J. Eukaryot. Microbiol., 56(4), 2009 pp. 348–356

DOI: 10.1111/j.1550-7408.2009.00408.x

The RAB Family GTPase Rab1A from *Plasmodium falciparum* Defines a Unique Paralog Shared by Chromalveolates and Rhizaria

MAREK ELIAS,[a] NICOLA J. PATRON[b] and PATRICK J. KEELING[b]

[a]*Department of Botany, Faculty of Science, Charles University in Prague, Benatska 2, 128 01 Prague 2, Czech Republic, and*
[b]*Department of Botany, Canadian Institute for Advanced Research, University of British Columbia, Vancouver, BC, Canada V6T 1Z4*

ABSTRACT. The RAB GTPases, which are involved in regulation of endomembrane trafficking, exhibit a complex but incompletely understood evolutionary history. We elucidated the evolution of the RAB1 subfamily ancestrally implicated in the endoplasmic reticulum-to-Golgi traffic. We found that RAB1 paralogs have been generated over the course of eukaryotic evolution, with some duplications coinciding with the advent of major eukaryotic lineages (e.g. Metazoa, haptophytes). We also identified a unique, derived RAB1 paralog, orthologous to the *Plasmodium* Rab1A, that occurs in stramenopiles, alveolates, and Rhizaria, represented by the chlorarachniophyte *Gymnochlora stellata*. This finding is consistent with the recently documented existence of a major eukaryotic clade (''SAR'') comprising these three lineages. We further found a Rab1A-like protein in the cryptophyte *Guillardia theta*, but it exhibits unusual features among RAB proteins: absence of a C-terminal prenylation motif and an N-terminal extension with two MSP domains; and its phylogenetic relationships could not be established convincingly due to its divergent nature. Our results nevertheless point to a unique membrane trafficking pathway shared by at least some lineages of chromalveolates and Rhizaria, an insight that has implications towards interpreting the early evolution of eukaryotes and the endomembrane system.

Key Words. Chlorarachniophyta, Cryptophyta, eukaryotic cell, evolution, membrane trafficking, phylogeny.

RECONSTRUCTION of the deepest evolutionary history of eukaryotes comprises at least two interrelated tasks: (1) defining the topology of the eukaryotic phylogenetic tree; and (2) identification of evolutionary novelties or apomorphies associated with the origin of individual monophyletic lineages. The first task has proven challenging, but the recent applications of large sequence datasets analysed with improved methods of phylogenetic inference suggest that a robust scheme of eukaryotic phylogeny may soon be available (Burki et al. 2007; Hackett et al. 2007; Hampl et al. 2009; Keeling et al. 2005; Patron, Inagaki, and Keeling 2007; Rodriguez-Ezpeleta et al. 2007). One of the most surprising results of these analyses is the emergence of a novel strongly supported clade, informally dubbed ''SAR'', comprising the stramenopiles, alveolates, and Rhizaria (Burki et al. 2007; Burki, Shalchian-Tabrizi, and Pawlowski 2008; Hackett et al. 2007; Hampl et al. 2009; Rodriguez-Ezpeleta et al. 2007). There is no specific cellular or molecular character known to be shared by these three lineages, and these groups are partially overlapping with another supergroup hypothesis, the chromalveolates.

The chromalveolate hypothesis postulates the monophyletic supergroup Chromalveolata uniting stramenopiles and alveolates with cryptophytes and haptophytes rather than with Rhizaria (Cavalier-Smith 1999). This hypothesis is based on the presence in most chromalveolate lineages of a red alga-derived secondary plastid that is postulated to have been present already in the common ancestor of the group (and secondarily lost in some representatives, e.g. ciliates). Rhizaria lack this plastid, so the potential close relationship of Rhizaria with stramenopiles and alveolates, if confirmed, would necessitate major revisions of our understanding of plastid evolution.

Identification of evolutionary novelties specific for many monophyletic lineages of eukaryotes has also been challenging, in particular hampered by the paucity of genomic data for many important groups (e.g. Rhizaria, cryptophytes, glaucophytes, jakobids, apusomonads). Several candidate molecular synapomorphies have been suggested for some major groupings (e.g. Elias 2008; Rice and Palmer 2006; Richards and Cavalier-Smith 2005), including the chromalveolates (Harper and Keeling 2003; Patron, Rogers, and Keeling 2004). However, broad sampling of eukaryotic diversity is a prerequisite for defining the actual phylogenetic distribution of any such candidate synapomorphies. A prominent example that has not withstood database expansion is the class II myosin subfamily, claimed by Richards and Cavalier-Smith (2005) to be restricted to the eukaryotic supergroups Opisthokonta and Amoebozoa and thus supporting their union into a grouping called the ''unikonts''. However, it has now been found in a ''bikont'', the heterolobosean *Naegleria gruberi* (Odronitz and Kollmar 2007). Caution is therefore necessary when interpreting phyletic patterns of presence/absence of particular genes inferred from a sparse sample of eukaryotic genomes.

One aspect of the eukaryotic cell that might have been shaped by many lineage-specific evolutionary innovations is the membrane trafficking network. Indeed, the endomembrane system has proven to be very plastic among different eukaryotic groups (Becker and Melkonian 1996; Dacks and Field 2004; Field, Gabernet-Castello, and Dacks 2007). It is not surprising that comparative analyses of proteins implicated in endomembrane dynamics, although still limited in their scope, have suggested possible molecular idiosyncrasies specific for some eukaryotic lineages (e.g. opisthokonts Elias 2008; Field et al. 2007; or alveolates Gould et al. 2008).

The RAB family of monomeric GTPases includes central components of the vesicle transport machinery. The family is extremely complex, with number of distinct paralogs ranging to over 60 in human (Pereira-Leal and Seabra 2001) or up to 300 in *Trichomonas vaginalis* (Carlton et al. 2007). The RAB proteins regulate various steps in the trafficking network, including the budding of vesicles from donor membranes, vesicle transport, and fusion with the target membrane (see reviews by Grosshans, Ortiz, and Novick 2006; Zerial and McBride 2001). The evolutionary relationships within the RAB family and functional specialisation of individual paralogs are not completely understood, but several ancient RAB subclasses, each with a specialised function, are probably conserved across a wide range of eukaryotes (Pereira-Leal and Seabra 2001).

One recent analysis of the evolutionary history of the RAB5 subfamily showed a number of independent gene duplications leading to RAB5 paralogs restricted to only particular eukaryotic subgroups (Dacks, Poon, and Field 2008). Here, we describe a detailed analysis of the RAB1 subfamily. The RAB1 subfamily as understood here corresponds to the ''RAB functional group I'' defined by Pereira-Leal and Seabra (2001) as comprising the Rab1 and Rab35 paralogs from Metazoa, Ypt1 genes from yeasts, and

Corresponding Author: M. Elias, Department of Botany, Faculty of Science, Charles University in Prague, Benatska 2, 128 01 Prague 2, Czech Republic—Telephone number: 0042 02219 51648; FAX number: 0042 02219 51645; e-mail: melias@natur.cuni.cz

the RabD group from *Arabidopsis thaliana*. Despite the different naming conventions, all these RABs appear to represent one orthology group, characterised by a conserved common function in the anterograde transport between the endoplasmic reticulum (ER) and the Golgi complex (Batoko et al. 2000; Dhir, Goulding, and Field 2004; Grosshans et al. 2006; Morsomme and Riezman 2002). The RAB1 orthologs have since been found in an array of diverse eukaryotes (Ackers, Dhir, and Field 2005; Eisen et al. 2006; Lal et al. 2005; Langford et al. 2002; Montsant et al. 2007; Quevillon et al. 2003; Saito-Nakano et al. 2005; Weeks, Gaudet, and Insall 2005). However, some taxa are known to possess multiple RAB1 paralogs whose evolutionary origin has not yet been addressed. Based on a broad phylogenetic analysis of the RAB1 subfamily, we find a unique RAB1 paralog typified by Rab1A from *Plasmodium falciparum*. The paralog has an intriguing phylogenetic distribution that further supports close relationship of Rhizaria to stramenopiles and alveolates.

MATERIALS AND METHODS

Assembling the sequence dataset. The sequences analysed in this study came from our extensive database of RAB sequences (ME., unpubl. data) derived from completed and ongoing genome projects and expressed sequence tag (EST) surveys including an ongoing EST survey of the chlorarachniophyte *Gymnochlora stellata* (for the list of sequences and the source databases see the Supporting Information Table S1). The full-length coding sequence of the Rab1A gene from *Guillardia theta* (EU069499) was assembled using EST sequences from our own survey and from EST sequences kindly made available by the DOE Joint Genome Institute. Reads of ESTs or whole-genome shotgun (WGS) from the same species likely representing the same gene (i.e. sharing $\geq 96\%$ identity) were clustered using the CAP3 Sequence Assembly Program (Huang and Madan 1999; pbil.univ-lyon1.fr/cap3.php) and the assemblies were manually refined whenever necessary; in several cases the sequences were corrected after inspecting the original sequencing chromatograms available from the Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/) or TBestDB (http://amoebidia.bcm.umontreal.ca/pepdb). Clones corresponding to several incomplete ESTs were fully sequenced; the sequences were deposited at GenBank under the accession numbers EU069499–EU069504. The exon–intron structure of previously un-annotated genes was determined relying on the information from EST/cDNA sequences (if available) and conservation of intron positioning and encoded protein sequences. If evidence from homology or transcript sequences argued for the less common GC-AG intron borders, they were allowed in addition to the canonical GT-AG borders. We rechecked our predictions as well as predictions retrieved from databases after inspecting protein multiple alignments and corrected the gene models before producing final alignments used for phylogenetic analyses. New and corrected protein predictions and sequences inferred from assembled ESTs and WGS reads are available upon request from M.E.

Sequence analyses. To obtain a starting set of potential RAB1 orthologs, we compared all RAB sequences to all RAB GTPases from *Homo sapiens* using BLASTP (BLOSUM62 scoring matrix, low-complexity filter off; Altschul et al. 1997). All sequences giving known RAB1 subfamily members (i.e. Rab1a [NP_004152.1], Rab1b [NP_112243.1], or Rab35 [NP_006852.1]) as best hits were retrieved and aligned together with representatives of several other conserved RAB subfamilies (RAB8, RAB18, RAB2, RAB11, RAB5) using Clustal X (Thompson et al. 1997). The alignment was further manually edited in GeneDoc (Nicholas and Nicholas 1997; http://www.psc.edu/biomed/genedoc/) following experimentally determined secondary structures of RAB1 subfamily proteins. Poorly conserved regions were removed yielding 160 positions in the final alignment used in phylogenetic analyses.

The original dataset contained 377 sequences including many potentially RAB1-unrelated and/or highly divergent ones, which might negatively influence phylogenetic analyses. To identify such sequences and to reduce the extent of the dataset so that it can be analysed with computationally demanding methods, we constructed a working BioNJ (Gascuel 1997) tree based on distances calculated with the JTT matrix using *PROTDIST* (PHYLIP 3.6 package; Felsenstein 2004). We placed the root of the tree arbitrarily on the branch leading to the RAB5 subfamily, which is distantly related to the RAB1 subfamily (e.g. Pereira-Leal and Seabra 2001). We then defined RAB1 subfamily sequences as those belonging to the most inclusive clade including the human Rab1a but excluding the RAB8 subfamily sequences. In addition, we identified and removed divergent or long-branch sequences, which were defined as those with a cumulative branch length from the root to the leaf longer than a threshold arbitrarily selected as the leaf of the *P. falciparum* Rab1A sequence (CAD51503.1; Quevillon et al. 2003). Finally, we excluded sequences from several metazoan and fungal species having other close relatives in the dataset and highly similar paralogs from the same species. This selection procedure yielded a final alignment of 160 sequences that were subjected to a maximum likelihood (ML) phylogenetic analysis implemented in PhyML-aLRT v1.1 (Anisimova and Gascuel 2006; Guindon and Gascuel 2003) using the WAG+G+I substitution model with eight rate categories and all parameters estimated from the data. Replicates for bootstrap analysis were produced with the SEQBOOT program from the PHYLIP 3.6 package and bootstrap trees were inferred with the ML method with PhyML-aLRT using the WAG+G+I substitution model with parameters as estimated for the original dataset. Bootstrap support for the tree topology was also assessed with the rapid ML bootstrapping algorithm (-x option, PROTGAMMAIWAG substitution model) implemented in RAxML 7.0.3 (Stamatakis, Hoover, and Rougemont 2008). Consensus trees with bootstrap values were obtained using the program CONSENSE from the PHYLIP 3.6 package. Trees with extra sequences added to the core dataset (see ''Results'' and ''Discussion'') were inferred with the same procedure as described above.

RESULTS

A complex evolutionary history of the RAB1 subfamily. We noted by reciprocal BLASTP searches a possible relationship between certain RAB1-like sequences from stramenopiles and alveolates, including the Rab1A gene from *P. falciparum*. This suggested there might exist a specific RAB subfamily that is unique to alveolates and stramenopiles, a possibility that we sought to test using phylogenetic analyses. Figure 1 shows the result of a ML analysis of a broad set of RAB1-related sequences. Even with the exclusion of the most divergent sequences (see ''Materials and Methods''), the tree still suffers from extensive ranges of evolutionary rates. Because RABs are small proteins that provide a limited number of positions for phylogenetic inference, only 160 positions in our case, there is low bootstrap support for most nodes. Nevertheless, the tree topology reflects a tendency of sequences from most eukaryotic groups to branch together.

Hence, all fungal RAB1 sequences, except the more divergent sequence from the yeast *Saccharomyces cerevisiae*, form a single clade, although without bootstrap support (Fig. 1). Metazoan sequences except the derived Rab35 paralog (see below) and a handful of other sequences (some from species known to harbour rapidly diverging sequences such as nematodes or *Schistosoma*) branch together with genes from the related unicellular opisthokonts (i.e. *Monosiga*, *Capsaspora*, *Amoebidium*), though again

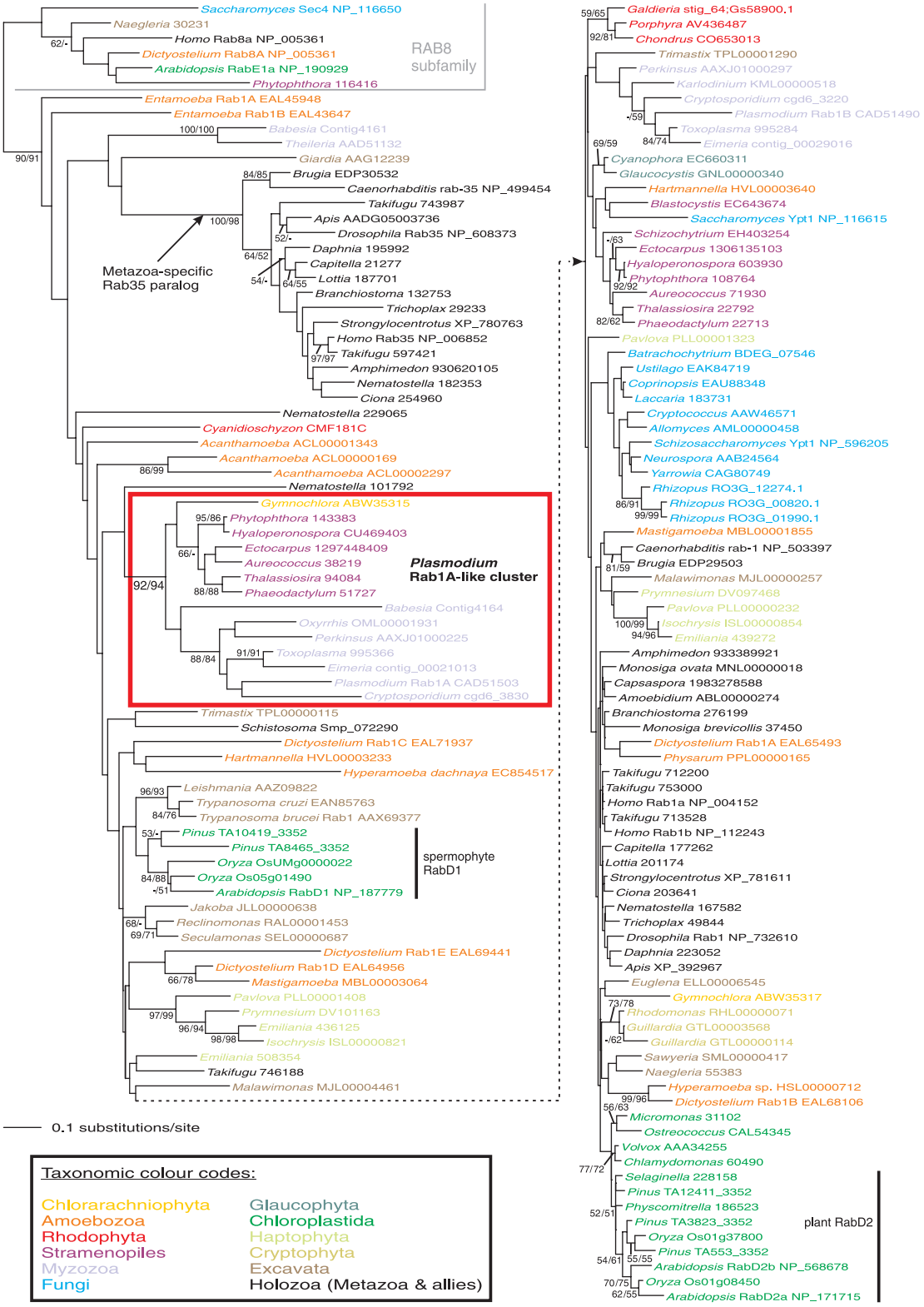

Fig. **1.** Maximum likelihood phylogenetic analysis of the RAB1 subfamily of RAB GTPases. The tree was inferred from an alignment of 160 amino acid positions of RAB1 sequences using PhyML-aLRT 1.1 (WAG+G+I, eight rate categories, loglk = − 13914.99831, $\alpha = 0.642$, proportion of invariable sites = 0.017). Bootstrap values based on 100 replicates were calculated with the ML method employing PhyML-aLRT 1.1 with the WAG+G+I model with parameters as estimated for the original tree and with the rapid bootstrap algorithm as implemented in RAxML 7.0.3. Note that only values higher than 50% are shown. Taxonomic affiliation of sequences is indicated with different colours. Only generic names are shown for most taxa, full species names and source databases are provided in Table S1. For genes that have been previously given specific names, these are inserted between the taxon names and accession numbers. The root of the tree is placed between the RAB1 subfamily and representatives of the closely related RAB8 subfamily.

without bootstrap support and with two amoebozoan sequences nested within the group. The Rab35 paralog forms a strongly supported cluster comprising sequences only from Metazoa. Several apparently derived signatures are conserved in all Rab35 sequences (e.g. a deletion of one residue in the loop connecting the helix2 and strand5; data not shown), and the long stem of the Rab35 clade indicates an accelerated evolution before metazoan radiation.

RAB1 sequences from the green lineage (Chloroplastida, including Embryophyta and several green algae) form two clusters (both without support). One comprises at least one gene per each species examined, whereas the other is restricted only to sequences from spermatophytes (*Arabidopsis*, rice, pine). Three of four rhodophyte RAB1 sequences analysed in our study cluster weakly together, but that from *Cyanidioschyzon* is displaced toward the base of the tree occupied mostly by rapidly evolving sequences (Fig. 1). Two sequences from *Cyanophora* and *Glaucocystis*, species representing the remaining lineage of the supergroup Archaeplastida, also form a weakly supported cluster (Fig. 1).

Sequences from other supergroups are more scattered. Amoebozoan RAB1-related sequences are found throughout the tree (Fig. 1). In all species with the genome sequences or extensive EST collections available, multiple RAB1 paralogs occur. For example, five RABs (Rab1A–Rab1E) from *Dictyostelium discoideum* have been assigned to the Rab1 group by Weeks et al. (2005) and other genes that have been excluded from the present analysis because of their divergence are perhaps also related to the RAB1 subfamily (not shown). Some paralogs appear to be shared by distantly related amoebozoans (e.g. the Rab1B gene from *Dictyostelium* and *Hyperamoeba* or *Dictyostelium* Rab1D and one of the paralogs from *Mastigamoeba*).

Sequences from trypanosomatids, *Euglena*, jakobids, heteroloboseans, *Malawimonas* (two paralogs), *Trimastix* (two paralogs), and *Giardia* each form a unique group that do not display any supported relationship to each other or to any other sequences (Fig. 1). Some other RAB1 sequences from these taxa and all RAB1 orthologs identified in *Trichomonas* by Lal et al. (2005) were discarded from the tree analysis due to their high rate of substitution.

Multiple RAB1 paralogs were uncovered also in haptophytes. Majority of them fall into two strongly supported clusters, each comprising representatives from both major haptophyte classes (i.e. Pavlovophyceae and Prymnesiophyceae, Fig. 1), indicating that both paralogs were present in the common ancestor of extant haptophytes.

A RAB1 paralog shared by stramenopiles, alveolates, and Rhizaria. The most important result of our analysis is the observation that RAB1 sequences from stramenopiles, Myzozoa (a subgroup of alveolates comprising apicomplexans, perkinsids, and dinoflagellates), and the chlorarachniophyte *G. stellata* fall into two broad categories. One category represents RABs that are similar to non-divergent RAB1 sequences of other eukaryotes and cluster with them in the tree in a manner mirroring the species groupings (with the exception of a sequence from the parasitic stramenopile *Blastocystis* and two highly divergent RABs from piroplasmids *Babesia* and *Theileria*; Fig. 1). The second category corresponds to a strongly supported clade closer to the base of the RAB1 subfamily comprising one sequence from each completely sequenced stramenopile genome, one sequence from majority of completely sequenced apicomplexan genomes (except *Theileria*, see below), one gene from the genome of *Perkinsus marinus*, an EST from the dinoflagellate *Oxyrrhis marina*, and an EST from *Gymnochlora* (Fig. 1). Because this clade includes a gene from *P. falciparum* previously designated as Rab1A (Quevillon et al. 2003), we refer to it as ''*Plasmodium* Rab1A-like cluster''.

Inspecting the RAB1 alignment showed sequence signatures unique *Plasmodium* Rab1A-like proteins: positions 49, 50, 79, and 85 (as in the human Rab1a) are typically occupied by K, I, S, and H residues, respectively, but R, F, A, and D residues are found at these positions in all sequences of the Rab1A cluster (Fig. 2).

The presence of Rab1A in most myzozoans raises a question whether it is conserved in alveolates in general. Because for the analysis presented above (Fig. 1) we had omitted a number of sequences forming long branches in a preliminary distance tree, including a sequence from the piroplasmid *Theileria* and sequences from ciliates, we asked whether any of those divergent sequences could be related to the Rab1A cluster. Indeed, when we added the *Theileria* sequence to the dataset, it was robustly resolved as a sister of the Rab1A sequence from another piroplasmid, *Babesia* (Fig. 3a). This sequence also shares three of the four Rab1A signature substitutions (Fig. 2), suggesting that it is indeed orthologous to Rab1A from other myzozoans. Next, we added a group of relatively divergent sequences from the ciliates *Paramecium*, *Tetrahymena*, and *Oxytricha*, which were sistergroup to the Rab1A cluster in the preliminary distance tree (not shown). The resulting tree again shows the ciliate sequences as a divergent sister to the ''core'' Rab1A cluster with weak support (Fig. 3b). Inspecting individual bootstrap trees shows that the low bootstrap values (i.e. below 50%) for the core Rab1A cluster in this tree are because it is often disrupted either by the ciliate Rab1A-like sequences or by the relatively divergent *Babesia* Rab1A sequence branching with the ciliate Rab1A sequences elsewhere in the tree. Most of the ciliate sequences also share the four signature residues with Rab1A, with the exception of a few sequences from *Paramecium* deviating at one or two positions (Fig. 2).

After the analyses reported here were completed, a draft genome assembly from the on-going sequencing project for another chlorarachniophyte became available, *Bigelowiella natans* (see http://www.jgi.doe.gov/sequencing/why/50026.html). The *Bigelowiella* genome encodes a protein highly similar and apparently orthologous to the *Gymnochlora* Rab1A: it contains all four Rab1A signature residues, and branches with the *G. stellata* Rab1A with 100% bootstrap support and within the Rab1A cluster with 98% support in ML analysis (using RAxML: data not shown). These preliminary results thus indicate wider conservation of the Rab1A paralog in the chlorarachniophytes.

A structurally unique Rab1A-like protein in the cryptophyte *Guillardia theta*. Searching for possible relatives of Rab1A among the RAB1 sequences omitted from the core analysis (Fig. 1), we noticed that one sequence from the cryptophyte *G. theta* exhibits higher BLASTP similarity to the stramenopile Rab1A sequences than to any other RABs in our database. It also features three of the four Rab1A signature residues, with the fourth signature position occupied by a W rather than an F residue, which can be considered as a conservative substitution (Fig. 2). When we added this *Guillardia* sequence to the main dataset, we found that it indeed branches as sister to the core Rab1A cluster with weak support (Fig. 3c). As with the ciliate Rab1A sequences, the *Guillardia* sequence either nested within Rab1A or the *Babesia* Rab1A sequence grouped with *Guillardia* elsewhere in the tree in individual bootstrap replicates, weakening the support for Rab1A (data not shown). Overall, the *Guillardia* Rab1A-like sequence appears to be an ortholog of the stramenopile, alveolate, and rhizarian Rab1A. Two more *Gullardia* RAB1 paralogs cluster together with an EST from another cryptophyte, *Rhodomonas salina*, and represent less divergent members of the RAB1 subfamily (Fig. 1).

The vast majority of RABs possess a hypervariable C-terminal tail bearing two (or infrequently one) cysteine residues that are modified by geranylgeranyl moieties anchoring the RAB protein in membranes (Pereira-Leal and Seabra 2001; Pfeffer and Aivazian 2004). The *Guillardia* Rab1A-like sequence, however, lacks a C-terminal cysteine geranylgeranylation motif and instead con-

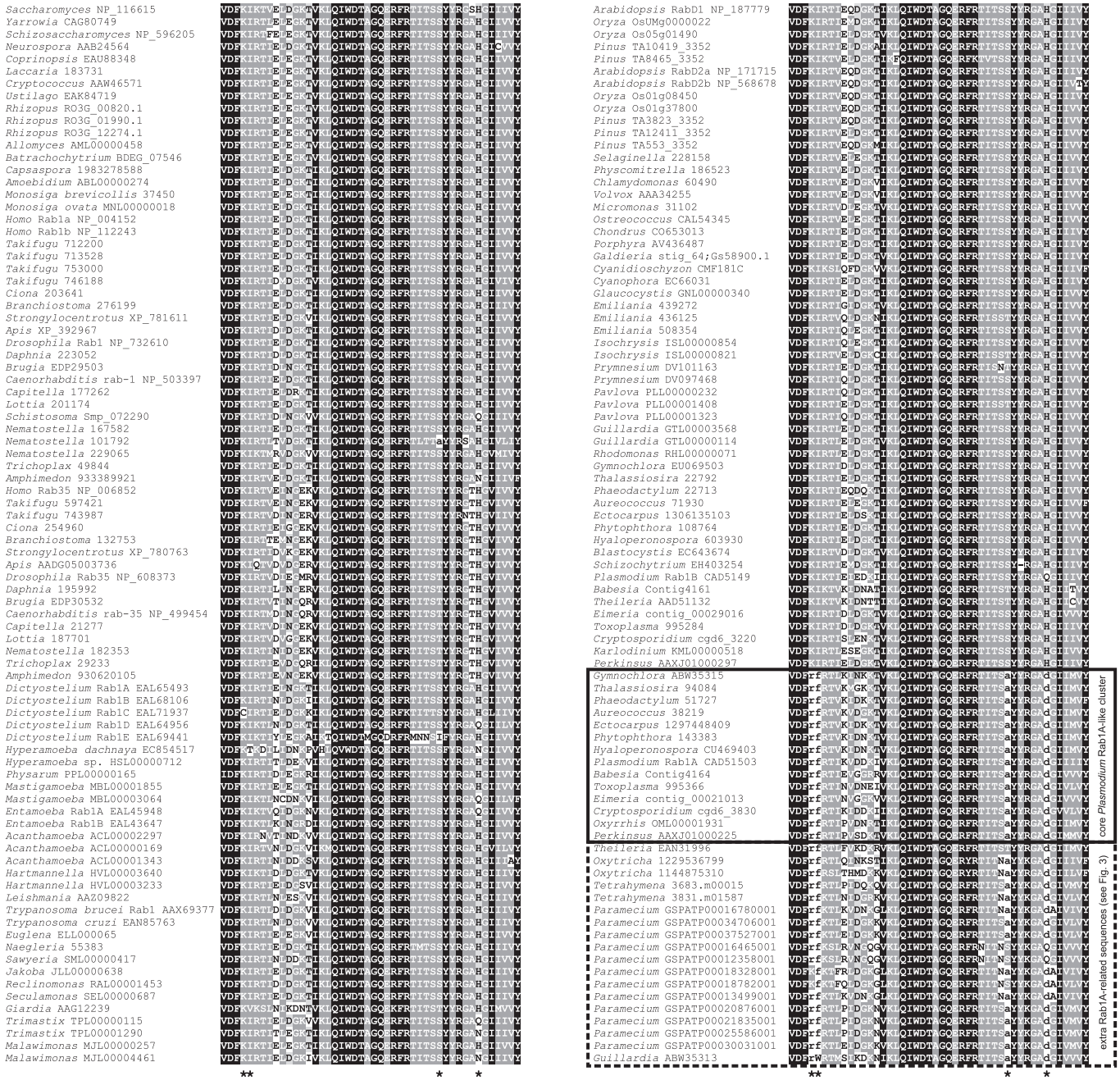

Fig. **2.** Signature residues defining the *Plasmodium* Rab1A-like paralog. The figure shows a part of a multiple alignment of RAB1 subfamily sequences corresponding to the strand2 to strand4 region. The four signature positions for Rab1A are marked with asterisks at the bottom of the alignment; residues at these positions conforming to the Rab1A-specific pattern are displayed in lower case (r, f, a, and d, respectively). Sequences forming the *Plasmodium* Rab1A-like cluster (see Fig. 1) are boxed with a solid line. Additional sequences related to this cluster, based on other analyses (see Fig. 3) are boxed with a dashed line.

tains an N-terminal extension of $\approx$ 310 residues (Fig. 3c), which according to Search Pfam (Finn et al. 2006) contains a tandem of two Major sperm protein (MSP) domains (with *E*-values of 3.9e − 10 and 4.4e − 11, respectively) separated by a short linker (28 residues). The MSP domain is present in a number of proteins including the mammalian ER-associated VAMP-associated protein, in which it mediates recruitment of FFAT motif-containing proteins to the ER membrane (Kaiser et al. 2005). We speculate that cryptophytes have modified their Rab1A protein by replacing geranylgeranyl-mediated membrane association with a novel mechanism involving the interaction of the MSP domains with a membrane protein (possibly containing the FFAT motif).

DISCUSSION

RAB1 evolution in opisthokonts, plants, and amoebozoa. Elucidating the phylogenetic relationships within a family of proteins so short as RABs is difficult, but our analyses do nevertheless allow some interesting insights. For example, the metazoan

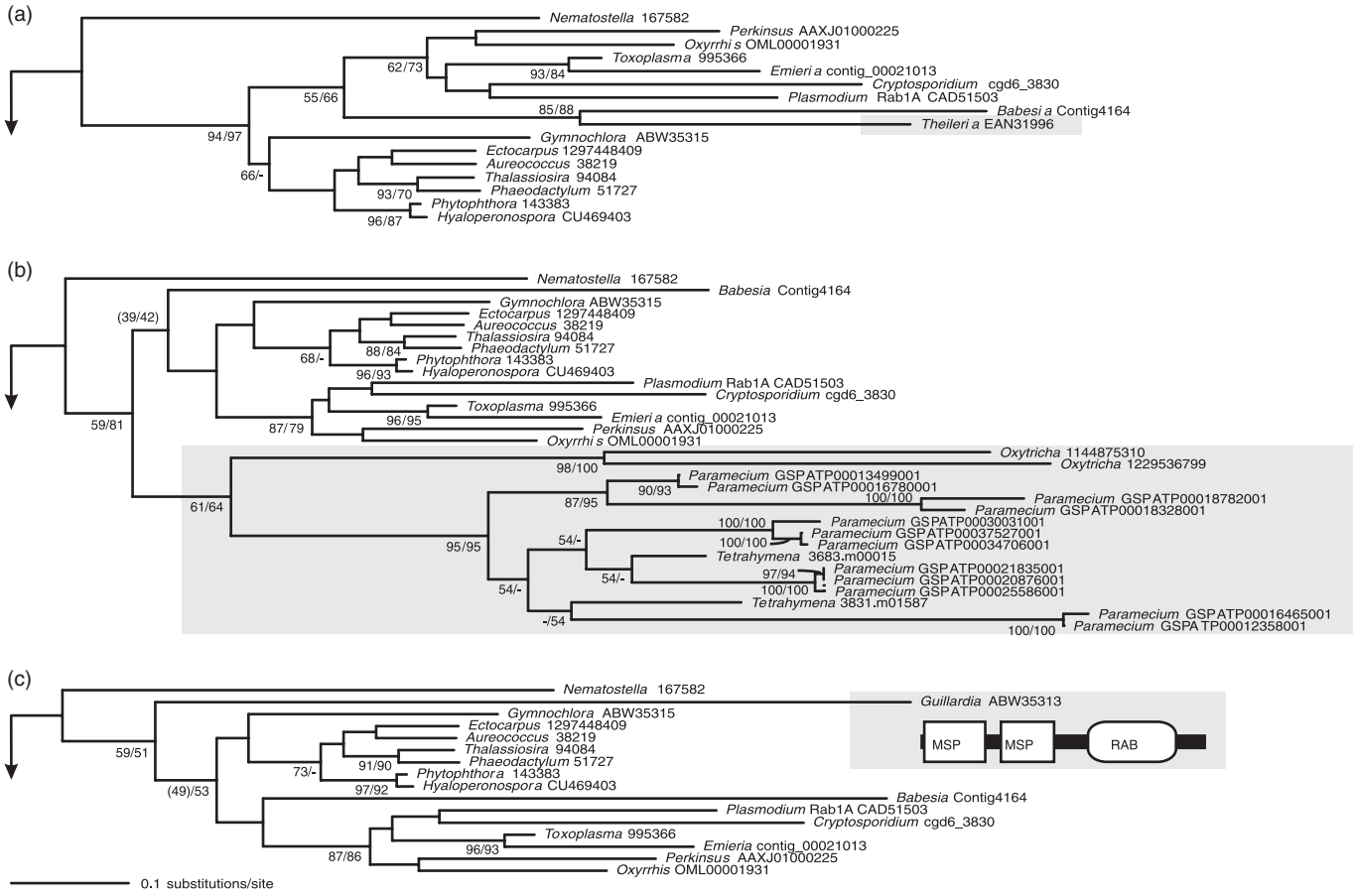

Fig. **3.** Reanalyses of the RAB1 phylogeny with extra sequences added to the core dataset. The analyses were performed with the same approach as used for constructing the first tree (see Fig. 1). Extra sequences added to the core dataset are highlighted by a grey background. The overall topologies of the trees were very similar to that of Fig. 1, hence only parts of the trees around the *Plasmodium* Rab1A-like cluster are shown. The arrows point to the hidden parts of the trees. The domain architecture of the *Guillardia* Rab1A (ABW35313) is schematically shown according to the output from Pfam search (MPS—Major sperm protein domain).

RAB1 subfamily sequences are probably best interpreted as reflecting the existence of two conserved RAB1 paralogs ancestral to multicellular animals, including sponges: the prototypical Rab1 and the derived Rab35. Rab35-related genes are not present in genomes or ESTs from any non-metazoan eukaryotes sampled, including other members of the Holozoa (i.e. the choanoflagellates *Monosiga brevicollis* and *Monosiga ovata*, the ichthyosporean *Amoebidium parasiticum*, and *Capsaspora owczarzaki*) or the closely related Fungi, suggesting that Rab35 arose via duplication and rapid divergence of the ancestral RAB1 gene specifically in the metazoan lineage. The increase in the evolutionary rate of Rab35 may reflect neofunctionalisation, because Rab35 is no longer considered to act in ER-to-Golgi traffic (the ancestral function of the RAB1 proteins), but serves instead in the endosomal recycling pathway (Kouranti et al. 2006; Sato et al. 2008).

Similarly, we find that the two distinct RAB1 subgroups previously reported for angiosperms, RabD1 and RabD2 (Rutherford and Moore 2002), were already present before the radiation of sperm plants. In this case, the RabD1 paralog is apparently the derived one, whereas RabD2 sequences have remained more like the only RAB1 gene in spore plants and algae. Data from ferns and allies are needed to pinpoint the actual origin of RabD1 within embryophytes and functional characterisation of RabD1 genes is necessary to test for their potentially novel cellular function.

Amoebozoan species are characterised by harbouring extended families of RAB1 paralogs whose evolutionary origin is difficult to trace from the present analysis. Because there is no evidence in our tree for a specific relationship of any amoebozoan RAB1 sequence with sequences from other eukaryotic lineages, the most parsimonious assumption is that the paralogs have emerged from duplications within this supergroup. The existence of some paralogs shared by distantly related representatives (e.g. *Dictyostelium* and *Mastigamoeba*) indicates that Amoebozoa may have duplicated their RAB1 genes relatively early in their history, but much better sampling of the supergroup is required to learn more about the evolution of its RAB1 subfamily. The same conclusion applies to the diversity of RAB1 paralogs in Excavata.

A new RAB1 paralog unique to the chromalveolates and rhizarians. The most interesting result from our RAB1 phylogeny is the evidence for a broader occurrence of a RAB1 paralog typified by Rab1A from *P. falciparum*. Our data conclusively demonstrate orthologs of the Rab1A gene in other myzozoans, stramenopiles, and the chlorarachniophyte *Gymnochlora*, representing the supergroup Rhizaria. We also found candidate Rab1A genes in ciliates, although the bootstrap support for the relationship is not strong. However, because ciliates are a well-established sistergroup of Myzozoa and the respective RABs exhibit the Rab1A-specific signature residues, we suggest that they are

divergent but *bonafide* Rab1A orthologs. The multiple rounds of duplications and extensive divergence observed for the ciliate Rab1A genes are a fate rather characteristic for ciliate genes in general (see, e.g. Aury et al. 2006; Zufall et al. 2006). Hence, we conclude that the Rab1A paralog has been retained in every species of alveolates, stramenopiles, and Rhizaria with the genome sequence available for examination.

It has been noticed already that *Plasmodium* Rab1A is a specialised, derived paralog compared with *Plasmodium* Rab1B and prototypical RAB1 sequences from other eukaryotes (Quevillon et al. 2003). This is consistent with our results, as the non-Rab1A sequences from stramenopiles, myzozoans, and *Gymnochlora* have generally shorter branches. The Rab1A paralog is also differentiated by several unique substitutions. Interestingly, they map to regions of the RAB protein mediating its interaction with specific effectors (e.g. see Ostermeier and Brunger 1999). The Rab1A paralog may thus use a new set of effectors and potentially serves a new function compared with the prototypical Rab1B paralog. The intracellular localisation of Rab1A in *Plasmodium* (Quevillon et al. 2003) is consistent with a function in the early secretory pathway (as it is expected for a RAB1 subfamily member), so additional experimental data are needed to test for the predicted functional differences between Rab1A and Rab1B, and how consistent the function is between Rab1A proteins in different taxa.

The specific relationship of the chlorarachniophyte Rab1A to genes from stramenopiles and alveolates is not without precedents. Several previous studies showed that a number of genes from *Bigelowiella natans*, most of them encoding plastid-targeted proteins, are closely related to genes from stramenopiles/alveolates, and this phylogenetic pattern was interpreted as indicating horizontal gene transfer (HGT) to the chlorarachniophyte lineage (Archibald et al. 2003; Li et al. 2006; Obornik and Green 2005; Petersen et al. 2006; Rogers et al. 2007; Teich et al. 2007). However, recent results from phylogenomic analyses demonstrating that stramenopiles, alveolates, and Rhizaria (including chlorarachniophytes) form a clade (Burki et al. 2007, 2008; Hackett et al. 2007; Hampl et al. 2009; Rodriguez-Ezpeleta et al. 2007) have raised the possibility that at least some of these genes were vertically inherited from the common ancestor of this ''SAR'' clade (see the discussion in Lane and Archibald 2008). The two chlorarachniophyte Rab1A genes could still be due to HGT, but, notably, they do not exhibit a specific relationship to any stramenopile or alveolate sublineage in our analysis, so the donor group in the case of HGT scenario remains hypothetical at present. We therefore prefer the explanation invoking vertical inheritance from an ancestor shared with stramenopiles and alveolates, in agreement with the phylogenomic evidence for the SAR clade. Better sampling, especially of additional rhizarian lineages (foraminiferans, radiolarians, cercomonads etc.), will help to distinguish between the alternative explanations for the occurrence of Rab1A in chlorarachniophytes.

In addition to species of the SAR clade, we found potential ortholog of the *Plasmodium* Rab1A in the cryptophyte *G. theta*, but elucidating the relationship of this gene to other RAB1 genes is complicated by its divergent nature and the lack of genome data from other cryptophytes. The occurrence of an ortholog of the stramenopile/alveolate Rab1A in cryptophytes would be appealing given the fact that it would add support to relationship of these taxa postulated by the chromalveolate hypothesis (Cavalier-Smith 1999). Recent evidence indicates that cryptophytes are specifically related to haptophytes (Burki et al. 2007; Hackett et al. 2007; Patron et al. 2007; Rice and Palmer 2006) and a few more poorly studied lineages (kathablepharids, *Telonema*, and perhaps picobiliphytes; Kim, Simpson, and Graham 2006; Not et al. 2007; Okamoto and Inouye 2005; Shalchian-Tabrizi et al. 2006), but the position of this clade in relation to alveolates and stramenopiles remains contentious (Burki et al. 2008; Hampl et al. 2009; Kim and Graham 2008). If the cryptophyte RAB1A-like gene were truly orthologous to Rab1A of the SAR clade, we would expect the Rab1A paralog in haptophytes, too. However, the draft genome sequence of the haptophyte *Emiliania huxleyi* apparently lacks it, necessitating a secondary loss or divergence beyond recognition in at least this haptophyte representative, or some model involving HGT.

We tried to find among the divergent RAB1 sequences additional candidates for *Plasmodium* Rab1A-like genes but have not found any BLASTP searches, looking for the Rab1A-specific signature sequences, or phylogenetic analyses. The *Guillardia* gene thus remains as currently the only candidate for the *Plasmodium* Rab1A-like paralog outside the SAR clade in the available sample of RAB sequences.

Concluding remarks. Phylogenetic analysis of the RAB1 subfamily suggests that the multiple RAB1 paralogs in most eukaryotic species stem mainly from independent, lineage-specific duplication events. This finding reflects considerable evolutionary dynamics of the RAB1 subfamily and flexibility with which regulation of the endomembrane trafficking has been modified over the course of eukaryotic evolution. The best understood lineage-specific expansion within the RAB1 subfamily is the duplication resulting in Rab1 and Rab35 early in metazoan evolution. We now present evidence that the *Plasmodium* Rab1A and Rab1B may also derive from an ancient duplication possibly taking place before the radiation of stramenopiles, alveolates, Rhizaria (i.e. the SAR clade), and perhaps also cryptophytes. The shared possession of two clearly distinct RAB1 paralogs points to a shared trafficking route mediated by the specialised Rab1A paralog. Increasing the taxonomic sampling of RAB1 data will be needed to determine whether Rab1A is widespread in Rhizaria, whether the divergent *Guillardia* Rab1A-like sequence is indeed orthologous to the Rab1A, and whether such an ortholog has been indeed lost from haptophytes.

ACKNOWLEDGMENTS

This work was supported by the research project no. 21620828 of Czech Ministry of Education, the PEP project of Genome Canada/Genome Atlantic, and a grant from the Natural Sciences and Engineering Research Council of Canada (to P.J.K.). P.J.K. is a Fellow of the Canadian Institute for Advanced Research and a Senior Scholar of the Michael Smith Foundation for Health Research. We thank the Joint Genome Institute's Community Sequencing Program (http://www.jgi.doe.gov/sequencing/why/50026.html) for their ongoing efforts to sequence the nuclear genomes of *G. theta* and *Bigelowiella natans*, K. Barry and E. Lindquist of the JGI for project management and data availability, and J.M. Archibald, M.W. Gray, G.I. McFadden, and C.E. Lane for their contributions to the project. We also thank TBestDB for support with EST sequence analysis. Preliminary sequence data from *Toxoplasma gondii* were obtained from The Institute for Genomic Research website at http://www.tigr.org. Sequencing of *T. gondii* was funded by the National Institute of Allergy and Infectious Disease. Genome sequence data from *Eimeria tenella* and *Babesia bigemina* were provided by the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/). Genome sequence data from *Aureococcus anophagefferens*, *E. huxleyi*, *Micromonas pusilla*, and *N. gruberi* were produced by the U.S. Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication only. Genome sequence data from *Hyaloperonospora parasitica* were produced by the Genome Sequencing Center at Washington University School of Medicine in St. Louis and can be obtained from http://genome.wustl.edu/pub/organism/Fungi/Hyaloperonospora_parasitica/. Sequence data

from *Galdieria sulphuraria* were obtained from the Michigan State University Galdieria Database (http://genomics.msu.edu/galdieria). We also highly acknowledge the opportunity to use unassembled WGS sequencing reads from *Amphimedon queenslandica* (= *Reniera* sp.) produced by DOE JGI, from *Oxytricha trifallax* produced by GSC WUSTL, from *Capsaspora owczarzaki* produced by Broad Institute, from *Acanthamoeba castellanii* produced by BCM HGSC, and from *Ectocarpus siliculosus* produced by Genoscope.

LITERATURE CITED

Ackers, J. P., Dhir, V. & Field, M. C. 2005. A bioinformatic analysis of the RAB genes of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **141**:89–97.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**:3389–3402.

Anisimova, M. & Gascuel, O. 2006. Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. *Syst. Biol.*, **55**:539–552.

Archibald, J. M., Rogers, M. B., Toop, M., Ishida, K. & Keeling, P. J. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc. Natl. Acad. Sci. USA*, **100**:7678–7683.

Aury, J. M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Câmara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A. M., Kissmehl, R., Klotz, C., Koll, F., Le Mouël, A., Lepère, G., Malinsky, S., Nowacki, M., Nowak, J. K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Bétermier, M., Weissenbach, J., Scarpelli, C., Schächter, V., Sperling, L., Meyer, E., Cohen, J. & Wincker, P. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, **444**:171–178.

Batoko, H., Zheng, H. Q., Hawes, C. & Moore, I. 2000. A rab1 GTPase is required for transport between the endoplasmic reticulum and Golgi apparatus and for normal golgi movement in plants. *Plant Cell*, **12**:2201–2218.

Becker, B. & Melkonian, M. 1996. The secretory pathway of protists: spatial and functional organization and evolution. *Microbiol. Rev.*, **60**:697–721.

Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjaeveland, A., Nikolaev, S. I., Jakobsen, K. S. & 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.

Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Muller, S., Dessi, D., Fiori, P. L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F. D., Simoes-Barbosa, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Pertea, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osoegawa, K., de Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M. **Jr.**, Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y. S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M. & Johnson, P. J. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science*, **315**:207–212.

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.

Dacks, J. B. & Field, M. C. 2004. Eukaryotic cell evolution from a comparative genomic perspective: the endomembrane system. *In*: Hirt, R. & Horner, D. (ed.), Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics. Taylor & Francis Books, London. p. 309–334.

Dacks, J. B., Poon, P. P. & Field, M. C. 2008. Phylogeny of endocytic components yields insight into the process of non-endosymbiotic organelle evolution. *Proc. Natl. Acad. Sci. USA*, **105**:588–593.

Dhir, V., Goulding, D. & Field, M. C. 2004. TbRAB1 and TbRAB2 mediate trafficking through the early secretory pathway of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **137**:253–265.

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Smith, R. K. **Jr.**, Garg, J., Pearlman, R. E., Karrer, K. M., Sun, L., Manning, G., Elde, N. C., Turkewitz, A. P., Asai, D. J., Wilkes, D. E., Wang, Y., Cai, H., Collins, K., Stewart, B. A., Lee, S. R., Wilamowska, K., Weinberg, Z., Ruzzo, W. L., Wloga, D., Gaertig, J., Frankel, J., Tsao, C. C., Gorovsky, M. A., Keeling, P. J., Waller, R. F., Patron, N. J., Cherry, J. M., Stover, N. A., Krieger, C. J., del Toro, C., Ryder, H. F., Williamson, S. C., Barbeau, R. A., Hamilton, E. P. & Orias, E. 2006. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.*, **4**:e286.

Elias, M. 2008. The guanine nucleotide exchange factors Sec2 and PRONE: candidate synapomorphies for the Opisthokonta and the Archaeplastida. *Mol. Biol. Evol.*, **25**:1526–1529.

Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the Author. Department of Genome Sciences, University of Washington, Seattle. Available at http://evolution.genetics.washington.edu/phylip.html (accessed 2004).

Field, M. C., Gabernet-Castello, C. & Dacks, J. B. 2007. Reconstructing the evolution of the endocytic system: insights from genomics and molecular cell biology. *Adv. Exp. Med. Biol.*, **607**:84–96.

Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L. & Bateman, A. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.*, **34**:D247–251.

Gascuel, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.*, **14**:685–695.

Gould, S. B., Tham, W. H., Cowman, A. F., McFadden, G. I. & Waller, R. F. 2008. Alveolins, a new family of cortical proteins that define the protist infrakingdom Alveolata. *Mol. Biol. Evol.*, **25**:1219–1230.

Grosshans, B. L., Ortiz, D. & Novick, P. 2006. Rabs and their effectors: achieving specificity in membrane traffic. *Proc. Natl. Acad. Sci. USA*, **103**:11821–11827.

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

Hackett, J. D., Yoon, H. S., Li, S., Reyes-Prieto, A., Rummele, S. E. & Bhattacharya, D. 2007. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of 'Rhizaria' with chromalveolates. *Mol. Biol. Evol.*, **24**:1702–1713.

Hampl, V., Hug, L., Leigh, J. W., Dacks, J. B., Lang, B. F., Simpson, A. G. B. & Roger, A. J. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic ''Supergroups''. *Proc. Natl. Acad. Sci. USA*, **106**:3859–3864.

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.

Huang, X. & Madan, A. 1999. CAP3: a DNA Sequence Assembly Program. *Genome Res.*, **9**:868–877.

Kaiser, S. E., Brickner, J. H., Reilein, A. R., Fenn, T. D., Walter, P. & Brunger, A. T. 2005. Structural basis of FFAT motif-mediated ER targeting. *Structure*, **13**:1035–1045.

Keeling, P. J., Burger, G., Durnford, D. G., Lang, B. F., Lee, R. W., Pearlman, R. E., Roger, A. J. & Gray, M. W. 2005. The tree of eukaryotes. *Trends Ecol. Evol.*, **20**:670–676.

Kim, E. & Graham, L. E. 2008. EEF2 analysis challenges the monophyly of Archaeplastida and Chromalveolata. *PLoS ONE*, **3**:e2621.

Kim, E., Simpson, A. G. & Graham, L. E. 2006. Evolutionary relationships of apusomonads inferred from taxon-rich analyses of 6 nuclear encoded genes. *Mol. Biol. Evol.*, **23**:2455–2466.

Kouranti, I., Sachse, M., Arouche, N., Goud, B. & Echard, A. 2006. Rab35 regulates an endocytic recycling pathway essential for the terminal steps of cytokinesis. *Curr. Biol.*, **16**:1719–1725.

Lal, K., Field, M. C., Carlton, J. M., Warwicker, J. & Hirt, R. P. 2005. Identification of a very large Rab GTPase family in the parasitic protozoan *Trichomonas vaginalis*. *Mol. Biochem. Parasitol.*, **143**:226–235.

Lane, C. E. & Archibald, J. M. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol. Evol.*, **23**:268–275.

Langford, T. D., Silberman, J. D., Weiland, M. E., Svard, S. G., McCaffery, J. M., Sogin, M. L. & Gillin, F. D. 2002. *Giardia lamblia* identification and characterization of Rab and GDI proteins in a genome survey of the ER to Golgi endomembrane system. *Exp. Parasitol.*, **101**:13–24.

Li, S., Nosenko, T., Hackett, J. D. & Bhattacharya, D. 2006. Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. *Mol. Biol. Evol.*, **23**:663–674.

Montsant, A., Allen, A. E., Coesel, S., De Martino, A., Falciatore, A., Mangogna, M., Siaut, M., Heijde, M., Jabbari, K., Maheswari, U., Rayko, E., Vardi, A., Apt, K. E., Berges, J. A., Chiovitti, A., Davis, A. K., Thamatrakoln, T., Hadi, M. Z., Lane, T. W., Lippmeier, J. C., Martinez, D., Parker, M. S., Pazour, G. J., Saito, M. A., Rokhsar, D. S., Armbrust, E. V. & Bowler, C. 2007. Identification and comparative genomic analysis of signaling and regulatory components in the diatom *Thalassiosira pseudonana*. *J. Phycol.*, **43**:585–604.

Morsomme, P. & Riezman, H. 2002. The Rab GTPase Ypt1p and tethering factors couple protein sorting at the ER to vesicle targeting to the Golgi apparatus. *Dev. Cell*, **2**:307–317.

Nicholas, K. B. & Nicholas, H. B. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. (Distributed by the authors). Available at http://www.nrbsc.org/gfx/genedoc/index.html (accessed 2004).

Not, F., Valentin, K., Romari, K., Lovejoy, C., Massana, R., Tobe, K., Vaulot, D. & Medlin, L. K. 2007. Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science*, **315**:253–255.

Obornik, M. & Green, B. R. 2005. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol. Biol. Evol.*, **22**:2343–2353.

Odronitz, F. & Kollmar, M. 2007. Drawing the tree of eukaryotic life based on the analysis of 2,269 manually annotated myosins from 328 species. *Genome Biol.*, **8**:R196.

Okamoto, N. & Inouye, I. 2005. The katablepharids are a distant sister group of the Cryptophyta: a proposal for Katablepharidophyta divisio nova/Kathablepharida phylum novum based on SSU rDNA and beta-tubulin phylogeny. *Protist*, **156**:163–179.

Ostermeier, C. & Brunger, A. T. 1999. Structural basis of Rab effector specificity: crystal structure of the small G protein Rab3A complexed with the effector domain of rabphilin-3A. *Cell*, **96**:363–374.

Patron, N. J., Inagaki, Y. & Keeling, P. J. 2007. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr. Biol.*, **17**:887–891.

Patron, N. J., Rogers, M. B. & Keeling, P. J. 2004. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot. Cell*, **3**:1169–1175.

Pereira-Leal, J. B. & Seabra, M. C. 2001. Evolution of the Rab family of small GTP-binding proteins. *J. Mol. Biol.*, **313**:889–901.

Petersen, J., Teich, R., Brinkmann, H. & Cerff, R. 2006. A ''green'' phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts, and dinoflagellates. *J. Mol. Evol.*, **62**:143–157.

Pfeffer, S. & Aivazian, D. 2004. Targeting Rab GTPases to distinct membrane compartments. *Nat. Rev. Mol. Cell. Biol.*, **5**:886–896.

Quevillon, E., Spielmann, T., Brahimi, K., Chattopadhyay, D., Yeramian, E. & Langsley, G. 2003. The *Plasmodium falciparum* family of Rab GTPases. *Gene*, **306**:13–25.

Rice, D. W. & Palmer, J. D. 2006. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol.*, **4**:31.

Richards, T. A. & Cavalier-Smith, T. 2005. Myosin domain evolution and the primary divergence of eukaryotes. *Nature*, **436**:1113–1118.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burger, G., Roger, A. J., Gray, M. W., Philippe, H. & Lang, B. F. 2007. Towards resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr. Biol.*, **17**:1420–1425.

Rogers, M. B., Watkins, R. F., Harper, J. T., Durnford, D. G., Gray, M. W. & Keeling, P. J. 2007. A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer. *BMC Evol. Biol.*, **7**:89.

Rutherford, S. & Moore, I. 2002. The *Arabidopsis* Rab GTPase family: another enigma variation. *Curr. Opin. Plant Biol.*, **5**:518–528.

Saito-Nakano, Y., Loftus, B. J., Hall, N. & Nozaki, T. 2005. The diversity of Rab GTPases in *Entamoeba histolytica*. *Exp. Parasitol.*, **110**: 244–252.

Sato, M., Sato, K., Liou, W., Pant, S., Harada, A. & Grant, B. D. 2008. Regulation of endocytic recycling by *C. elegans* Rab35 and its regulator RME-4, a coated-pit protein. *EMBO J.*, **27**:1183–1196.

Shalchian-Tabrizi, K., Eikrem, W., Klaveness, D., Vaulot, D., Minge, M. A., Le Gall, F., Romari, K., Throndsen, J., Botnen, A., Massana, R., Thomsen, H. A. & Jakobsen, K. S. 2006. Telonemia, a new protist phylum with affinity to chromist lineages. *Proc. Biol. Sci.*, **273**:1833–1842.

Stamatakis, A., Hoover, P. & Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.*, **57**:758–771.

Teich, R., Zauner, S., Baurain, D., Brinkmann, H. & Petersen, J. 2007. Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in Plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses. *Protist*, **158**:263–276.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**:4876–4882.

Weeks, G., Gaudet, P. & Insall, R. H. 2005. The Small GTPase Superfamily. *In*: Loomis, W. F. & Kuspa, A. (ed.), *Dictyostelium* Genomics. Horizon Scientific Press, Norfolk. p. 173–210.

Zerial, M. & McBride, H. 2001. Rab proteins as membrane organizers. *Nat. Rev. Mol. Cell Biol.*, **2**:107–117.

Zufall, R. A., McGrath, C. L., Muse, S. V. & Katz, L. A. 2006. Genome architecture drives protein evolution in ciliates. *Mol. Biol. Evol.*, **23**:1681–1687.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. RAB1 subfamily genes analysed.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received: 01/19/09; accepted: 02/20/09