patterson DJ, Katz LA (2009) Multigene evidence for the placement of a heterotrophic amoeboid lineage *Leukarachnion* sp. among photosynthetic stramenopiles. Protist **160**:376–385

Horn S, Ehlers K, Fritzsch G, Gil-Rhodríguez MC, Wilhelm C, Schnetter R (2007) *Synchroma grande* spec. nov. (Synchromophyceae class. nov., Heterokontophyta): An amoeboid marine alga with unique plastid complexes. Protist **158**:277–293

Janouškovec J, Horak A, Obornik M, Jukes J, Keeling PJ (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc Natl Acad Sci USA **107**:10949–10954

Kai A, Yoshii Y, Nakayama T, Inouye I (2008) Aurearenophyceae classis nova, a new class of Heterokontophyta based on a new marine unicellular alga *Aurearena cruciata* gen. et sp. nov. inhabiting sandy beaches. Protist **159**:435–457

Kawachi M, Inouye I, Honda D, O'Kelly CJ, Bailey JC, Bidigare RR, Andersen RA (2002) The Pinguiophyceae classis nova, a new class of chromophyte algae whose members produce large amounts of omega-3 fatty acids. Phycol Res **50**:31–47

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol **6**:29

Le Corguillé G, Pearson G, Valente M, Viegas C, Gschloessl B, Corre E, Bailly X, Peters A, Jubin C, Vacherie B, Cock JM, Leblanc C (2009) Plastid genomes of two brown algae, *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal derived plastids. BMC Evol Biol **9**:253

Leipe DD, Wainright PO, Gunderson JH, Porter D, Patterson DJ, Valois F, Himmerich S, Sogin ML (1994) The stramenopiles from a molecular perspective: 16S-like rRNA sequences from *Labyrinthuloides minuta* and *Cafeteria roenbergensis*. Phycologia **33**:369–377

Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Mol Biol Evol **25**:1307–1320

Moestrup Ø (2000) The Flagellate Cytoskeleton. Introduction of a General Terminology for Microtubular Flagellar Roots in Protists. In Leadbeater BSC, Green JC (eds) The Flagellates. Unity, Diversity and Evolution. Taylor and Francis, London, pp 69–94

Moriya M, Nakayama T, Inouye I (2002) A new class of the stramenopiles, Placididea classis nova: description of *Placidia cafeteriopsis* gen. et sp. nov. Protist **153**:143–156

Mostaert AS, Karsten U, Hiroki M, Hara Y, Watanabe MM (1998) Pigments and fatty acids of marine raphidophyte: a chemotaxonomic re-evaluation. Phycol Res **46**:205–220

O'Kelly CJ (1989) The Evolutionary Origin of the Brown Algae; Information from Studies of Motile Cell Structure. In Green JC, Leadbeater BSC, Diver WL (eds) The Chromophyte Algae: Problems and Persepectives. Clarendon Press, Oxford, UK, pp 255–278

Ong HG, Wilhelm SW, Gobler CJ, Bullerjahn G, Jacobs MA, McKay J, Sims EH, Gillett WG, Zhou Y, Haugen E, Rocap G, Cattolico RA (2010) Analyses of the complete chloroplast genome sequences of two members of the Pelagophyceae: *Aureococcus anophagefferens* CCMP 1984 and *Aureoumbra lagunensis* CCMP 1507. J Phycol **46**:602–615

Page RDM, Charleston MA (1997) Reconciled Trees and Incongruent Gene and Species Trees. in: Mirkin B, McMorris FR, Roberts FS, Rzhetsky A. (eds), DIMACS. Series in

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001

Discrete Mathematics and Theoretical Computer Science. 37 Am Math Soc, Rutgers, New Brunswick, NJ, pp 57–70

Potter D, Saunders GW, Andersen RA (1997) Phylogenetic relationships of the Raphidophyceae and Xanthophyceae as inferred from nucleotide sequences of the 18S ribosomal RNA gene. Am J Bot **84**:962–972

Riisberg I, Orr RJS, Kluge R, Shalchian-Tabrizi K, Bowers HA, Patil V, Edvardsen B, Jakobsen KS (2009) Seven gene phylogeny of heterokonts. Protist **160**:191–204

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572–1574

Sang T, Zhong Y (2000) Testing hybridization hypotheses based on incongruent gene trees. Syst Biol **49**:422–434

Saunders GW, Potter D, Andersen RA (1997) Phylogenetic affinities of the Sarcinochrysidales and Chrysomeridales (Heterokonta) based on analyses of molecular and combined data. J Phycol **33**:310–318

Saunders GW, Potter D, Paskind MP, Andersen RA (1995) Cladistic analyses of combined traditional and molecular data sets reveal an algal lineage. Proc Natl Acad Sci USA **92**:244–248

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Syst Biol **51**:492–508

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylgoentic tree selection. Bioinformatics **17**:1246–1247

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688–2690

Wilkinson M, McInerney JO, Hirt RP, Foster PG, Embley TM (2007) Of clades and clans: terms for phylogenetic relationships in unrooted trees. Trends Ecol Evol **22**:114–115

Yang EC, Boo SM (2004) Evidence for two independent lineages of *Griffithsia* (Ceramiaceae, Rhodophyta) based on plastid protein-coding *psa*A, *psb*A, and *rbc*L gene sequences. Mol Phylogenet Evol **31**:680–688

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**:11724–11729

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

Please cite this article in press as: Yang EC, et al. Supermatrix Data Highlight the Phylogenetic Relationships of Photosynthetic Stramenopiles. Protist (2011), doi:10.1016/j.protis.2011.08.001