Protist, Vol. 160, 191—204, May 2009
http://www.elsevier.de/protis
Published online date 9 February 2009

Protist

ORIGINAL PAPER

# Seven Gene Phylogeny of Heterokonts

**Ingvild Riisberg[a,d,1], Russell J.S. Orr[b,d,1], Ragnhild Kluge[b,c,2], Kamran Shalchian-Tabrizi[d], Holly A. Bowers[e], Vishwanath Patil[b,c], Bente Edvardsen[a,d], and Kjetill S. Jakobsen[b,d,3]**

[a]Marine Biology, Department of Biology, University of Oslo, P.O. Box 1066, Blindern, NO-0316 Oslo, Norway
[b]Centre for Ecological and Evolutionary Synthesis (CEES),Department of Biology, University of Oslo, P.O. Box 1066, Blindern, NO-0316 Oslo, Norway
[c]Department of Plant and Environmental Sciences, P.O. Box 5003, The Norwegian University of Life Sciences, N-1432, Ås, Norway
[d]Microbial Evolution Research Group (MERG), Department of Biology, University of Oslo, P.O. Box 1066, Blindern, NO-0316, Oslo, Norway
[e]Center of Marine Biotechnology, 701 East Pratt Street, Baltimore, MD 21202, USA

Submitted May 23, 2008; Accepted November 15, 2008
Monitoring Editor: Mitchell L. Sogin

**Nucleotide ssu and lsu rDNA sequences of all major lineages of autotrophic (Ochrophyta) and heterotrophic (Bigyra and Pseudofungi) heterokonts were combined with amino acid sequences from four protein-coding genes (actin, β-tubulin, *cox1* and *hsp90*) in a multigene approach for resolving the relationship between heterokont lineages. Applying these multigene data in Bayesian and maximum likelihood analyses improved the heterokont tree compared to previous rDNA analyses by placing all plastid-lacking heterotrophic heterokonts sister to Ochrophyta with robust support, and divided the heterotrophic heterokonts into the previously recognized phyla, Bigyra and Pseudofungi. Our trees identified the heterotrophic heterokonts Bicosoecida, *Blastocystis* and Labyrinthulida (Bigyra) as the earliest diverging lineages. A separate analysis of the phototrophic lineages, by adding the *rbcL* gene, further resolved the Ochrophyta lineages by increased support for several important nodes. Except for the positioning of Chrysophyceae, Eustigmatophyceae, Raphidophyceae and Pinguiophyceae, all main branches of Ochrophyta were resolved. Our results support the transfer of classes Dictyochophyceae and Pelagophyceae from subphylum Phaeista to Khakista. Based on all our trees, in combination with current knowledge about ultrastructure of heterokonts we suggest that a more advanced flagellar apparatus originated at one occasion in the ancestor of Phaeista whereas, Khakista independently reduced their flagellar apparatus and gained chlorophyll $c_3$.**


**Key words:** heterokont phylogeny; covarion; Bayesian; maximum likelihood; multigene; stramenopiles.

## Introduction

Heterokonta was established as a phylum by Cavalier-Smith (1986), comprising all eukaryotic motile biflagellate cells having typically a forward directed flagellum (cilium) with tripartite rigid tubular flagellar hairs (mastigonemes) and a trailing hairless (smooth) flagellum, plus all their descendants having secondarily lost one or both

e-mail k.s.jakobsen@bio.uio.no (K.S. Jakobsen).
[1]These authors contributed equally to this work.
[2]Present address: Norwegian Pollution Control Authority, Box 8100 Dep, 0032 Oslo, Norway.
[3]Corresponding author; fax +47 22854001.


doi:10.1016/j.protis.2008.11.004

flagella. The diversity among heterokonts is striking; they range from large multicellular seaweeds to tiny unicellular species, they are present in freshwater, marine and terrestrial habitats and embrace many ecologically important algal (e.g. diatoms, brown algae, chrysophytes), and heterotrophic (e.g. oomycetes) groups. Due to the diversity in Heterokonta it was later raised to infrakingdom (Cavalier-Smith 1997) with two main groups; Ochrophyta (Cavalier-Smith 1986) consisting mainly of autotrophic heterokonts, and a purely heterotrophic group, which was again further subdivided in two phyla: Bigyra and Pseudofungi (Cavalier-Smith and Chao 2006).

Due to their diversity and ecological significance, a number of molecular phylogenetic studies have been performed on separate heterokonts groups; predominantly single gene inferences (Andersen et al. 1999; Guillou et al. 1999; Moriya et al. 2000; Negrisolo et al. 2004; Potter et al. 1997). Investigations on the global phylogeny of heterokonts are limited to relatively few studies of nuclear-encoded ribosomal RNAs (rDNA) or chloroplast-encoded *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase) (Ben Ali et al. 2001; Daugbjerg and Guillou 2001; Edvardsen et al. 2007). Recently, the most species-rich phylogeny of all three heterokont phyla (Ochrophyta, Bigyra, Pseudofungi) employing a comprehensive ssu rDNA dataset was performed (Cavalier-Smith and Chao 2006). The ssu rDNA study clarified some sister taxa relationships, and displayed the heterotrophic groups as deep branching heterokonts (Cavalier-Smith and Chao 2006). The phylogenetic inferences of heterokonts have been congruent in placing the plastid-free heterotrophic species basal to the plastid-containing forms (Cavalier-Smith and Chao 2006), but the statistical support for this phylogeny has been only moderate. According to the chromalveolate hypothesis heterokonts, haptophytes, cryptophytes and alveolates acquired their red-algal derived plastids in a single secondary endosymbiotic event (Cavalier-Smith 1999). However, recurrent plastid loss and gains have been suggested for several chromalveolate groups (Bachvaroff et al. 2006; Bodyl 2005; Sanchez-Puerta and Delwiche 2008; Shalchian-Tabrizi et al. 2006b). A robust phylogeny for Ochrophyta and the entirely heterotrophic heterokont groups may improve our understanding of the type of processes forming the current plastid distribution.

Multigene approaches for resolving phylogenetic relationships using a moderate number of protein encoding genes have during the last few years been successfully applied to several protist groups (Fast et al. 2002; Kim et al. 2006; Nosenko and Bhattacharya 2007; Shalchian-Tabrizi et al. 2006a; Simpson et al. 2006). For heterokonts, two-gene analyses (lsu+ssu rDNA) have been performed — but, with the main emphasis on Ochrophyta (Ben Ali et al. 2002; Edvardsen et al. 2007). In the present study, we have performed a two-gene lsu+ssu rDNA analysis of all three heterokont phyla (Ochrophyta, Pseudofungi and Bigyra). In order to further enhance the phylogenetic resolution we have combined the rDNA data with amino acids from five protein encoding genes (actin, beta-tubulin [β-tubulin], cytochrome oxidase subunit I (*cox1*), heat-shock protein 90 (*hsp90*) and *rbcL*). Our largest concatenated alignment consisted of 5643 characters (3886 nucleotides and 1757 amino acids) thereby representing the most gene- and character-rich heterokont alignment published to date.

## Results

Thirty-six new heterokont nucleotide sequences including lsu (rRNA), ssu (rRNA), actin, β-tubulin, *cox1* and *hsp90* were PCR-amplified for the purpose of this study and have been made publicly available with accession numbers from FJ030880 to FJ030915. In addition, we have extracted all available accessions from GenBank for the various heterokont groups including the above-mentioned genes plus *rbcL* (Table 1).

### Global Phylogeny of Ochrophyta and Heterotrophic Heterokonts

Our dataset included 35 taxa representing ten heterokont classes (Bacillariophyceae, Chrysophyceae, Dictyochophyceae, Eustigmatophyceae, Pelagophyceae, Phaeophyceae, Phaeothamniophyceae, Pinguiophyceae, Raphidophyceae, Xanthophyceae) as well as the parasitic taxa *Blastocystis hominis* and *Developayella elegans.* Thus, all major lineages of Ochrophyta, Bigyra and Pseudofungi (heterotrophic heterokonts) except Bolidiophyceae, Chrysomerophyceae, Placididea and Opalinidea (excluded due to lack of available sequence data or lack of available cultures or DNA) were represented.

For all our analyses the Bayesian covarion sequence evolution model (Huelsenbeck 2002) fitted the data better than competing models with the highest marginal likelihoods. Our phylogenetic analysis based on lsu+ssu rDNA sequence data

**Table 1.** Taxa, genes and accession numbers of sequences used in the phylogenetic inferences (Figs 1—3)

| Taxon | lsu | ssu | actin | β-tubulin | *cox1* | *hsp90* | *rbcL* | Missing % |
|---|---|---|---|---|---|---|---|---|
| Aconoidasida | AF013419[1*] | AF013418[1] | ABO69629[2] | XP_001347369[2] | AAC63390[2] | BM162404[3] CX020168[3] | Heterotroph | 0.4 |
| *Aureococcus anophagereffens* | AF289042 | U40257 | EH117431 | AAP42295 | — | EH117437 | AAD39107 | 14.1 |
| Bacillariophycidae | EF553458[4] CT899870[4] | EF140622[4] | AAV32831[5] | AAW58083[4] | BAA86610[6] | AAX10945[4] | YP_874418[4] | 0.2 |
| *Blastocystis hominis* | CT899870 EC649592 EC651237 EC649837 EC649960 EC645290 | AB023499 | EC649064 | EC650343 EC650685 | — | — | Heterotroph | 23.2 |
| *Caecitellus parvulus* | FJ030883 | AY642126 | ABC85753 | — | FJ030895 | — | Heterotroph | 17.4 |
| *Chattonella* sp. | FJ030891[7] | AB217626[7] | — | — | AAC94976[7] | — | AAC02920[8] | 16.9 |
| Chlorarachniophyceae | AF289036[9] | AF054832[9] | AAK92363[9] | AAC68506[9] | BAA96355[10] | DR041547[9] DR040208[9] DR038899[9] | Heterotroph | 0.4 |
| *Cryptomonas* sp. | AY752990[11] | AF508271[12] | AAG31474[13] | ABR22551[11] | BAA96358[13] | — | CAJ19660[13] | 11.2 |
| *Developayella elegans* | FJ030882 | U37107 | — | — | — | — | Heterotroph | 25.3 |
| *Dictyocha speculum* | AF289046 | U14385 | — | — | AM850252 | — | AAK85735 | 19.5 |
| Ectocarpales | AF331155[14] | AY307398[15] | — | P30156[16] | AAC94980[17] | — | AAD22595[18] | 11.2 |
| Eimeriorina | AF044252[19] | AF009244[19] | AAC13766[20] | XP_627803[21] | ABL96290[22] | XP_626924[21] | Heterotroph | 4.8 |
| Fucales | AF331151[23] | AB011423[24] | Q39758[23] | DV670063[25] | YP_448623[23] | ACF06184[26] | CAC67518[27] | 9.7 |
| *Glossomastix chrysoplasta* | AF409128 | AF438324 | FJ030905 | — | FJ030899 | FJ030915 | AAM45877 | 9.8 |
| *Guillardia theta* | AF289037 | X57162 | AAG31473 | AAD02570 | — | AAX10949 | NP_050705 | 6.3 |
| Gymnodiniaceae | AY831412[28] | AY831412[29] | ABI14387[29] | ABV22200[29] | ABK57965[30] | CAJ75751[29] | Excluded | 0.2 |
| *Heterosigma akashiwo* | AF409124 | DQ191681 | AAW56948 | AAW58080 | — | AAX10940 | BAD151144 | 6.3 |
| *Hyphochytrium catenoides* | X80345 | AF163294 | — | — | — | — | Heterotroph | 25.1 |
| *Isochrysis galbana* | DQ202389 EC143993 EC146828 EC144527 EC142332 EC137274 | AJ246266 | AAW56949 | AAW58081 | BAA24905 | AAX10942 | BAB20783 | 6.5 |
| *Japonochytrium* sp. | FJ030887[31] | AB022104[31] | ABC85743[32] | ABC61493[32] | — | — | Heterotroph | 33.6 |
| Laminariaceae | AF331153[33] | AB022818[34] | ABB02445[35] | CN468112[33] | NP_659274[33] | CN468204[33] | AAR24450[33] | 10.0 |
| *Mallomonas* sp. | FJ030885[36] EF165147[37] | FJ030893[37] | AAW56950[36] | — | — | FJ030913[36] | ABN46956[38] | 15.8 |
| Mischococcales | FJ030881[39] | FJ030892[39] | — | — | BAA24967[40] | FJ030907[39] | CAE45700[39] | 11.8 |
| *Nannochloropsis gaditana* | FJ030880 | AF067957 | FJ030901 | — | — | FJ030909 | BAC10559 | 13.7 |
| *Nannochloropsis salina* | Y07975 | AF045048 | — | — | — | — | BAC10567 | 23.1 |
| *Nerada mexicana* | FJ030884 | AY520453 | FJ030902 | — | — | FJ030912 | Heterotroph | 15.7 |
| Ochromonadaceae | Y07977[41] | EF165136[42] | AAR27549[43] | AAW58087[43] | NP_066450[41] | AAR27540[43] | ABN46934[44] | 2.3 |
| *Pavlova* sp. | DQ202388[45] EC174948[45] EC174484[45] EC174288[45] EC177179[45] | AF102373[46] EC177131[47] | EC176662[47] | AAW58082[47] | AAD12068[47] | AAX10944[47] | BAA08281[48] | 12.9 |
| Pedinellales | AF289045[49] | U14387[50] | — | — | FJ030896[50] | FJ030910[50] | AAC02817[50] | 13.6 |
| *Pelagococcus subviridis* | FJ030889 | U14386 | — | — | — | — | AAC02919 | 23.4 |
| *Pelagomonas calceolata* | AF289047 | U14389 | ABX25970 | ABX25960 | — | — | AAC02816 | 14.3 |
| Peridiniales | AY112746[51] | AF077055[51] | AAO49340[52] | AAO49343[52] | AAM97694[51] | AAX10941[52] | Excluded | 6.6 |
| Phaeothamniales | FJ030890[53] | AF044846[54] | — | — | — | — | AAC62462[54] | 34.4 |
| *Phytophtora* sp. | X75631[55] | AY742761[56] | P22131[56] | ABV00081[57] | AAA32023[55] | AAX10946[58] | Heterotroph | 0.4 |
| *Pinguiococcus pyrenoidosa* | AF409130 | AF438325 | FJ030906 | — | FJ030900 | FJ030908 | AAM45878 | 9.9 |
| Prymnesiaceae | AF289038[59] | AB183612[60] | AAW56954[61] | ABJ80997[62] | — | AAX10951[61] | BAB20788[61] | 6.8 |
| *Pseudochattonella farcimen* | FJ030886 | AM075624 | FJ030903 | FJ030894 | FJ030897 | — | AM850235 | 27.9 |
| *Pythium* sp. | AY598647[63] EL777230[63] EL777164[63] EL774542 | EF418925[64] | AAW56955[65] | AAY40457[66] | AAM98738[67] | AAX10948[65] | Heterotroph | 0.3 |

Table 1. (*continued*)

| Taxon | lsu | ssu | actin | β-tubulin | *cox1* | *hsp90* | *rbcL* | Missing % |
|---|---|---|---|---|---|---|---|---|
| *Rhizosolenia setigera* | AF289048 | M87329 | — | — | BAA86611 | — | AAC02907 | 16.9 |
| Saprolegniaceae | AF235943[68] AF320587[68] | M32705[69] | P26182[69] | P20802[70] | YP_052915[71] | AAM90674[72] | Heterotroph | 20.2 |
| *Synura* sp. | DQ980475[73] | U73221[73] | ABJ80925[73] | ABJ81004[73] | — | FJ030911[74] | ABN46952[75] | 9.5 |
| Thalassiosirales | Y11512[76] FC497182[79] FC536095[79] | AJ632209[76] | ABC54738[77] | AAF81906[78] | YP_316586[79] | FC503950[79] | YP_874498[79] | 0.4 |
| *Thaumatomonas* sp. | DQ980477[80] | AF411259[81] | ABJ80927[80] | ABJ81007[80] | — | AAP72162[80] | Heterotroph | 7.5 |
| *Thraustochytrium* sp. | FJ030888[82] | AB022110[82] | FJ030904[82] | ABC61492[83] | FJ030898[82] | FJ030914[82] | Heterotroph | 18.8 |
| *Tribonema* sp. | Y07979[84] | AF083397[85] | — | — | BAA24975[84] | — | AAD02843[84] | 16.9 |
| *Vacuolaria virescens* | AF409125 | U41651 | — | — | — | — | AAC02921 | 24.9 |
| Character end position | 2473 | 3886 | 4134 | 4521 | 4873 | 5185 | 5643 | |
| Alignment accession | ALIGN_001291 | ALIGN_001292 | ALIGN_001293 | ALIGN_001294 | ALIGN_001295 | ALIGN_001296 | ALIGN_001297 | |

Calculated percentage of total missing characters for each taxon in the multigene alignment is presented. Note, percentages of missing characters for species not included in the phylogenetic inference represented in Fig. 3 are calculated without *rbcl* (labelled; Heterotroph). The dinoflagellate (Dinophyta) taxa, Gymnodiniaceae and Peridiniales were "Excluded" from the *rbcL* inference, as this gene is highly derived. All seven single-gene alignments have been made available under the provided accession numbers from the EMBL-Align database: ftp://ftp.ebi.ac.uk/pub/databases/embl/align/

*Composite (*in silico* chimeric) sequences constructed from closely related species indicated as follows: [1]*Theileria parva,* [2]*Plasmodium falciparum,* [3]*Plasmodium* sp., [4]*Phaeodactylum tricornutum,* [5]*Nitzschia thermalis,* [6]*Nitzschia frustulum,* [7]*Chattonella antiqua,* [8]*Chattonella subsalsa,* [9]*Bigelowiella natans,* [10]*Chlorarachnion reptans,* [11]*Cryptomonas paramecium,* [12]*Cryptomonas platyuris,* [13]*Cryptomonas ovata,* [14]*Streblonema maculans,* [15]*Ectocarpus siliculosus,* [16]*Ectocarpus variabilis,* [17]*Ectocarpus* sp., [18]*Myriotrichia clavaeformis,* [19]*Frenkelia microti,* [20]*Toxoplasma gondii,* [21]*Cryptosporidium parvum,* [22]*Eimeria acervulina,* [23]*Fucus vesiculosus,* [24]*Fucus distichus,* [25]*Sargassum binderi,* [26]*Fucus serratus,* [27]*Sargassum muticum,* [28]*Akashiwo sanguinea,* [29]*Karlodinium micrum,* [30]*Akashiwo* sp., [31]*Japonochytrium* sp., [32]*Japonochytrium marinum,* [33]*Laminaria digitata,* [34]*Laminaria angustata,* [35]*Saccharina japonica,* [36]*Mallomonas tonsurata,* [37]*Mallomonas* sp., [38]*Mallomonas rasilis,* [39]*Chlorellidium tetrabotrys,* [40]*Botrydiopsis alpina,* [41]*Ochromonas danica,* [42]*Ochromonas distigma,* [43]*Spumella uniguttata,* [44]*Ochromonas* sp., [45]*Pavlova* sp., [46]*Pavlova pinguis,* [47]*Pavlova lutheri,* [48]*Pavlova salina,* [49]*Apedinella radians,* [50]*Pseudopedinella elastica,* [51]*Pfiesteria piscicida,* [52]*Heterocapsa triquetra,* [53]*Phaeobotrys solitaria,* [54]*Phaeothamnion confervicola,* [55]*Phytophthora megasperma,* [56]*Phytophthora infestans,* [57]*Phytophthora glovera,* [58]*Phytophthora palmivora,* [59]*Prymnesium patelliferum,* [60]*Prymnesium* sp., [61]*Prymnesium parvum,* [62]*Chrysochromulina* sp., [63]*Pythium* sp., [64]*Pythium helicoids,* [65]*Pythium graminicola,* [66]*Pythium macrosporum,* [67]*Pythium aphanidermatum,* [68]*Achlya* sp., [69]*Achlya bisexualis,* [70]*Achlya klebsiana,* [71]*Saprolegnia ferax,* [72]*Achlya ambisexualis,* [73]*Synura sphagnicola,* [74]*Synura curtispina,* [75]*Synura petersenii,* [76]*Skeletonema pseudocostatum,* [77]*Skeletonema costatum,* [78]*Thalassiosira weissflogii,* [79]*Thalassiosira pseudonana,* [80]*Thaumatomonas* sp., [81]*Thaumatomonas seravini,* [82]*Thraustochytrium aureum,* [83]*Thraustochytrium kinnei,* [84]*Tribonema aequale,* [85]*Tribonema intermixtum.*

(3886 nucleotides, Fig. 1) supported Ochrophyta as a monophyletic clade with high support (Bayesian posterior probability of 1.00, maximum likelihood bootstrap value of 99 (Treefinder) and 100 (RAxML), hereafter written 1.00/99/100). The lsu+ssu rDNA tree (Fig. 1) included two weakly supported heterotrophic branches, Bigyra (0.64/—/—) and Pseudofungi (0.82/—/—). Within Ochrophyta, all classes were highly supported (1.00/>90/>90). In addition, the phylogenetic relationship of the classes Pelagophyceae and Dictyochophyceae as sister taxa was well resolved (1.00/93/90) — likewise that of Phaeophyceae as sister taxa to Xanthophyceae (1.00/100/99).

By adding protein sequences from four genes (actin, β-tubulin, *cox1* and *hsp90*; in total 1299 amino acid characters) to the lsu+ssu rDNA dataset, we generated the most gene-rich heterokont alignment to date. No taxa had more than 35% missing character data and the supermatrix proportion of missing data was <15% (Table 1).

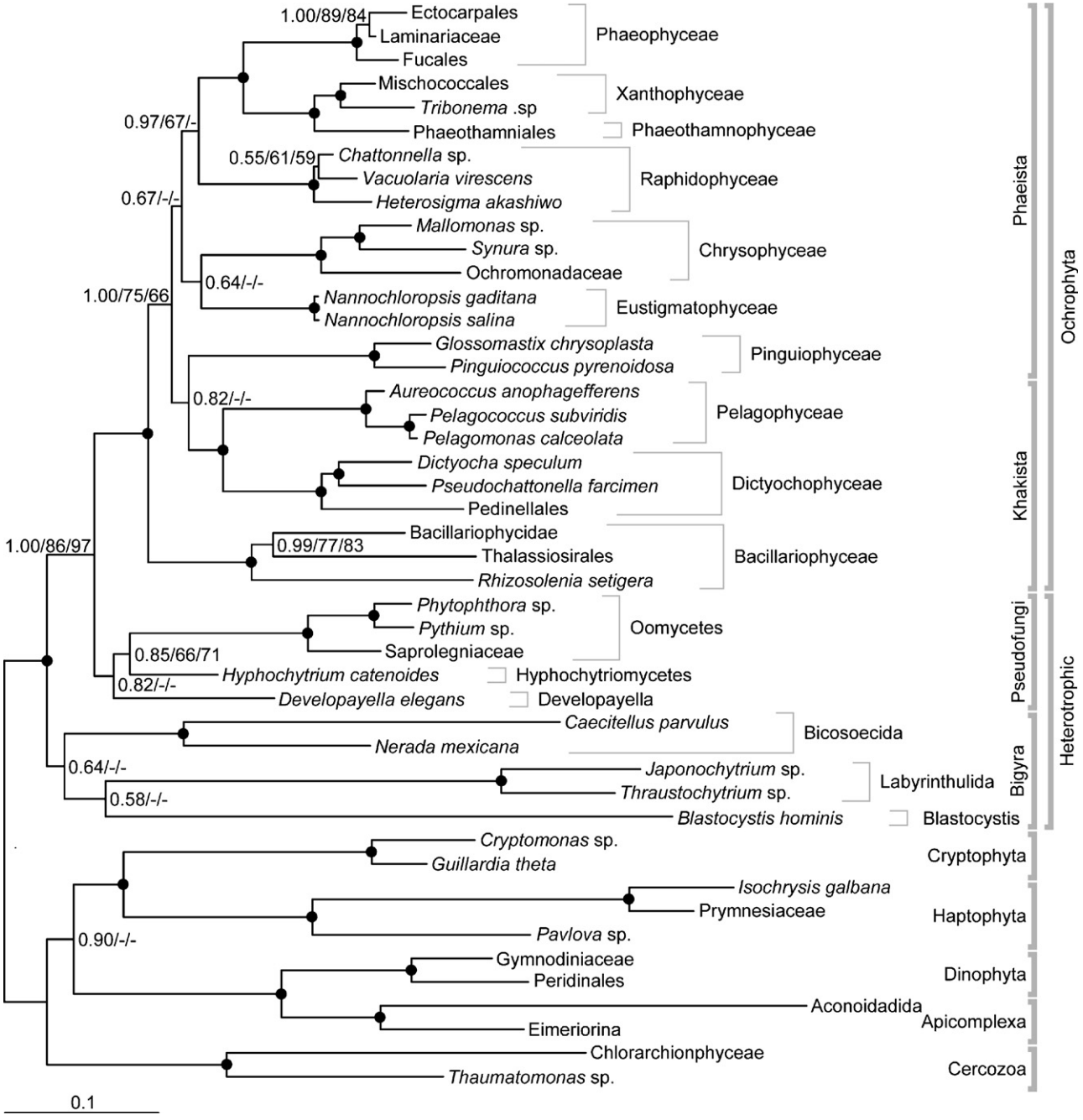


**Figure 1**. rDNA phylogeny of heterokonts. Combined lsu and ssu phylogeny of Ochrophyta, Bigyra and Pseudofungi (3886 characters). The tree is reconstructed with Bayesian inference (MrBayes). Numbers on the internal nodes represent posterior probability and bootstrap values (>50%) for Treefinder and RAxML (ordered; MrBayes/Treefinder/RAxML). Black circles indicate support values of 1.00 posterior probability and bootstrap over 90%. See Table 1 for inferred composite (in silico chimeric) sequences.

The combined rDNA and protein tree (Supplementary Fig. S1) showed a similar topology to the lsu+ssu rDNA tree (Fig. 1), suggesting that the protein and rDNA data contain congruent phylogenetic signals. However, some branching patterns remained elusive, and the heterotrophic taxa *Developayella elegans* moved from a basal position in the Pseudofungi to a position basal to Ochrophyta (0.98/63/68). Additionally, Bigyra split to form two clades consisting of Labyrinthulida

and Blastocystis/Bicosoecida (0.96/—/51). To investigate the effect of the fastest evolving heterotrophic taxa on the topology, PAML was used to remove fast-evolving classes of sites from the alignment (site class 8) (Yang 1997, 2007; Kumar et al. unpublished). The resulting tree (Fig. 2) again supports Ochrophyta as a monophyletic group, congruent with Figures 1 and S1 (support values 1.00/100/99). The groupings of Pelagophyceae with Dictyochophyceae (1.00/91/87) and

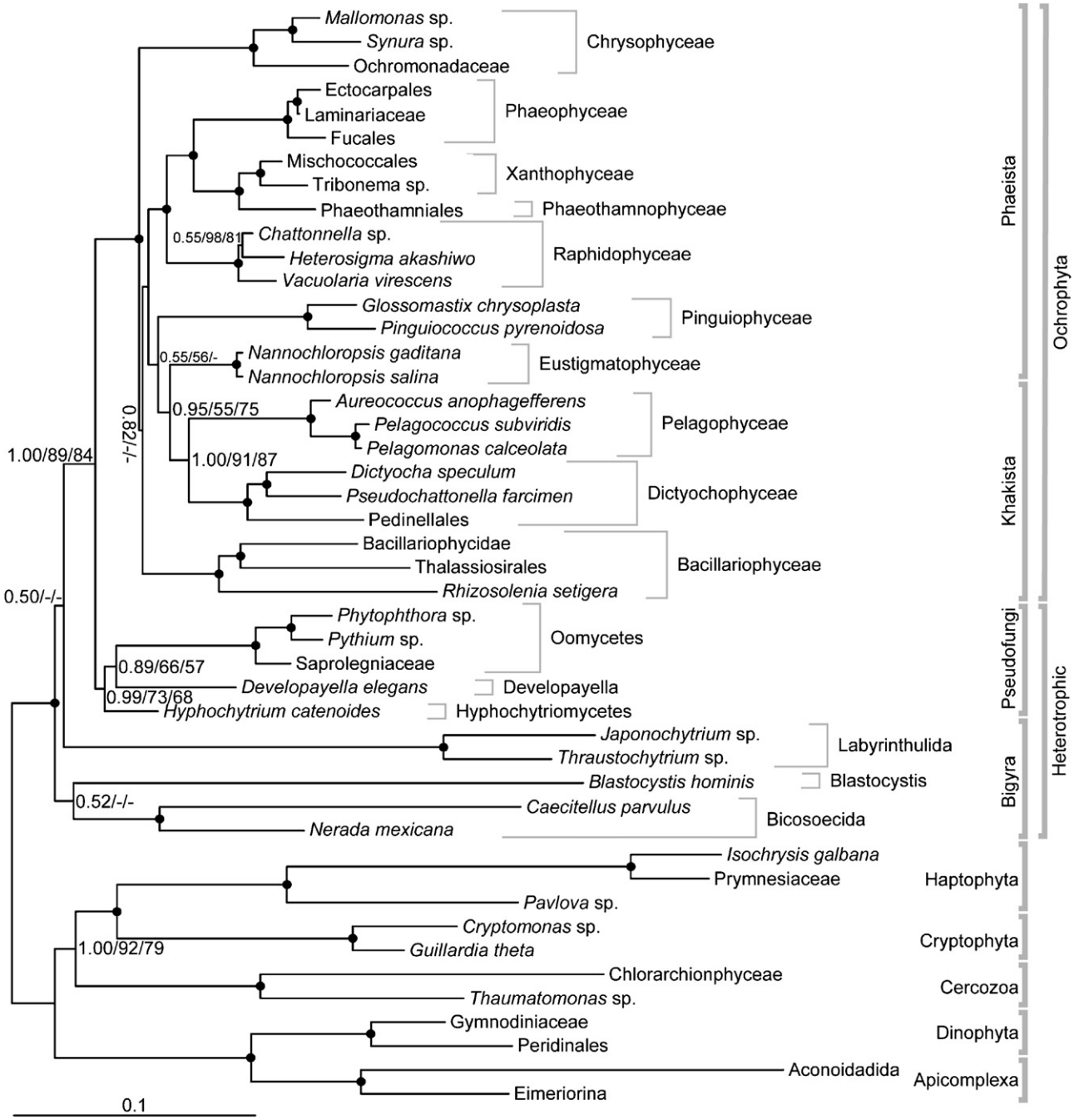


**Figure 2**. Protein and rDNA phylogeny of heterokonts. Combined lsu, ssu, actin, β-tubulin, *cox1* and *hsp90* phylogeny of Ochrophyta, Bigyra and Pseudofungi. PAML fast-evolving site category 8 for both rDNA and proteins has been removed (Yang 1997, 2007; Kumar et al. unpublished) (4632 characters). For further information, see legend for Figure 1.

Phaeophyceae as sister taxa to Xanthophyceae were still highly supported (1.00/100/99). Further, Bacillariophyceae clustered at a deep Ochrophyta node, and the placement of Raphidophyceae was consistent with Figure 1. The combined rDNA and protein tree did not resolve the position of Pinguiophyceae, Chrysophyceae and Bacillariophyceae; all positions received weak support. In contrast to the tree including fast evolving sites (Fig. S1) and in congruence with Figure 1, *Developayella elegans* clustered monophyletically with the Pseudofungi to form a moderately supported clade (0.99/73/68). The branching pattern for Bigyra remained unresolved. As branching pattern for the heterotrophic taxa remained an issue, the approximately unbiased (AU) test (Shimodaira 2002) was applied to compare several user-defined trees as well as the optimal topologies obtained from the Bayesian analyses. The AU test rejected a single plastid loss (heterotrophic monophyly) hypothesis ($p$-value $= 0.00$). However, it could not reject a heterotrophic heterokont evolution resulting in 2, 3 or 4 separate clades. See Supplementary data; AU tests.

### Protein Tree is Congruent with the rDNA and Combined rDNA and Protein Trees

A separate four-gene protein analysis (actin, β-tubulin, *cox1*, *hsp90*) of taxa having less than 50% missing character data was performed (25 taxa, 1299 characters, Fig. S2). As this dataset contained fewer taxa than Figures 1 and 2, the tree does not reflect the relationship between all classes, but shows an overall similarity by dividing the Ochrophyta and the heterotrophic groups (1.00/99/92). The deviations from the other topologies in Figures 1 and 2 were not significantly supported. Thus, it seems that there are no significant conflicts in the phylogenetic signal in the rDNA and protein datasets. Additionally, the single protein gene trees for actin, β-tubulin, *cox1* and *hsp90* did not reveal any supported topological conflicts with the protein concatenated trees (see Methods and Figs S4—S9).

### Addition of *rbcL* Sequence Data for Inference of Ochrophyta Phylogeny

To further improve the resolution of Ochrophyta, we added the *rbcL* gene to the protein and rDNA sequences (458 amino acids, in total the alignment constituted 1757 amino acids) and performed a separate analysis of these lineages only (in total 30 taxa, 5643 characters). The heterotrophic lineages (Bigyra and Pseudofungi) not possessing the autotrophic *rbcL* gene were excluded and the phylogeny was inferred both with and without an outgroup (Figs 3 and S3). The resulting tree with outgroup (Fig. 3) showed many similarities with Figures 1 and 2, but a notable change was the placement of Pelagophyceae/Dictyochophyceae together with Bacillariophyceae (Khakista excluded from Phaeista with 1.00/60/—). The exclusion of Khakista from Phaeista was further highlighted when the relationship was inferred without an outgroup (1.00/80/58) (Fig. S3). The branching pattern within Phaestia was unresolved with Eustigmatophyceae, Raphidophyceae, Chrysophyceae and Pinguiophyceae receiving no topological support. The single gene tree for *rbcL* did not reveal any supported topological conflicts with the concatenated trees (see Methods and Fig. S10).

## Discussion

### Multigene Phylogeny and Covarion Analyses Strongly Divide Ochrophyta and Heterotrophs

In previous studies, usually only a few of the affiliated classes have clustered together with statistical support, and many of the deeply diverging nodes have remained unresolved (Ben Ali et al. 2001, 2002; Daugbjerg and Andersen 1997; Edvardsen et al. 2007). Here we present a considerable number of sequences from all the major classes of the heterokonts, and increase the support for class-level phylogenies. Compared to previous combined lsu+ssu rDNA analyses (Ben Ali et al. 2002; Edvardsen et al. 2007) we increased the number of taxa, and instead of using only partial sequences including the D1 and D2 domains (Edvardsen et al. 2007) we sequenced and compared the full length of the lsu sequences (Fig. 1).

In all our analyses no taxa had more than 35% missing character data in the inferred alignments (for further details, see Table 1). According to Wiens, taxa that are missing up to 50% character data can subdivide long branches and improve accuracy as well as those with no missing character data (Wiens 2005, 2006). In addition, the resulting supermatrices proportion of missing data never exceeded 15% (14% missing data in Fig. 3), a figure significantly lower than that

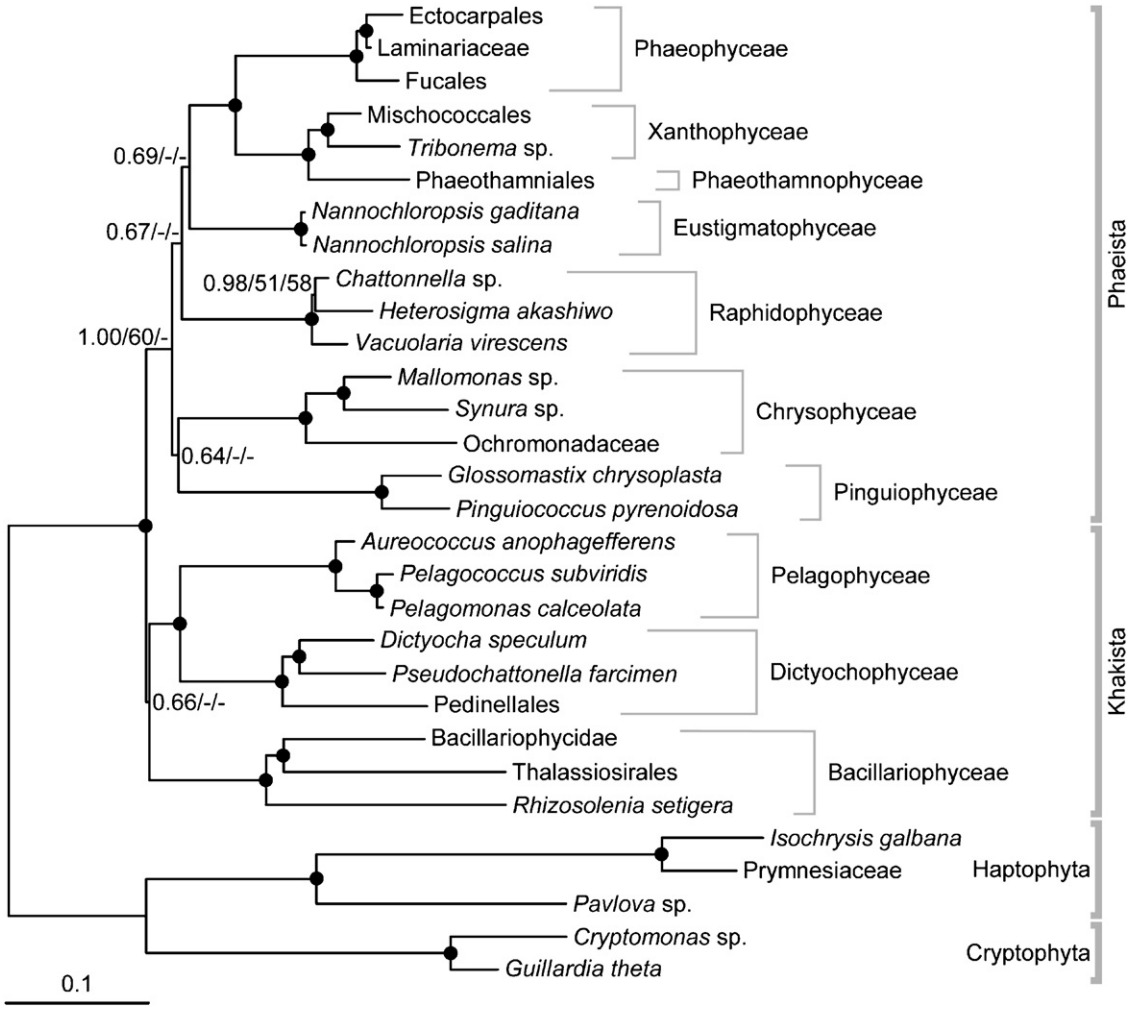


**Figure 3**. Protein and rDNA phylogeny of photosynthetic heterokonts. Combined lsu, ssu, actin, β-tubulin, *cox1*, *hsp90* and *rbcL* phylogeny of Ochrophyta (5643 characters). For further information, see legend for Figure 1.

scrutinized by others (e.g., 25%: Philippe et al. 2004; 75%: Wiens and Reeder 1995; 95%: McMahon and Sanderson 2006). For this reason, the Bayesian topology was preferred since this has been documented to be less prone to missing characters than both likelihood and parsimony methods (Wiens 2005, 2006; Wiens and Moen 2008). Additionally, the covarion model received significantly higher marginal likelihood values than competing models.

When comparing the amounts of rDNA data versus protein data in our alignments, rDNA constituted the majority (Table 1). Still, the separate protein tree (with taxa having less than 50% missing character data) was congruent with the rDNA tree (Figs 1 and S2). In our global heterokont trees (Figs 1 and 2) the split between heterotrophic heterokonts (Bigyra and Pseudofungi) and Ochrophyta was strongly supported which shows that the resolution and support for our preferred topology is greater than in previous published analyses (Ben Ali et al. 2002; Cavalier-Smith and Chao 2006; Edvardsen et al. 2007).

The heterokont trees embracing the three phyla Ochrophyta, Bigyra and Pseudofungi (Figs 1 and S1) indicated that the heterotrophic heterokonts (Bigyra and pseudofungi) have higher evolutionary rates (inferred from branch length) than Ochrophyta. This evolutionary rate was considerably reduced with the exclusion of the fastest-evolving nucleotide and amino acid sites from the alignment (Fig. 2). Among all tested models of

sequence evolution the covarion model (Huelsenbeck 2002) fitted the data with the highest log likelihood. The covarion model is most appropriate for this dataset due to changes in site-rate pattern, in which sites switch between invariable and variable states across the heterokont tree — similarly to the uneven rates observed among the lineages of other protist nuclear and plastid genome sequences (Shalchian-Tabrizi et al. 2006 a,b,c). Application of the covarion model on the multigene data resulted in a tree with better support for the basal branches of heterokonts than earlier shown in studies of global heterokont phylogeny (Ben Ali et al. 2002; Cavalier-Smith and Chao 2006; Edvardsen et al. 2007).

In conclusion, some nodes in Ochrophyta were well resolved (Figs 1 and 2), but despite the amount of data (up to 5185 characters), concatenating six genes in a multigene analysis, we were not able to resolve the phylogeny of heterotrophic heterokonts. Our four gene protein tree (Fig. S2) as well as the combined rDNA+protein tree (Fig. 2) were in congruence with the rDNA tree (Fig. 1). In all our global heterokont trees, Ochrophyta was placed as a monophyletic group clearly divided from the heterotrophic phyla (Bigyra and Pseudofungi).

## Proposed Relationships of Ochrophyta

In spite of several attempts to resolve the evolutionary relationships among heterokonts, the phylogeny among classes in Ochrophyta has been unclear. Previous analyses (Ben Ali et al. 2002; Cavalier-Smith and Chao 2006; Daugbjerg and Andersen 1997; Edvardsen et al. 2007) clarified some interclass relationships in Ochrophyta such as Dictyochophyceae clustering with Pelagophyceae (Ben Ali et al. 2001, 2002; Edvardsen et al. 2007) and likewise Phaeophyceae as sister taxa to Xanthophyceae (Ben Ali et al. 2002). These sister taxa relationships are strongly supported in our analyses (Figs 1 and 2) by receiving maximum support values. When adding *rbcL* the separate analysis of Ochrophyta displayed the three heterokont classes, Bacillariophyceae, Pelagophyceae and Dictyochophyceae (Khakista) as being excluded from the Phaeista lineages with moderately high support (Fig. 3). This topology is contradictory to Figures 1 and 2 as well as Cavalier-Smith and Chao (2006), where the Bacillariophyceae (together with Bolidophyceae) formed a basal group within Ochrophyta. Figure 3 shows Phaeothamniophyceae cluster strongly with Xanthophyceae. These two clades may belong to a single class as earlier suggested (Goertzen and Theriot 2003; Negrisolo et al. 2004). However, more data from deeply diverging xanthophytes are needed to test this scenario.

The position of Pinguiophyceae, Chrysophyceae, Eustigmatophyceae and Raphidophyceae within Phaeista varied in almost all trees, and these seem to be the clades with weakest affiliation among all ochrophytes.

## Proposed Relationships between Heterotrophic Heterokonts (Bigyra and Pseudofungi)

The basal heterotrophic branch (Figs 1 and 2) consisting of Bigyra (Bicoecida, *Blastocystis* and Labyrinthulida) is consistent with previous studies (Cavalier-Smith and Chao 2006) but only weakly supported. The branch consisting of oomycetes and *Hyphochytrium catenoides* (forming Pseudofungi) was strongly supported, and the monophyletic inclusion, with moderate support, of *Developayella elegans* within the pseudofungi was achieved when the class of fastest-evolving sites was excluded from the analysis. Both Bigyra and Pseudofungi were excluded from the Ochrophyta congruent with earlier studies (Cavalier-Smith and Chao 2006; Leipe et al. 1994, 1996), but here shown with high support in Bayesian and maximum likelihood analyses.

## Molecular Phylogeny and Ultrastructure

Taken all our trees, plus known ultrastructural traits into consideration, we propose that Bacillariophyceae, Dictyochophyceae and Pelagophycae are related and placed within the subphylum Khakista, and separated from subphylum Phaeista (Cavalier-Smith and Chao 2006). This is consistent with the morphological features as the lineages belonging to Khakista have reduced flagellar apparatus (Cavalier-Smith and Chao 2006; Leipe et al. 1994) with no or vestigial flagellar roots (Guillou et al. 1999) and usually no distal transitional helix (members of the dictyochophyte order Florenciellales is an exception; Edvardsen et al. (2007), but may have a proximal transitional helix (found in some dictyochophytes and pelagophytes) in the forwardly directed flagellum (Guillou et al. 1999). Furthermore, they lack the pigment violaxanthin and possess chlorophyll $c_3$ (Andersen 2004; Bjørnland and Liaaen-Jensen 1989; Edvardsen et al. 2007). The basal bodies are placed in a

depression in the nucleus in Dictyochophyceae and Pelagophyceae (Edvardsen et al. 2007; Honda et al. 1995) and in Bacillariophyceae, only one basal body may be present in the flagellate stage (Preisig 1989). In the remaining taxa, the basal bodies are placed anterior to the nucleus. Contrary to the original definition of Khakista (Cavalier-Smith and Chao 2006), both the molecular and ultrastructural data indicate that Pelagophyceae and Dictyochophyceae should be moved from subphylum Phaeista to Khakista. The existence of a monophyletic group consisting of Bacillariophyceae, Pelagophyceae and Dictyochophyceae has previously been suggested from ultrastructural, biochemical and ssu rDNA data (Potter et al. 1997; Saunders et al. 1995, Saunders et al. 1997).

This study indicated at least two heterotrophic heterokont clades (Fig. 1) — and possibly more (Figs 2 and S1). Since the number of heterotrophic heterokont taxa is limited in this study, more data from other groups will probably further resolve the heterotrophic heterokont tree. Assuming two heterotrophic clades, as our data may suggest, two independent plastid losses seem to be the most likely scenario given a common photosynthetic ancestor of heterokonts (Cavalier-Smith 1999). Alternatively, assuming a non-photosynthetic origin of heterokonts (see Sanchez-Puerta and Delwiche 2008), only a single acquisition of a red algal (or red algal-derived) plastid in the ancestral Ochrophyta needs to be invoked. Conversely, the evolution of plastids in heterokonts may be more complex, for example, invoked independent plastid losses in some Ochrophyta lineages (e.g. *Oikomonas* with the Chrysophyceae clade; see Cavalier-Smith and Chao 2006).

This study has further resolved the phylogeny of heterokonts. In particular, the topology within Ochrophyta was better resolved and their separation from heterotrophic heterokonts was robustly demonstrated. However, despite the substantial increase in sequence data in this work, there are some ambiguities among the heterotrophic phyla, Bigyra and Pseudofungi. As most of the character information in our data came from the rDNA genes we believe it is worth generating an even larger alignment with more genes and increased number of taxa preferably from the classes Bolidiophyceae, Chrysomerophyceae, Placididea and several of the heterotrophic classes including Opalinidea as well as the recently identified pico-sized MAST clades (Kolodziej and Stoeck 2007; Massana et al. 2006). Extending our analyses with improved taxon and gene sampling combined with ultrastructural characters will enable us to better understand the evolution of this diverse group of organisms.

## Methods

**Algal cultures**: In total 19 different algal strains (Supplementary Table 1) were grown in media and temperature according to the specific recommendations from culture collections. The dictyochophyte *Pseudochattonella farcimen* strain UIO110 was grown in IMR1/2 medium (Eppley et al. 1967) added 10 nM selenite and with salinity 25, at the temperature 14—15 °C and irradiance of about 100 μmol photons $m^{-2}$ $s^{-1}$.

**DNA isolation and PCR amplification**: Algal cells were harvested directly from the culture (1 ml) by centrifugation (Eppendorf 5415R, Eppendorf, Hamburg, Germany) at 3220*g* for 5 min. Organisms grown on solid surfaces (bottom of Petri dishes or agar surfaces) were scraped off and dissolved in molecular biology grade water prior to centrifugation. DNA from the algal pellets was isolated using standard procedure of Dynabeads DNA direct Universal kit (Invitrogen, CA, USA) according to manufactures recommendation.

The primers used for amplifications by PCR are shown in Supplementary Table 2. For degenerate primers the final concentration was doubled (20 μM). When nested PCR reactions were required, the initial PCR products were diluted 1:20 in molecular biology grade water. Amplifications of PCR products were optimized and carried out in a Master Cycler (Eppendorf). Two different DNA polymerases HotMaster (Eppendorf) and Phusion (Finnzymes OY, Espoo, Finland) were used (Supplementary Table 2).

**Molecular cloning and sequencing**: All PCR products were examined for correct length on a 1.0% agarose gel stained with ethidium bromide. PCR products were purified from agarose gels, with WizardSV Gel and PCR Clean-Up System (Promega, WI, USA). Depending on polymerase used in PCR we cloned PCR products into TOPO cloning pCR 2.1 vector (Invitrogen, CA, USA) or PJET vector (Fermentas, St. Leon-Rot, Germany) and transformed into TOP10 competent bacterial cells (Invitrogen). Plasmids were purified by use of Wizard plus SV miniprep (Promega) and bidirectionally sequenced using the standard M13 sequencing primers or PJET sequencing primers (Fermentas). Internal sequencing primers (Supplementary Table 2) were also used. One to five clones from each PCR product were sequenced and compared.

**Single and multigene alignments**: Additional non-amplified sequences used during this study were obtained from the publicly available GenBank databases (http://www.ncbi.nlm.nih.gov/). Sequences were compared and manually edited in Vector NTI (Invitrogen). Identification of different paralogous gene copies was performed using Pfam gene family seed alignments (Finn et al. 2006) and non-orthologous gene copies were subsequently removed prior to concatenation. Sequences were aligned to the Pfam alignments using ClustalW (Thompson et al. 1994) and subsequently edited manually in BioEdit v.7.0.5 (Tippmann 2004) and MacClade (Vazquez 2004). The single gene alignments were then inferred using RAxML version 7.0.3 (Stamatakis 2006) with 100 bootstrap replicates, random starting trees and either the GTRMIX (rDNA) or PROTMIXrtREV(F) (protein) substitution model. This determined if the genes had phylogentically compatible histories (no strongly supported conflicting clades). In all cases, no paralogous evidence and no strongly

supported (i.e. bootstrap support >55%) topological conflicts were observed (Figs S4—S10). The only exceptions to this rule were the placement of *Ochromonas danica* (Ochromonadaceae) as sister to *Caecitelius parvulus* in the *cox1* tree (Fig. S8) and the placement of *Nerada mexicana* within the outgroup in the *hsp90* tree (Fig. S9). Both these topologies and high bootstrap support values are most likely due to artefacts caused by biased taxon sampling; for *cox1*, *Ochromonas danica* and *Caecitelius* are the singular representatives of Chrysophyceae and Bigyra, whilst for *hsp90*, *Nerada mexicana* is adversely affected as being the only representative from the basal Bigyra taxa (Bicosoecida and Blastocystis). Removing *Ochromonas danica cox1* and *Nerada mexicana hsp90* sequences for the multigene inferences did not change the topology nor maximum likelihood bootstrap values significantly from that presented in Figures 1 and 2 (results not shown).

The choice of outgroup has been a contentious issue in resolving heterokont evolutionary relationships (Cavalier-Smith and Chao 2006). An outgroup distant to the ingroup can result in long branch attraction (LBA) affecting the ingroup relationships (Hillis 1998; Rannala et al. 1998). However, recent articles that have inferred global eukaryotic phylogenies based upon 123+ genes (29908+ amino acid characters) show a clear relationship between the heterokonts, Rhizaria and the alveolates (ciliates, Apicomplexa and dinoflagellates), which form the SAR clade (Burki et al. 2007, 2008). It was therefore important to include taxa from these lineages in addition to the chromist divisions of Cryptophyta and Haptophyta, which make up the chromalveolates with the heterokonts and alveolates, for the enhanced resolution of heterokont evolution.

Two genes (β-tubulin and *rbcL*) from the dictyochophyte *Pseudochattonella farcimen* strain UIO110 originated from an EST library of this organism (Riisberg et al. unpublished; accessions FJ030894 and AM850235, respectively).

Taxa were chosen to prevent missing character data for the multiple gene alignments. However, as missing character data was extensive for some species, composite (in silico chimeric) sequences were constructed from closely related species. All taxa had less than 35% missing character data in the resulting alignments (for further details, see Table 1) and in the resulting supermatrices the proportion of missing data never exceeded 15% (14.37% missing data in Fig. 3).

**Phylogenetic analyses and AU test**: The evolutionary models were chosen on the basis of BIC and AIC criterions implemented in ProtTest version 1.3, Modeltest and Treefinder (Abascal et al. 2005; Jobb et al. 2004; Posada and Crandall 1998). The best fitting evolutionary models for the protein partitions were estimated to be the rtREV+F (frequency) model. For the rDNA partition the GTR evolutionary model was preferred. For all analyses we accounted for invariable sites and among-site rate variation by assuming a gamma distribution of site rates approximated with four rate categories. In analyses of concatenated data was divided into rDNA and protein data partitions and the best fitting models of evolution for each partition applied.

Bayesian analyses were carried out with MrBayes MPI version 3.1.2 (Huelsenbeck et al. 2001; Ronquist and Huelsenbeck 2003) with the best fitting BIC model estimated in Prottest, Modeltest and Treefinder. Bayesian covarion trees were generated from two runs with one heated and three cold chains in the Metropolis-coupled Monte Carlo Markov Chain (MCMC) with 4,000,000 generations implemented. Both covarion and non-covarion phylogenies were inferred, however the covarion model, which allows for different between-site evolutionary rates was preferred in all cases with significantly higher marginal likelihoods for the runs. Burn-in trees were set based on the assessment of likelihood plots and convergence diagnostics implemented in MrBayes. Tree sampling was done every 100 generations, and after burn-in the consensus of the sampled trees used to calculate the posterior probability of the trees topology.

Maximum likelihood (ML) analyses were performed with Treefinder (Jobb et al. 2004) and RAxML version 7.0.3 (Stamatakis 2006). Bootstrap analyses were done by 100 pseudoreplicates with the same evolutionary model as the initial search. Trees were inferred from RAxML under GTR(PROT)MIX, with an initial inference under the GTR(PROT)CAT approximation (optimization of individual per-site substitution rates and classification of those individual rates under 25 specified rate categories), before the GTR(PROT)GAMMA model with 4 discrete GAMMA rate categories was used to evaluate the final tree topology, such that it yielded stable likelihood values (Stamatakis 2006).

To investigate the impact of the fast-evolving heterotrophic taxa on the phylogenetic inference, fast-evolving sites were identified using PAML (Yang 1997, 2007). The rDNA (lsu and ssu) dataset was assessed by estimation with baseML, whilst the protein dataset (actin, β-tubulin, *cox1* and *hsp90*) was assessed by estimation with codonML. For both estimations, sites were classified under 8-rate categories, and subsequent site removal script applied to the alignment (Kumar et al. unpublished). Alignments were constructed with sites 8 and sites 7 and 8 removed; however, subsequent analysis showed that removal of more than site 8 resulted in a highly reduced alignment and therefore phylogenetic resolution (results not shown) (Kumar et al. unpublished).

The paired-site approximately unbiased (AU) test (Shimodaira 2002) was used to compare incongruent heterotrophic topologies and implemented in Treefinder (Jobb et al. 2004). The heterotrophic topologies to be tested were reoptimized (multifurcations resolved) in Treefinder before AU testing. The alignment used was based on Figure S1 (lsu, ssu, actin, β-tubulin, *cox1* and *hsp90*), where no fast evolving were sites excluded. Taxon choice for the autotrophic heterokonts was also reduced, with the exclusion of Phaeista taxa from the dataset. The AU test was performed with 1,000,000 bootstrap replicates with the same evolutionary models and partitions as described earlier (GTR and rtREV+F). The tested topologies and results are shown in Supplementary data.

All model estimation and phylogenetic analyses were done on the freely available Bioportal at the University of Oslo (http://www.bioportal.uio.no/).

## Note added in proof

Tsui et al. (2008) recently published a global heterokont phylogeny using three protein-coding genes focusing mainly on the relationship between Labyrinthulida and Bicosoecida (Tsui et al. 2008). In congruence with our topologies they obtained statistical support for the separation of Bigyra from Pseudofungi and separation between the heterotrophs and Ochrophyta. However in Tsui et al., Labyrinthulida was moderately supported as sister to Bicosoecida. In our trees,

based on seven gene sequences and different taxon sampling, we found the latter two as weakly supported paraphyletic groups.

## Acknowledgements

We would like to thank Thomas Cavalier-Smith for discussions and Ema Chao for DNA from *Nerada mexicana*, Cedric Berney for lsu and ssu alignments of available heterokont sequences, and Ave Tooming-Klunderud for PCR amplifying and cloning of the actin sequence for the Labyrinthulida *Thraustochytrium aureum.* We thank the Bioportal at the University of Oslo for computer resources. The study was financially supported by grants from the Norwegian Research Council (Projects no. 166555 and 172572) to KSJ, PhD and by the University of Oslo and the Norwegian University of Life Sciences for strategic funding (Trippelallliansen). The seven single-gene alignments used in this study have been made freely available for public access at the EMBL-Align database: ftp://ftp.ebi.ac.uk/pub/databases/embl/align/ under the accession numbers ALIGN_001291—ALIGN_001297.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.protis.2008.11.004.

## References

**Abascal F, Zardoya R, Posada D** (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**: 2104—2105

**Andersen RA, Van de Peer Y, Potter D, Sexton JP, Kawachi M, LaJeunesse T** (1999) Phylogenetic analysis of the SSU rRNA from members of the Chrysophyceae. Protist **150**: 71—84

**Andersen RA** (2004) Biology and systematics of heterokont and haptophyte algae. Am J Bot **91**: 1508—1522

**Bachvaroff TR, Sanchez-Puerta MV, Delwiche CF** (2006) Rate variation as a function of gene origin in plastid-derived genes of peridinin-containing dinoflagellates. J Mol Evol **62** 42-U27

**Ben Ali A, De Baere R, Van der Auwera G, De Wachter R, Van de Peer Y** (2001) Phylogenetic relationships among algae based on complete large-subunit rRNA sequences. Int J Syst Evol Microbiol **51**: 737—749

**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. Systemat Assoc, Clarendon Press, Oxford, pp 37—60

**Bodyl A** (2005) Do plastid-related characters support the chromalveolate hypothesis? J Phycol **41**: 712—719

**Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J** (2007) Phylogenomics reshuffles the eukaryotic supergroups. PLoS ONE **2**: e790

**Burki F, Shalchian-Tabrizi K, Pawlowski J** (2008) Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett **4**: 366—369

**Cavalier-Smith T** (1986) The Kingdom Chromista: Origin and Systematics. In Round FE, Chapman DJ (eds) Progress in Phycological Research, Vol. 4. Biopress Ltd., Bristol, pp 309—347

**Cavalier-Smith T** (1997) Sagenista and Bigyra, two phyla of heterotrophic heterokont chromists. Arch Protistenkd **148**: 253—267

**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—366

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

**Daugbjerg N, Guillou L** (2001) Phylogenetic analyses of Bolidophyceae (Heterokontophyta) using rbcL gene sequences support their sister group relationship to diatoms. Phycologia **40**: 153—161

**Daugbjerg N, Andersen RA** (1997) A molecular phylogeny of the heterokont algae based on analyses of chloroplast-encoded *rbc*L sequence data. J Phycol **33**: 1031—1041

**Edvardsen B, Eikrem W, Shalchian-Tabrizi K, Riisberg I, Johnsen G, Naustvoll L, Throndsen J** (2007) *Verrucophora farcimen* gen. et sp. nov. (Dictyochophyceae, Heterokonta) — a bloom-forming ichthyotoxic flagellate from the Skagerrak, Norway. J Phycol **43**: 1054—1070

**Eppley R, Holmes RW, E P** (1967) Periodicity in cell division and physiological behaviour of *Ditylum brightwellii*, a marine planktonic diatom during growth in light—dark cycles. Arch Mikrobiol **56**: 305—323

**Fast NM, Xue LR, Bingham S, Keeling PJ** (2002) Re-examining alveolate evolution using multiple protein molecular phylogenies. J Eukaryot Microbiol **49**: 30—37

**Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A** (2006) Pfam: clans, web tools and services. Nucleic Acids Res **34**: D247—D251

**Goertzen LR, Theriot EC** (2003) Effect of taxon sampling, character weighting, and combined data on the interpretation of relationships among the heterokont algae. J Phycol **39**: 423—439

**Guillou L, Chretiennot-Dinet MJ, Medlin LK, Claustre H, Loiseaux-de Goer S, Vaulot D** (1999) *Bolidomonas*: a new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). J Phycol **35**: 368—381

**Hillis DM** (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst Biol **47**: 3—8

**Honda D, Kawachi M, Inouye I** (1995) *Sulcochrysis biplastida* gen. et sp. nov.: cell structure and absolute configuration of the flagellar apparatus of an enigmatic chromophyte alga. Phycol Res **43**: 1—16

**Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP** (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. Science **294**: 2310—2314

**Huelsenbeck JP** (2002) Testing a covariotide model of DNA substitution. Mol Biol Evol **19**: 698—707

**Jobb G, von Haeseler A, Strimmer K** (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol **4**: 18

**Kim E, Simpson AGB, Graham LE** (2006) Evolutionary relationships of apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. Mol Biol Evol **23**: 2455—2466

**Kolodziej K, Stoeck T** (2007) Cellular identification of a novel uncultured marine stramenopile (MAST-12 clade) small-subunit rRNA gene sequence from a Norwegian estuary by use of fluorescence *in-situ* hybridization-scanning electron microscopy. Appl Environ Microbiol **73**: 2718—2726

**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 ribosomal-RNA sequences from *Labyrinthuloides minuta* and *Cafeteria roenbergensis*. Phycologia **33**: 369—377

**Leipe DD, Tong SM, Goggin CL, Slemenda SB, Pieniazek NJ, Sogin ML** (1996) 16S-like rDNA sequences from *Developayella elegans*, *Labyrinthuloides haliotidis*, and *Proteromonas lacertae* confirm that the stramenopiles are a primarily heterotrophic group. Eur J Protistol **32**: 449—458

**Massana R, Terrado R, Forn I, Lovejoy C, Pedros-Alio C** (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. Environ Microbiol **8**: 1515—1522

**McMahon MM, Sanderson MJ** (2006) Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. Syst Biol **55**: 818—836

**Moriya M, Nakayama T, Inouye I** (2000) Ultrastructure and 18S rDNA sequence analysis of *Wobblia lunata* gen. et an. nov., a new heterotrophic flagellate (Stramenopiles, Incertae sedis). Protist **151**: 41—55

**Negrisolo E, Maistro S, Incarbone M, Moro I, Dalla Valle L, Broady PA, Andreoli C** (2004) Morphological convergence characterizes the evolution of Xanthophyceae (Heterokontophyta): evidence from nuclear SSU rDNA and plastidial *rbc*L genes. Mol Phylogenet Evol **33**: 156—170

**Nosenko T, Bhattacharya D** (2007) Horizontal gene transfer in chromalveolates. BMC Evol Biol **7**: 173

**Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D** (2004) Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol **21**: 1740—1752

**Posada D, Crandall KA** (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics **14**: 817—818

**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**: 966—972

**Preisig HR** (1989) The Flagellar Base Ultrastructure and Phylogeny of Chromophytes. In Green JC, Leadbeater BSC, Diver WL (eds) The Chromophyte Algae: Problems and Perspectives. Clarendon Press, Oxford, pp 167—187

**Rannala B, Huelsenbeck JP, Yang ZH, Nielsen R** (1998) Taxon sampling and the accuracy of large phylogenies. Syst Biol **47**: 702—710

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

**Sanchez-Puerta MV, Delwiche CF** (2008) A hypothesis for plastid evolution in chromalveolates. J Phycol **44**: 1097—1107

**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

**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

**Shalchian-Tabrizi K, Eikrem W, Klaveness D, Vaulot D, Minge MA, Le Gall F, Romari K, Throndsen J, Botnen A, Massana R, Thomsen HA, Jakobsen KS** (2006a) *Telonemia*, a new protist phylum with affinity to chromist lineages. Proc R Soc Biol Sci B **273**: 1833—1842

**Shalchian-Tabrizi K, Minge MA, Cavalier-Smith T, Nedreklepp JM, Klaveness D, Jakobsen KS** (2006b) Combined heat shock protein 90 and ribosomal RNA sequence phylogeny supports multiple replacements of dinoflagellate plastids. J Eukaryot Microbiol **53**: 217—224

**Shalchian-Tabrizi K, Skanseng M, Ronquist F, Klaveness D, Bachvaroff TR, Delwiche CF, Botnen A, Tengs T, Jakobsen KS** (2006c) Heterotachy processes in rhodophyte-derived secondhand plastid genes: Implications for addressing the origin and evolution of dinoflagellate plastids. Mol Biol Evol **23**: 1504—1515

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

**Simpson AGB, Inagaki Y, Roger AJ** (2006) Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol Biol Evol **23**: 615—625

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

**Thompson JD, Higgins DG, Gibson TJ** (1994) Clustal-W — improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap

penalties and weight matrix choice. Nucleic Acids Res **22**: 4673—4680

**Tippmann HF** (2004) Analysis for free: comparing programs for sequence analysis. Brief Bioinform **5**: 82—87

**Tsui CK, Marshall W, Yokoyama R, Honda D, Lippmeier JC, Craven KD, Peterson PD, Berbee ML** (2008) Labyrinthulomycetes phylogeny and its implications for the evolutionary loss of chloroplasts and gain of ectoplasmic gliding. Mol Phylogenet Evol **50**: 129—140

**Vazquez J** (2004) Phylogenetics — MacClade 4: analysis of phylogeny and character evolution, version 4.06. Am Biol Teacher **66**: 511—512

**Wiens JJ** (2005) Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Syst Biol **54**: 731—742

**Wiens JJ** (2006) Missing data and the design of phylogenetic analyses. J Biomed Inform **39**: 34—42

**Wiens JJ, Moen DS** (2008) Missing data and the accuracy of Bayesian phylogenetics. J Syst Evol **46**: 307—314

**Wiens JJ, Reeder TW** (1995) Combining data sets with different numbers of taxa for phylogenetic analysis. Syst Biol **44**: 548—558

**Yang Z** (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci **13**: 555—556

**Yang Z** (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24**: 1586—1591