International Journal of Systematic and Evolutionary Microbiology (2005), **55**, 487–496 **DOI** 10.1099/ijs.0.63216-0

On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes

James T. Harper, Esmé Waanders and Patrick J. Keeling

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, Canada V6T 1Z4

Correspondence
Patrick J. Keeling
pkeeling@interchange.ubc.ca

A global phylogeny of major eukaryotic lineages is a significant and ongoing challenge to molecular phylogenetics. Currently, there are five hypothesized major lineages or 'supergroups' of eukaryotes. One of these, the chromalveolates, represents a large fraction of protist and algal diversity. The chromalveolate hypothesis was originally based on similarities between the photosynthetic organelles (plastids) found in many of its members and has been supported by analyses of plastid-related genes. However, since plastids can move between eukaryotic lineages, it is important to provide additional support from data generated from the nuclear-cytosolic host lineage. Genes coding for six different cytosolic proteins from a variety of chromalveolates (yielding 68 new gene sequences) have been characterized so that multiple gene analyses, including all six major lineages of chromalveolates, could be compared and concatenated with data representing all five hypothesized supergroups. Overall support for much of the phylogenies is decreased over previous analyses that concatenated fewer genes for fewer taxa. Nevertheless, four of the six chromalveolate lineages (apicomplexans, ciliates, dinoflagellates and heterokonts) consistently form a monophyletic assemblage, whereas the remaining two (cryptomonads and haptophytes) form a weakly supported group. Whereas these results are consistent with the monophyly of chromalveolates inferred from plastid data, testing this hypothesis is going to require a substantial increase in data from a wide variety of organisms.

INTRODUCTION

A well-resolved global phylogeny of eukaryotes has been a major goal in understanding eukaryotic evolution and biodiversity because eukaryotic macroevolutionary events can only be interpreted accurately within the context of the relationships among these organisms. For example, plastids (the photosynthetic organelles of plants and algae) have spread between eukaryotic lineages via secondary endosymbiosis, but without a clear understanding of the relationships between plastid-bearing eukaryotes, it is very difficult to trace the history of plastid gains and losses accurately (Archibald & Keeling, 2002). Indeed, plastids illustrate this need especially well, since much of what is known about algal relationships is based on plastid-encoded genes or genes for plastid-targeted proteins (Fast *et al.*, 2001; Harper & Keeling, 2003; Yoon *et al.*, 2002b), which can evolve independently of the host. This does not invalidate such inferences, but it is desirable to buttress these conclusions with supporting data from host lineages.

The first large-scale molecular phylogenies of eukaryotes, based on small-subunit rRNA (SSU rRNA), consisted of a basal ladder of protists followed by an explosive 'crown' of other eukaryotes – animals, fungi and plants (Cavalier-Smith & Chao, 1996; Kumar & Rzhetsky, 1996; Sogin & Silberman, 1998; Van de Peer & De Wachter, 1997). The first protein-coding gene phylogenies gave similar results (Brown & Doolittle, 1995; Hashimoto *et al.*, 1994; Kamaishi *et al.*, 1996), but also contained important incongruencies and over time, these contradictions have added up to serious questions about the 'shape' of the eukaryotic tree. Comparative phylogenetics based on several well-sampled protein-coding genes has shown that the overall ladder–crown structure of the tree is not a true representation of eukaryotic evolution. Nearly all of the statistically well-supported 'deep-branching' eukaryotes are now known to be artifacts of heterogeneous evolutionary rates or long-branch attraction (Embley & Hirt, 1998; Gribaldo & Philippe, 2002; Hirt *et al.*, 1999; Morin, 2000; Philippe, 2000; Van de Peer *et al.*, 2000b). All individual gene phylogenies seem to contain a few misleading branches and problematic taxa, but fortunately different trees seem

Abbreviations: EF-1 alpha, elongation factor-1 alpha; EST, expressed sequence tag; gDNA, genomic DNA; HSP, heat-shock protein; ML, maximum-likelihood; SSU rRNA, small-subunit rRNA.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AY729814–AY729874 and AY739894–AY739895.

Individual ProML phylogenies (Figs A–E) are available as supplementary material in IJSEM Online.

to have different problems. Thus, reconstructing a global eukaryotic tree seems to remain within reach.

A new hypothesis is emerging based on several different kinds of mutually reinforcing data from many different genes. These data can take the following forms: a single phylogeny based on multiple concatenated genes (Bapteste *et al.*, 2002; Yoon *et al.*, 2002b); several individual phylogenies that all converge on the same relationship (Fast *et al.*, 2002; Simpson *et al.*, 2002b); non-phylogenetic molecular evidence such as insertions, deletions, gene duplications or gene fusions (Archibald *et al.*, 2002; Baldauf & Palmer, 1993; Keeling & Palmer, 2001; Stechmann & Cavalier-Smith, 2002); and/or non-molecular data such as ultrastructure (Simpson *et al.*, 2002a). The hypothesis emerging from this approach is that there are five major divisions or 'supergroups' of eukaryotes representing most or perhaps even all eukaryotic diversity (Keeling, 2004). Many of these do not have formal taxonomic names that are universally accepted, so for simplicity they can be informally referred to as the rhizaria, chromalveolates, excavates, unikonts and plants.

The chromalveolate supergroup represents a large fraction of known eukaryotic diversity, ranging from tiny microbial parasites to kelps as long as 50 m. In fact, chromalveolates account for about half of the recognized species of protists and algae (Cavalier-Smith, 2004). The chromalveolates comprise six subgroups, each of which is a large and diverse group in its own right. The alveolates contain the apicomplexans, ciliates and dinoflagellates, whereas the chromists are made up of the cryptomonads, haptophytes and heterokonts (also called stramenopiles). Various permutations including some, but not all, of these lineages have been put forward in the past (e.g. Chadefaud, 1950; Christensen, 1989), but these were largely restricted to photosynthetic members and included some groups (e.g. euglenids) that are now known to be very distantly related to other putative chromalveolates. Moreover, the current chromalveolate hypothesis is based on a very specific model of plastid evolution. Cryptomonads, haptophytes, heterokonts and dinoflagellates all have members that contain plastids derived from secondary endosymbiosis with a red alga, whereas apicomplexans contain a cryptic plastid of putative red algal origin (Blanchard & Hicks, 1999; Fast *et al.*, 2001; Funes *et al.*, 2002; Köhler *et al.*, 1997; McFadden & Waller, 1997; Williamson *et al.*, 1994; Zhang *et al.*, 2000). Based on this, it was proposed that these plastids originated from a single endosymbiotic event in their common ancestor (Cavalier-Smith, 1999).

A single endosymbiotic event also leads to the prediction that both the host lineages and their plastids share a common evolutionary history. The only molecular evidence, in addition to other kinds of evidence, supporting this supergroup comes from plastid-targeted and plastid-encoded proteins. In the first instance, plastid-targeted GAPDH has been shown to be the product of a rare gene duplication of the cytosolic homologue, which has replaced the ancestral, cyanobacterial form (Fast *et al.*, 2001; Harper & Keeling, 2003). This replacement is a strong marker for a common origin of chromalveolate plastids. From the plastid genome itself, a phylogeny based on five concatenated genes supports the union of cryptomonads, haptophytes and heterokonts (Yoon *et al.*, 2002a). Given that plastids can move between lineages, however, these conclusions will be significantly stronger if the host lineages can also be demonstrated to be related. Molecular data have confirmed the alveolates as a group (Fast *et al.*, 2002; Van de Peer *et al.*, 2000a; Van de Peer & De Wachter, 1997) and there are data from a four-gene concatenated set indicating that ciliates, apicomplexans and heterokonts form a group (Baldauf *et al.*, 2000). However, no large-scale analysis of nuclear-cytoplasmic genes has included all major chromalveolate lineages and, moreover, none has included representation from the remaining four supergroups, which is desirable to make certain than none of these are positioned between members of the supergroup being tested.

Here, the monophyly of the chromalveolates has been examined within the context of a global phylogeny of eukaryotes using six nucleus-encoded genes for cytoplasmic proteins: actin, alpha-tubulin, beta-tubulin, elongation factor-1 alpha (EF-1 alpha), heat-shock protein 70 (HSP70) and heat-shock protein 90 (HSP90). Homologues of these genes were characterized from a variety of chromalveolates so that each gene was known from at least one representative of all six chromalveolate subgroups. In addition, genes from representatives of two other supergroups (rhizarians and excavates) have also been characterized since molecular data from these groups are sparse. Consequently, this is the first analysis that includes multiple gene data from all five supergroups.

METHODS

Strains and culture conditions. Unialgal axenic cultures of the cryptomonads *Guillardia theta* CCMP 327 and *Rhodomonas salina* CCMP 1319, the dinoflagellate *Heterocapsa triquetra* CCMP 449, the haptophytes *Isochrysis galbana* CCMP 1323, *Pavlova lutheri* CCMP 1325 and *Prymnesium parvum* CCMP 1926 and the heterokonts *Cyclotella cryptica* CCMP 333, *Mallomonas rasilis* CCMP 478 and *Phaeodactylum tricornutum* CCMP 1327 were obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (McKown Point, West Boothbay Harbor, ME, USA) and grown in f/2-Si medium at 16 °C (12 : 12 light : dark cycle).

The heterokont *Spumella uniguttata* was the prey organism for a culture of the apicomplexan *Colpodella edax* isolated from a freshwater pond near Borok (Yaroslavskaja, Russian Federation). Genes were obtained from *Spumella uniguttata* in several attempts to sequence the *Colpodella edax* homologues (Kuvardina *et al.*, 2002).

Genomic DNAs (gDNAs) from the oomycete heterokonts *Apodachlya brachynema* CBS 557.69, *Brevilegnia macrospora* CBS 132.37, *Phytophthora palmivora* CBS 236.30, *Plectospira myriandra* CBS 523.87, *Pythium graminicola* CBS 327.62 and *Thraustotheca clavata* CBS 343.33 were kindly donated by A. W. DeCock (Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands) and gDNA from the raphidophyte heterokont *Heterosigma akashiwo* was kindly donated by K. Ishida (Department of Biology, Kanazawa University, Kanazawa, Japan).

Table 1. Taxa and genes represented in analyses of the four-gene and six-gene concatenated datasets

Species	Actin	alpha-Tubulin	beta-Tubulin	HSP90	EF-1 alpha	HSP70
Chromalveolates						
Apicomplexans						
Cryptosporidium parvum	+	+	+	+	+	+
Eimeria tenella	+	+[a]*	+	+	+[b]	+[a]
Plasmodium falciparum	+	+	+	+	+	+
Ciliates						
Paramecium tetraurelia	+	+	+	+	+	−
Tetrahymena pyriformis	+	+	+	+	+	−
Tetrahymena thermophila	+	+	+	+	−	+
Dinoflagellates						
Crypthecodinium cohnii	+	+[c]	+	+	−	+
Heterocapsa triquetra	+	+	+	+	−†	+
Oxyrrhis marina	+	+	+	+	−	−
Perkinsus marinus	+	+	+	+	−	−
Cryptomonads						
Guillardia theta	+	+	+	+	+[d]	+
Haptophytes						
Isochrysis galbana	+	+	+	+	−†	−
Pavlova lutheri	−	+	+	+	−	−
Prymnesium parvum	+	+	+	+	−	−
Heterokonts						
Apodachlya brachynema	+	+	+	+	+	−
Brevilegnia macrospora	+	−	+	+	+	+
Heterosigma akashiwo	+	+	+	+	+	+
Mallomonas rasilis	+	+	+	+	+	−
Phaeodactylum tricornutum	+	+	+	+	+	−
Phytophthora palmivora	+	+	+	+	+	+
Plectospira myriandra	+	+	+	+	+	−
Pythium graminicola	+	+	+	+	−	−
Spumella uniguttata	+		+	+	+	+
Thalassiosira pseudonana	+	+	+	+	+	+
Thraustotheca clavata	+	+	+	+	−	−
Unikonts						
Animals						
Caenorhabditis elegans	+	+	+	+	+	+
Danio rerio	+	+	+	+	+	+
Drosophila melanogaster	+	+	+	+[e]	+	+
Homo sapiens	+	+	+	+	+	+
Rattus norvegicus	+	+	+	+	+	+
Fungi						
Ajellomyces capsulatus	+	+	+	+	+	+
Candida albicans	+	+	+	+	+	+
Neurospora crassa	+	+	+	+	+	+
Saccharomyces cerevisiae	+	+	+	+	+	+
Schizosaccharomyces pombe	+	+	+	+	+	+
Mycetozoans						
Dictyostelium discoideum	+	+	+	+	+	+
Rhizarians						
Chlorarachniophytes						
Bigelowiella natans	+	+	+	+	−†	+
Plants						
Land plants						
Arabidopsis thaliana	+	+	+	+	+	+
Oryza sativa	+	+	+	+	+	+

Table 1. cont.

Species	Actin	alpha-Tubulin	beta-Tubulin	HSP90	EF-1 alpha	HSP70
Zea mays	+	+	+	+	+	+
Green algae						
Chlamydomonas reinhardtii	+	+	+	+	−†	+
Red algae						
Porphyra yezoensis	+	+[f]	+[g]	+	+[g]	+
Excavates						
Diplomonads						
Giardia intestinalis	+	+	+	+[h]	+	+
Parabasalia						
Trichomonas vaginalis	+	+	+	+	+	+
Euglenozoans						
Euglena gracilis	+	+	+	+	+	−
Leishmania donovani	+[i]	+	+[j]	+	+[l]	+[i]
Trypanosoma brucei	+	+	+	+	+	+
Trypanosoma cruzi	+	+	+	+	+	+
Heterolobosea						
Naegleria gruberi	+	+	+	+	+[m]	−
Oxymonads						
Streblomastix strix	−	+	+	+	+	+

*Sequences obtained from species other than those listed are indicated as follows: *a*, *Eimeria acervulina*; *b*, *Eimeria bovis*; *c*, *Amphidinium herdmannii*; *d*, *Rhodomonas salina*; *e*, *Drosophila auraria*; *f*, *Prionitis lanceolata*; *g*, *Porphyra purpurea*; *h*, *Hexamita inflata*; *i*, *Leishmania major*; *j*, *Leishmania mexicana*; *l*, *Leishmania brasiliensis*; *m*, *Naegleria andersoni*.

†The sequence was determined, but was very divergent and was not used in phylogenetic analyses.

Algal cultures were harvested by centrifugation and cell pellets were lysed by grinding in a Knotes Duall 20 tissue homogenizer. gDNAs were extracted using the DNeasy Plant Mini kit (Qiagen) and total RNA was isolated using Trizol reagent (Invitrogen).

Characterization of new gene and cDNA sequences. PCR amplification was carried out using the following primers: actin, GGCCTGGAARCAYTTNCGRTGNAC and GAGAAGATGACNCARATH or TGGGAYGAYATGGARAARATHTGG; alpha-tubulin, CGCGGCCTCARGTNGGNAAYGCNTGYTGGGA and CGCGCCATNCCYTCNCCNACRTACCA; beta-tubulin, GCCTGCAGGNCARTGYGGNAAYCA and TCCTCGAGTRAAYTCCATYTCRTCCAT; EF-1 alpha, AACATCGTCGTGATHGGNCAYGTNGA and CTTGATCACNCCNACNGCNACNGT; HSP70, AAGATCATCGGNATHGAYYTN and CTGAACGATNCCRTTNGCATC; and HSP90, ACGTTYTAYWSNAAYAARGARAT and CGCCTTCATMATNCSYTCCATRTTNGC. PCR products were gel-purified and then cloned into the TOPO-TA vector pCR2.1 (Invitrogen) and multiple clones of each were sequenced on both strands with ABI BigDye terminator chemistry (Applied Biosystems).

HSP90 and HSP70 cDNAs from the host and endosymbiont of the chlorarachniophyte *Bigelowiella natans* CCMP 621 were obtained from an ongoing EST (expressed sequence tag) project and each was fully sequenced. HSP70 cDNA was identified in ongoing EST projects from *Isochrysis galbana* and was fully sequenced. HSP70 from the oxymonad *Streblomastix strix* was amplified using primers AAGATCATCGGNATHGAYYTNGGNACNAC and CTGAACGATNCRTTNGCRTCDATRTCRAA from DNA isolated from enriched hindgut material of the Pacific dampwood termite (*Zootermopsis angusticolis*) as described previously (Keeling & Leander, 2003).

Molecular phylogeny. Newly determined sequences were added to existing amino acid alignments. A total of eight alignments was subjected to phylogenetic analyses. Six of these were the respective alignments for the six protein-coding genes. The two remaining alignments were concatenations of different genes. One was a concatenation of actin, alpha-tubulin, beta-tubulin and HSP90 (AABH), whereas the second was a concatenation of all six genes (AABHHE). The resulting dimensions (number of taxa × number of amino acid positions) for all eight alignments were as follows: actin, 78 × 245; alpha-tubulin, 71 × 391; beta-tubulin, 73 × 380; EF-1 alpha, 48 × 422; HSP70, 45 × 505; HSP90, 61 × 516; AABH, 50 × 1554; and AABHHE, 50 × 2481. The amino acid alignments are available from the authors on request.

Distance analyses were performed on the individual datasets and the four-gene concatenated dataset only (AABH). Parsimony and maximum-likelihood (ML) analyses were performed on all datasets. For the six individual datasets and the four-gene concatenated dataset, ML distances were calculated using TREE-PUZZLE 5.0 (Strimmer & von Haeseler, 1996), using the WAG substitution matrix with the frequency of amino acid usage calculated from the data. Rate-across-site variation was modelled on a discrete gamma distribution with eight rate categories, estimating invariable sites and the shape parameter alpha from the data. Distance trees were constructed with weighted neighbour-joining using WEIGHBOR 1.0.1a (Bruno *et al.*, 2000) and Fitch–Margoliash using FITCH 3.6a (Felsenstein, 1993). Fitch–Margoliash trees were inferred using the global rearrangements option and 10 input-order jumbles. Weighted neighbour-joining and Fitch–Margoliash bootstrap trees were constructed (without global rearrangements and implementing two input-order jumbles in Fitch–Margoliash) from 100 resampled datasets with gamma-corrected distances (with the rate category parameters above) using PUZZLEBOOT 1.0.3 (by M. Holder and A. Roger; http://www.tree-puzzle.de).

Protein ML analyses were performed using ProML 3.6a (Felsenstein, 1993) and PhyML (Guindon & Gascuel, 2003). ProML was performed with amino acid frequencies estimated from the data and rate-across-site variation was modelled on a gamma curve using the -r option with eight variable rate categories and an invariable sites category (rates and frequencies estimated by TREE-PUZZLE). Trees were searched with global rearrangements and a randomized input order of sequences with two jumbles. ProML bootstrap trees were constructed as above, but with no site-to-site rate variation and a single randomized input order. PhyML was performed using an input tree generated by BIONJ, the JTT model of amino acid substitution, the proportion of variable rates estimated from the data and nine categories of substitution rates (eight variable and one variable; parameters estimated by TREE-PUZZLE). PhyML bootstrap trees were constructed using the same parameters as the individual ML trees.

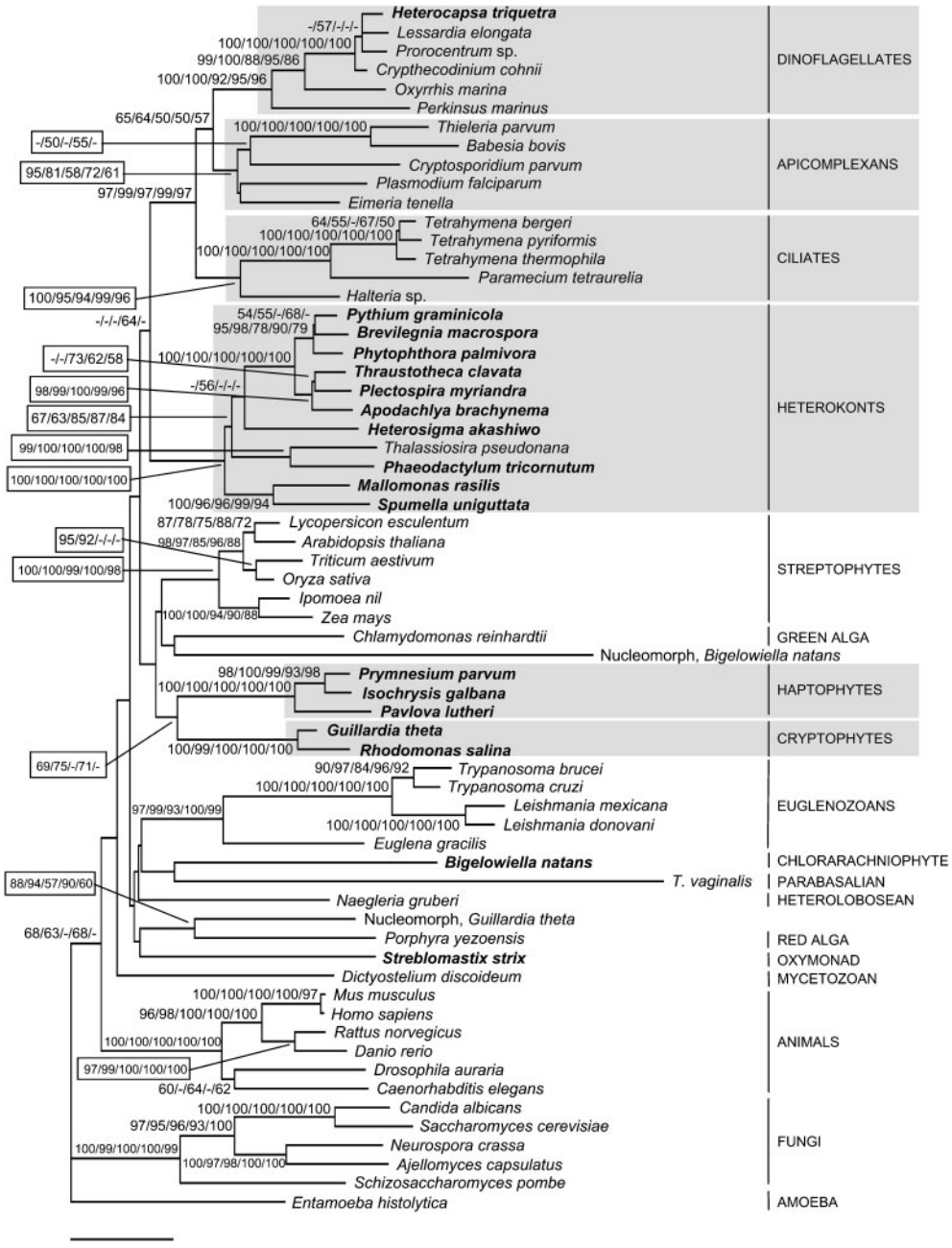

Fig. 1. Protein maximum-likelihood phylogeny (ProML) of HSP90. Bootstrap values are shown for nodes that received support over 50 % and are (left to right) weighted neighbour-joining, Fitch–Margoliash, ProML, PhyML and parsimony (dashes represent support lower than 50 %). Chromalveolate taxa are highlighted in shaded boxes and major groups are bracketed and labelled to the right. Newly determined sequence data are represented by taxon names in bold. Bar, 0·1 changes per site. Abbreviation: *T. vaginalis*, *Trichomonas vaginalis*.

Parsimony analysis was completed using PAUP 4.0b10 (50 random sequence additions; gaps treated as missing data) using a heuristic search with steepest descent and tree bisection-reconnection branch swapping (Swofford, 2002). Parsimony bootstraps were completed (100 replicates) using 10 random sequence additions.

Distance analyses were not performed on the six-gene concatenated dataset due to missing data for a number of taxa (see Table 1). Parsimony, ML and the respective bootstrap analyses were performed on this dataset as described above for the other gene analyses.

RESULTS AND DISCUSSION

Six genes representing five supergroups

Comparable datasets for six genes (actin, alpha-tubulin, beta-tubulin, EF-1 alpha, HSP90 and HSP70) were developed by sequencing 68 new genes or cDNAs from a variety of chromalveolates (primarily the most poorly sampled – dinoflagellates, cryptomonads, haptophytes and heterokonts) and other eukaryotes (Table 1). A table of GenBank/EMBL/DDBJ accession numbers for all taxa included in our various phylogenetic analyses is available as supplementary material in IJSEM Online. The overall aim was for all six major subgroups of chromalveolate and all five hypothesized supergroups to be represented in all six datasets and for representation to be sufficiently similar to allow genes to be concatenated. In addition, missing genes or cDNAs were also characterized from the excavate (oxymonad) *Streblomastix strix* and the rhizarian (the chlorarachniophyte cercozoan) *Bigelowiella natans*. No multiple gene phylogeny of eukaryotes has yet included a rhizarian (e.g. a cercozoan), which is an important omission when testing the monophyly of any one of the other supergroups.

Individual gene phylogenies

Of the six genes analysed here, the phylogeny of HSP90 (Fig. 1) has been investigated the least, but has shown promise. The phylogeny of alveolates has recently been examined with an expanded HSP90 dataset (Leander & Keeling, 2004). Moderate support was found for grouping alveolates and heterokonts, a result that has received varying levels of support from other molecular investigations (Baldauf *et al.*, 2000; Ben Ali *et al.*, 2001; Dacks *et al.*, 2002; Stechmann & Cavalier-Smith, 2003; Van de Peer & De Wachter, 1997). Sequence data for cryptomonads, haptophytes or a diverse sample of heterokonts, however, has been lacking from these analyses. With the addition of these data, our HSP90 phylogeny shows a weak relationship between the strongly supported alveolate and heterokont clades. Interestingly, the strongly supported cryptomonad and haptophyte clades branch together with intermediate to weak support. The positions of both these larger groups, however, remain equivocal.

Phylogenies of the tubulins, actin, EF-1 alpha and HSP70 have all been analysed more extensively than HSP90 (e.g. Baldauf & Palmer, 1993; Baldauf *et al.*, 2000; Bhattacharya & Weber, 1997; Hashimoto *et al.*, 1995; Keeling, 2001;

Table 2. Summary of support for single- and multiple-gene analyses

Except where indicated, bootstrap values are for weighted neighbour-joining, Fitch–Margoliash, ProML and PhyML, respectively. AABH, Concatenation of actin, alpha-tubulin, beta-tubulin and HSP90; AL, alveolates; AN, animals; CH, chromists; CR, cryptophytes; D, diplomonads; E, euglenozoa; F, fungi; G, green algae; HA, haptophytes; HK, heterokonts; HT, heterolobosea; P, parabasalia; R, red algae; and S, streptophytes. NA, Not available.

Gene	CR	HA	HK	CH	AL	AL-HK	AL-CH	S-G	AN-F	E-HT	D-P
HSP90	100/99/100/100	100/100/100/100	100/100/100/100	-/-/-/-	97/99/97/99	-/-/-/64	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-
Actin	96/90/81/74	98/97/99/82	76/56/64/80	-/-/-/-	-/-/-/-	-/65*/-/-	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-
alpha-Tubulin	81/60/95/81	98/92/100/100	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	93/60/84/72	-/-/-/-	-/-/-/-
beta-Tubulin	-/-/-/-	96/94/91/94	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	85/93/83/93	-/-/-/-	99/92/69/62†	72/81/73/57
EF-1 alpha	NA‡	NA§	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	73/82/92/80	-/-/-/-	90/91/67/66
HSP70	NA‡	NA‡	53/77/74/99	-/-/-/-	-/-/-/-	-/-/-/-	-/-/-/-	88/83/65/86	-/-/-/-	-/-/-/-	-/-/-/-
AABH	NA‡	100/100/100/100	99/100/100/100	-/-/-/-	98/97/98/100	66/53/-/59	-/-/-/-	100/100/100/100	-/-/-/78	85/56/61/83	100/98/97/86
AABHHE‖	NA‡	100/99	100/99	-/-	80/93	66/84	-/-	99/99	70/92	-/65	98/96

*Support for grouping apicomplexans, dinoflagellates and heterokonts.
†Support for grouping *Euglena gracilis* and *Naegleria gruberi*.
‡Only one sequence represented this taxon for this gene.
§The haptophyte sequence was unusually divergent and not included in this dataset (Keeling & Inagaki, 2004).
‖Distance analyses were not completed for the six-gene dataset.

Keeling & Doolittle, 1996; Leander & Keeling, 2004; Saldarriaga *et al.*, 2003) and the general characteristics of our phylogenies are similar to those given in these published reports. Bootstrap supports from these analyses are summarized in Table 2 [phylogenies (Figs A–E) are available as supplementary material in IJSEM Online]. Overall, the monophyly of each of the chromist lineages was recovered and strongly supported. As a group, the chromists were not recovered by any of the analyses of the six genes. Alveolates were resolved as monophyletic in actin, with the previously noted exception of ciliates (Keeling, 2001), and HSP70 phylogenies, but not with any appreciable bootstrap support. Not surprisingly, a chromalveolate clade was never recovered, but topologically, alveolates were positioned as a sister clade to the heterokonts in analyses of actin, EF-1 alpha, HSP70 and HSP90. For comparison, the support for four other more widely accepted groupings is also shown in Table 2 (green algae + streptophytes, animals + fungi, euglenozoans + heteroloboseans and diplomonads + parabasalia). Most of these groups are not overwhelmingly supported by individual phylogenies either and instead are supported erratically by different genes.

Six-gene and four-gene concatenated phylogenies

Few supported features of any of the individual phylogenies contradicted results from another gene and four of these genes have been concatenated in previous studies (Baldauf *et al.*, 2000). Accordingly, datasets of four- and six-gene concatenations were analysed to compare with the results of the individual phylogenies (see Table 1 for a complete list of taxa used in these concatenations). A four-gene concatenation consisting of actin, alpha-tubulin, beta-tubulin and HSP90 (AABH) was possible for 50 taxa representing the six subgroups of chromalveolates and the five eukaryotic supergroups. In many respects, the phylogeny inferred from these data (Fig. 2) is similar to that recovered for HSP90. There were, however, some notable differences (e.g. the monophyly of the discicristates: euglenozoans + heteroloboseans) and bootstrap support for most nodes was significantly higher (Table 2). Support for the various groupings of chromalveolate taxa was also similar to the results of HSP90 analyses. Alveolates and heterokonts were grouped with moderate support, whereas a cryptomonad

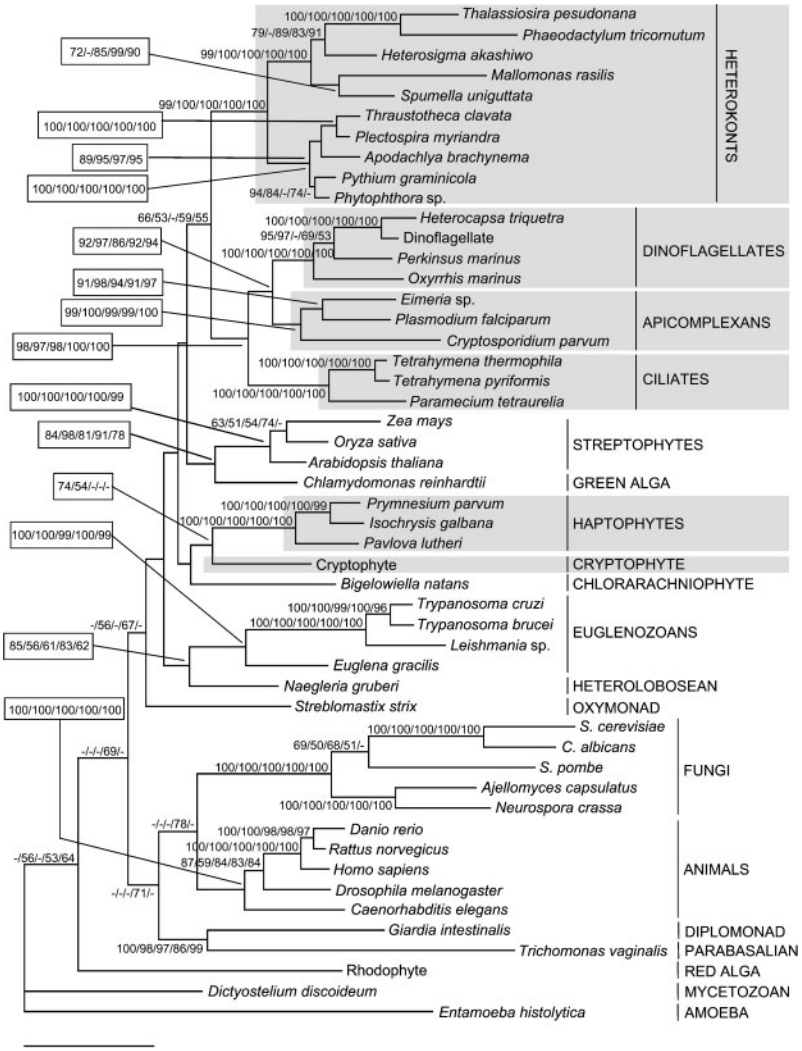

Fig. 2. Protein maximum-likelihood phylogeny (ProML) based on a concatenation of sequences for four nuclear-cytosolic protein-coding genes: actin, alpha-tubulin, beta-tubulin and HSP90. Bootstrap values and labelling formats are as for Fig. 1. The terms 'Cryptomonad', 'Dinoflagellate' and 'Rhodophyte' refer to concatenations of genes from more than one species representing these taxonomic groups (see Table 1). Bar, 0·1 changes per site. Abbreviations: *C. albicans*, *Candida albicans*; *S. cerevisiae*, *Saccharomyces cerevisiae*; *S. pombe*, *Schizosaccharomyces pombe*.

and haptophyte clade received weak support, a topology also recovered by Stechmann & Cavalier-Smith (2003). The individual chromalveolate groups all received strong support, with the exception of the cryptomonads, which were represented by a single set of concatenated genes.

Adding data for EF-1 alpha and HSP70 to the concatenated dataset for analysis yielded a tree (Fig. 3) with topology that was essentially identical to that recovered by analysis of the four-gene dataset (Fig. 2). Bootstrap support was also virtually identical to the four-gene tree with the following qualitative differences: a slight decrease in support for monophyly of the alveolates, a slight increase in support for an alveolate/heterokont clade and increased support for monophyly of the opisthokonts (animals + fungi).

Conclusions: chromalveolates within a global eukaryotic phylogeny

The impetus behind the chromalveolate theory was to limit the number of endosymbiotic events in plastid evolution by proposing that the chromalveolate plastids all share a common origin (Cavalier-Smith, 1999). Indeed, molecular investigations based on plastid-encoded genes and plastid-targeted proteins have provided overlapping results that strongly support this hypothesis (Fast *et al.*, 2001; Harper & Keeling, 2003; Yoon *et al.*, 2002b). There is now growing support from individual and multiple gene phylogenies based on nucleus-encoded cytosolic proteins for the monophyly of an alveolate/heterokont clade. Moreover, the weak relationship observed between cryptomonads and haptophytes (never seen in nuclear gene trees before) is also consistent with the chromalveolate hypothesis. Considering these data within the context of the original cellular characteristics that prompted the recognition of the chromalveolates suggests that these groups are all related and that, by extension, their plastids probably originated via a single secondary endosymbiotic event. Though there is presently no evidence from the phylogeny or cell biology of chromalveolates to suggest that they are not a monophyletic assemblage, it is also clear that more evidence from the nuclear-host lineage is necessary.

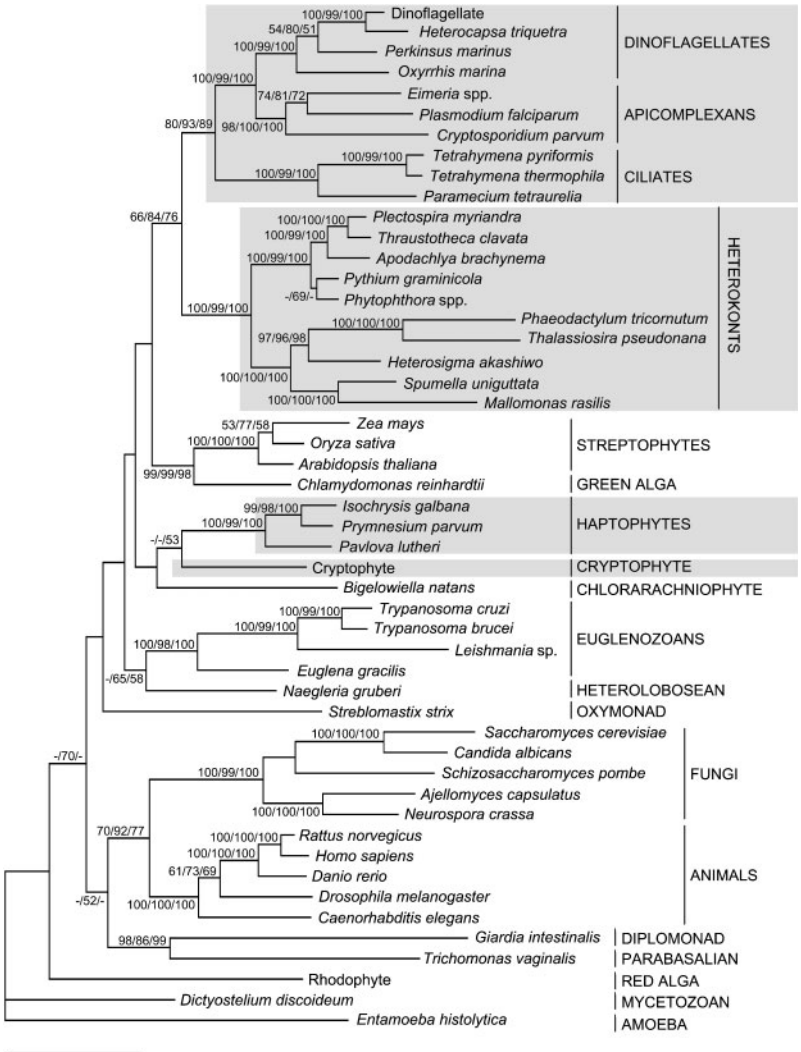

Fig. 3. Protein maximum-likelihood phylogeny (ProML) based on a concatenation of sequences for six nuclear-cytosolic protein-coding genes: actin, alpha-tubulin, beta-tubulin, EF-1 alpha, HSP70 and HSP90. Bootstrap values are shown for nodes that received support over 50 % and are (left to right) ProML, PhyML and parsimony (dashes represent support lower than 50 %). Labelling formats are as for Fig. 1. The terms 'Cryptomonad', 'Dinoflagellate' and 'Rhodophyte' refer to concatenations of genes from more than one species representing these taxonomic groups (see Table 1). Bar, 0·1 changes per site.

In general, uniting distant relatives (for, if the chromalveolates are a monophyletic group, they are an ancient group and distantly related) is a difficult problem that requires considerably more data than are currently available. A global phylogeny of eukaryotes based on nuclear-cytosolic proteins has been examined with four genes and 61 taxa (Baldauf *et al.*, 2000), providing reasonable resolution. However, adding more taxa (six lineages that had not been previously represented were analysed here – cercozoans, cryptomonads, dinoflagellates, oxymonads, *Entamoeba* and haptophytes) has undermined overall support for the tree considerably. Other multiple gene analyses of eukaryotic phylogeny (e.g. Bapteste *et al.*, 2002) have used many more genes (100) but fewer taxa (30) and were addressing important, but relatively narrow questions. A global phylogeny representing even an approximation of eukaryotic diversity remains to be achieved and will likely require a mixture of these approaches when more taxa have been sampled sufficiently to include a number of additional genes for analysis.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Canadian Institutes for Health Research (MOP-42517). *Isochrysis* EST sequencing was supported by Genome Canada/Genome Atlantic. *Bigelowiella* EST sequencing was supported by the Natural Sciences and Engineering Research Council of Canada. P. J. K. is a scholar of the CIAR and New Investigator of CIHR and MSFHR. We thank Brian Leander for *Spumella* clones, and Michelle McEwan and Brian Leander for critical reading of the manuscript.

REFERENCES

Archibald, J. M. & Keeling, P. J. (2002). Recycled plastids: a green movement in eukaryotic evolution. *Trends Genet* **18**, 577–584.

Archibald, J. M., Longet, D., Pawlowski, J. & Keeling, P. J. (2002). A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. *Mol Biol Evol* **20**, 62–66.

Baldauf, S. L. & Palmer, J. D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci U S A* **90**, 11558–11562.

Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977.

Bapteste, E., Brinkmann, H., Lee, J. A. & 8 other authors (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc Natl Acad Sci U S A* **99**, 1414–1419.

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.

Bhattacharya, D. & Weber, K. (1997). The actin gene of the glaucocystophyte *Cyanophora paradoxa*: analysis of the coding region and introns, and an actin phylogeny of eukaryotes. *Curr Genet* **31**, 439–446.

Blanchard, J. L. & Hicks, J. S. (1999). The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. *J Eukaryot Microbiol* **46**, 367–375.

Brown, J. R. & Doolittle, W. F. (1995). Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc Natl Acad Sci U S A* **92**, 2441–2445.

Bruno, W. J., Socci, N. D. & Halpern, A. L. (2000). Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol* **17**, 189–197.

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. (2004). Chromalveolate diversity and cell megaevolution: interplay of membranes, genomes and cytoskeleton. In *Organelles, Genomes and Eukaryotic Evolution*, pp. 71–103. Edited by R. P. Hirt & D. Horner. London: Taylor and Francis.

Cavalier-Smith, T. & Chao, E. E. (1996). Molecular phylogeny of the free-living archezoan *Trepomonas agilis* and the nature of the first eukaryote. *J Mol Evol* **43**, 551–562.

Chadefaud, M. (1950). Les cellules nageuses des Algues dans l'embranchement des Chromophycées. *C R Acad Sci Paris* **231**, 788–790 (in French).

Christensen, T. (1989). The Chromophyta, past and present. In *The Chromophyte Algae: Problems and Perspectives*. Systematics Association Special Vol. no. 38, pp. 1–12. Edited by J. C. Green, B. S. C. Leadbeater & W. L. Diver. Oxford: Clarendon Press.

Dacks, J. B., Marinets, B. A., Doolittle, W. F., Cavalier-Smith, T. & Logsdon, J. M., Jr (2002). Analyses of RNA polymerase II genes from free-living protists: phylogeny, long branch attraction, and the eukaryotic big bang. *Mol Biol Evol* **19**, 830–840.

Embley, T. M. & Hirt, R. P. (1998). Early branching eukaryotes? *Curr Opin Genet Dev* **8**, 624–629.

Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* **18**, 418–426.

Fast, N. M., Xue, L., Bingham, S. & Keeling, P. J. (2002). Re-examining alveolate evolution using multiple protein molecular phylogenies. *J Eukaryot Microbiol* **49**, 30–37.

Felsenstein, J. (1993). PHYLIP (phylogenetic inference package). Department of Genetics, University of Washington, Seattle, WA, USA.

Funes, S., Davidson, E., Reyes-Prieto, A., Magallón, S., Herion, P., King, M. P. & Gonzalez-Halphen, D. (2002). A green algal apicoplast ancestor. *Science* **298**, 2155.

Gribaldo, S. & Philippe, H. (2002). Ancient phylogenetic relationships. *Theor Popul Biol* **61**, 391–408.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

Harper, J. T. & Keeling, P. J. (2003). Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol* **20**, 1730–1735.

Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K. & Hasegawa, M. (1994). Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol Biol Evol* **11**, 65–71.

Hashimoto, T., Nakamura, Y., Kamaishi, T., Nakamura, F., Adachi, J., Okamoto, K. & Hasegawa, M. (1995). Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2. *Mol Biol Evol* **12**, 782–793.

Hirt, R. P., Logsdon, J. M., Jr, Healy, B., Dorey, M. W., Doolittle, W. F. & Embley, T. M. (1999). Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc Natl Acad Sci U S A* **96**, 580–585.

Kamaishi, T., Hashimoto, T., Nakamura, Y., Nakamura, F., Murata, S., Okada, N., Okamoto, K., Shimzu, M. & Hasegawa, M. (1996). Protein phylogeny of translation elongation factor EF-1α suggests microsporidians are extremely ancient eukaryotes. *J Mol Evol* **42**, 257–263.

Keeling, P. J. (2001). Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? *Mol Biol Evol* **18**, 1551–1557.

Keeling, P. J. (2004). Diversity and evolutionary history of plastids and their hosts. *Am J Bot* (in press).

Keeling, P. J. & Doolittle, W. F. (1996). Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol Biol Evol* **13**, 1297–1305.

Keeling, P. J. & Inagaki, Y. (2004). A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor-1α. *Proc Natl Acad Sci U S A* **101**, 15380–15385.

Keeling, P. J. & Leander, B. S. (2003). Characterisation of a non-canonical genetic code in the oxymonad *Streblomastix strix*. *J Mol Biol* **326**, 1337–1349.

Keeling, P. J. & Palmer, J. D. (2001). Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. *Proc Natl Acad Sci U S A* **98**, 10745–10750.

Köhler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J. M., Palmer, J. D. & Roos, D. S. (1997). A plastid of probable green algal origin in Apicomplexan parasites. *Science* **275**, 1485–1489.

Kumar, S. & Rzhetsky, A. (1996). Evolutionary relationships of eukaryotic kingdoms. *J Mol Evol* **42**, 183–193.

Kuvardina, O. N., Leander, B. S., Aleshin, V. V., Myl'nikov, A. P., Keeling, P. J. & Simdyanov, T. G. (2002). The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to apicomplexans. *J Eukaryot Microbiol* **49**, 498–504.

Leander, B. S. & Keeling, P. J. (2004). Early evolution of dinoflagellates and apicomplexans inferred from HSP90 and actin phylogeny. *J Phycol* **40**, 341–350.

McFadden, G. I. & Waller, R. F. (1997). Plastids in parasites of humans. *Bioessays* **19**, 1033–1040.

Morin, L. (2000). Long branch attraction effects and the status of 'Basal Eukaryotes': Phylogeny and structural analysis of the ribosomal RNA gene cluster of the free-living diplomonad *Trepomonas agilis*. *J Eukaryot Microbiol* **47**, 167–177.

Philippe, H. (2000). Opinion: long branch attraction and protist phylogeny. *Protist* **151**, 307–316.

Saldarriaga, J. F., McEwan, M. L., Fast, N. M., Taylor, F. J. R. & Keeling, P. J. (2003). Multiple protein phylogenies show that *Oxyrrhis marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. *Int J Syst Evol Microbiol* **53**, 355–365.

Simpson, A. G., Radek, R., Dacks, J. B. & O'Kelly, C. J. (2002a). How oxymonads lost their groove: an ultrastructural comparison of *Monocercomonoides* and excavate taxa. *J Eukaryot Microbiol* **49**, 239–248.

Simpson, A. G., Roger, A. J., Silberman, J. D., Leipe, D. D., Edgcomb, V. P., Jermiin, L. S., Patterson, D. J. & Sogin, M. L. (2002b). Evolutionary history of 'early-diverging' eukaryotes: the excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Mol Biol Evol* **19**, 1782–1791.

Sogin, M. L. & Silberman, J. D. (1998). Evolution of the protists and protistan parasites from the perspective of molecular systematics. *Int J Parasitol* **28**, 11–20.

Stechmann, A. & Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. *Science* **297**, 89–91.

Stechmann, A. & Cavalier-Smith, T. (2003). Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. *J Mol Evol* **57**, 408–419.

Strimmer, K. & von Haeseler, A. (1996). Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol* **13**, 964–969.

Swofford, D. L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (and other methods). Sunderland, MA: Sinauer Associates.

Van de Peer, Y. & De Wachter, R. (1997). Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site variation in 18S rRNA. *J Mol Evol* **45**, 619–630.

Van de Peer, Y., Baldauf, S. L., Doolittle, W. F. & Meyer, A. (2000a). An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. *J Mol Evol* **51**, 565–576.

Van de Peer, Y., Ben Ali, A. & Meyer, A. (2000b). Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene* **246**, 1–8.

Williamson, D. H., Gardner, M. J., Preiser, P., Moore, D. J., Rangachari, K. & Wilson, R. J. (1994). The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. *Mol Gen Genet* **243**, 249–252.

Yoon, H. S., Hackett, J. D. & Bhattacharya, D. (2002a). A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc Natl Acad Sci U S A* **99**, 11724–11729.

Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. (2002b). The single, ancient origin of chromist plastids. *Proc Natl Acad Sci U S A* **99**, 15507–15512.

Zhang, Z., Green, B. R. & Cavalier-Smith, T. (2000). Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids. *J Mol Evol* **51**, 26–40.