

Contrasting patterns in the evolution of the Rab GTPase family in Archaeplastida

Romana Petrželková, Marek Eliáš*

Department of Biology and Ecology, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic

Abstract

Rab GTPases are a vast group of proteins serving a role of master regulators in membrane trafficking in eukaryotes. Previous studies delineated some 23 Rab and Rab-like paralogs ancestral for eukaryotes and mapped their current phylogenetic distribution, but the analyses relied on a limited sampling of the eukaryotic diversity. Taking advantage of the recent growth of genome and transcriptome resources for phylogenetically diverse plants and algae, we reanalyzed the evolution of the Rab family in eukaryotes with the primary plastid, collectively constituting the presumably monophyletic supergroup Archaeplastida. Our most important novel findings are as follows: *(i)* the ancestral set of Rabs in Archaeplastida included not only the paralogs Rab1, Rab2, Rab5, Rab6, Rab7, Rab8, Rab11, Rab18, Rab23, Rab24, Rab28, IFT27, and RTW (=Rab12), as suggested previously, but also Rab14 and Rab34, because Rab14 exists in glaucophytes and Rab34 is present in glaucophytes and some green algae; *(ii)* except in embryophytes, Rab gene duplications have been rare in Archaeplastida. Most notable is the independent emergence of divergent, possibly functionally novel, in-paralogs of Rab1 and Rab11 in several archaeplastidial lineages; *(iii)* recurrent gene losses have been a significant factor shaping Rab gene complements in archaeplastidial species; for example, the Rab21 paralog was lost at least six times independently within Archaeplastida, once in the lineage leading to the “core” eudicots; *(iv)* while the glaucophyte *Cyanophora paradoxa* has retained the highest number of ancestral Rab paralogs among all archaeplastidial species studied so far, rhodophytes underwent an extreme reduction of the Rab gene set along their stem lineage, resulting in only six paralogs (Rab1, Rab2, Rab6, Rab7, Rab11, and Rab18) present in modern red algae. Especially notable is the absence of Rab5, a virtually universal paralog essential for the endocytic pathway, suggesting that endocytosis has been highly reduced or rewired in rhodophytes.

Keywords: Archaeplastida; Chloroplastida; endocytosis; evolution; Glaucophyta; Rab GTPases; Rhodophyta

Introduction

The family of Rab GTPases, constituting the largest subgroup of the Ras GTPase superfamily, is one of the hallmarks of the eukaryotic cell. Rabs serve as central regulators of membrane trafficking and are involved in maintaining identity of the various compartments of the membrane system and in ensuring specificity of the transport events between the compartments [1]. Our understanding of the Rab function derives primarily from studies on a few selected model systems, primarily mammalian cells and the yeast *Saccharomyces cerevisiae*, but more limited knowledge exists also for other species representing different distantly related eukaryotic lineages, for example the kinetoplastid *Trypanosoma brucei*, the ciliate *Tetrahymena thermophila*, and the plant *Arabidopsis thaliana* [2–4]. One of the puzzling aspects of the biology of the Rab family is the fact that the

total number of Rab genes may differ profoundly between different species: whereas some eukaryotic cells are able to secure their proper functioning with less than ten Rabs, other species exhibit tens or even hundreds of Rab paralogs [3,5].

Comparative genomic and phylogenetic studies revealed that a large number of distinct Rab paralogs have been established early in the evolution of eukaryotes [3,5,6]. Reconstructions of the Rab complement in the deepest point of the phylogeny of extant eukaryotes, i.e. the last eukaryotic common ancestor (LECA), suggest the existence of over 20 paralogs [5,6]. This is consistent with the presence of a highly elaborate endomembrane system in the LECA, in line with the emerging view of the LECA as a fully fledged and surprisingly complex eukaryotic cell [7,8].

However, the exact number of ancestral eukaryotic Rab paralogs still remains uncertain due to three factors. Firstly, classification of some Rab-like proteins lacking the C-terminal tail with a prenylation motif, such as RTW (=Rab12; [6,9]) or IFT27 (=Rab14; [6,10]), as bona fide Rab family members is controversial due to the poor resolution of Ras superfamily phylogenies. Second, the inference on

* Corresponding author. Email: marek.elias@osu.cz

Handling Editor: Andrzej Bodyl

the ancestral Rab complement depends on the position of the root of the eukaryotic phylogeny; several competing hypotheses have been recently discussed in the literature, but no consensus currently exists on where the root actually lies (see [11]). Third, the accuracy of the reconstruction of the Rab complement in the LECA significantly depends on the sampling of the eukaryotic phylogenetic diversity. Indeed, the analyses published so far [5,6,12] relied on only a limited number of genome sequences or only transcriptomic (expressed sequence tag – EST) data for many crucial eukaryotic lineages, while other lineages have not been studied at all. Recent progress in DNA sequencing technologies has enabled to dramatically improve the sampling of the eukaryotic phylogenetic diversity by full genome or deep transcriptome sequencing, although important gaps still persist [13,14].

Archaeplastida, often called Plantae, are a major eukaryotic supergroup defined by the synapomorphic presence of a primary plastid, i.e. a direct product of the original endosymbiotic acquisition of a cyanobacterial ancestor of eukaryotic plastids [15]. Such a plastid is found in three living eukaryotic lineages – glaucophytes (Glaucophyta or Glaucocystophyta), rhodophytes (Rhodophyta, Rhodophyceae, or Rhodoplantae), and the “green lineage” comprising green algae and their descendants land plants (Chloroplastida or Viridiplantae) [16,17]. In the most parsimonious scenario, these three lineages constitute a monophyletic grouping to the exclusion of other eukaryotes. However, phylogenomic analyses have so far failed to provide conclusive evidence for this hypothesis, because some other lineages, specifically haptophytes and/or cryptists, tend to disrupt the monophyly of the three archaeplastidial groups in some analyses (see, e.g., [18,19]). This would suggest a secondary loss of the primary plastid from some eukaryotes. Regardless these controversies, the monophyly of the Archaeplastida sensu Adl et al. [15] remains the preferred working hypothesis that will also be assumed in this study.

Our knowledge about the cell biology of the different archaeplastidial lineages is extremely uneven and biased towards Chloroplastida, particularly towards land plants (embryophytes) and the model green alga *Chlamydomonas reinhardtii*. In rhodophytes, *Cyanidioschyzon merolae* representing the basal lineage of red algae (Cyanidiophyceae) has been established as a highly useful model system for addressing diverse cell biological questions and it happened to become the first alga with a sequenced nuclear genome [20]. However, it is questionable to what extent this unicellular extremophilic species may be representative for red algae as a whole. *Cyanophora paradoxa* is a glaucophyte that has been used as a model organism for the whole group [21], but the knowledge on this species lags far behind the model systems of the other two archaeplastidial lineages.

However, a lot of key insights into the cell biology, biochemistry or physiology of any organisms can be obtained by computational analyses of their genetic blueprints. Fortunately, recent years have witnessed a rapid accumulation of genomic or transcriptomic data from both red algae and glaucophytes. These include draft genome sequences of the phylogenetically diverse rhodophytes *Chondrus crispus* [22], *Porphyridium purpureum* [23], *Galdieria sulphuraria* [24], and *Pyropia yezoensis* [25], and of the glaucophyte

C. paradoxa [26]. Transcriptomes of an even broader set of red algal and glaucophyte species have been deeply sequenced thanks to the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) [14]. Hence, we now have a unique opportunity to quickly improve our knowledge about the molecular underpinnings of red algal and glaucophyte cells by exploring the wealth of these data by computational analyses.

Substantial effort has been put into studying Rab GTPases in model plant species, primarily *A. thaliana*, which enabled to demonstrate that the plant Rab complement exhibits at the functional level both features shared with other eukaryotes as well as novel, plant specific features [2,27]. Comparative genomic and phylogenetic studies have additionally shown that land plants have retained only a subset of the presumed ancestral eukaryotic Rab paralogs, but secondarily expanded the Rab family by extensive gene duplications [6,28,29]. Much less is known about Rabs in algae. Except occasional early studies (e.g. [30]), attempts to functionally characterize algal Rab proteins are virtually lacking. Phylogenetic studies on algal Rabs have been also limited. Concerning Archaeplastida, the most comprehensive analysis published to date [6] included data from only four green algal genomes (*C. reinhardtii*, *Chlorella variabilis*, *Ostreococcus lucimarinus* and *Micromonas pusilla* CCMP1545) and two red algal genomes (*C. merolae*, *G. sulphuraria*), while only partial transcriptomic data from classical (Sanger sequencing-based) EST surveys were available for glaucophytes (*C. paradoxa* and *Glaucocystis nostochinearum*).

The aim of this study is to improve our knowledge about the diversity and evolution of Rab GTPases in Archaeplastida, focusing specifically on algal lineages. A significantly expanded sampling of the archaeplastidial diversity offered many new important insights helping us to refine the view of the cellular evolution in this highly significant assemblage of eukaryotic organisms.

Material and methods

Species and sequence data analyzed

The sequence dataset used in this study resulted as an expansion of the dataset of Rab sequences analyzed previously by Elias et al. [6], which included, in addition to the sequences from algal species mentioned above, Rabs from the eudicot *A. thaliana* and the lycophyte *Selaginella moellendorffii*. This dataset was revised and expanded using newly available data. We replaced the incomplete representation of Rab sequences from *C. paradoxa* as derived from an EST survey by a (presumably) complete set of sequences deduced from a recently reported draft genome assembly. We added sequences from a deeply sequenced transcriptome of the glaucophyte *Cyanoptylche gloeocystis* and we instead removed the extremely fragmentary set of sequences from the glaucophyte *G. nostochinearum* (which was anyway not informative beyond what was implied by the *C. paradoxa* and *C. gloeocystis* datasets). We also tested two independent transcriptome assemblies from the glaucophyte *Gloeochaete wittrockiana* made available by the MMETSP project (MMETSP0308 and MMETSP1089;

<http://camera.calit2.net/mmetasp/list.php>), but it turned out that both assemblies are heavily contaminated by another organism, most likely an amoebozoan (data not shown), perhaps due to contamination of the original culture (SAG 46.84). We therefore omitted *G. wittrockiana* from our analyses. We expanded the sampling of the rhodophyte diversity (only the class Cyanidiophyceae had been represented in the original dataset) by extracting Rab sequences from three newly released genome sequences representing three additional rhodophyte classes (Florideophyceae, Bangiophyceae, Porphyridiophyceae). Transcriptome assemblies additionally allowed us to include one representative of each of the remaining three classes (Compsopogonophyceae, Stylonematophyceae, Rhodellophyceae). Although transcriptome assemblies have recently become available for a number of phylogenetically diverse green algae, we decided not to include them in our analysis due to frequent contamination issues and to keep the size of the dataset within reasonable limits (we nevertheless used the assemblies for certain targeted analyses, see below). However, to improve the coverage of the Chloroplastida group in our analysis, we added Rab sequences from four green algal genomes and two land plant genomes. All species systematically analyzed are listed in [Tab. 1](#). Links to the sources of the sequence data are available in [Tab. S1](#).

Extraction and curation of Rab sequences

We used the program BLAST and its appropriate variants (blastp, tblastn, blastn) [31] to identify sequences of candidate Rab genes and proteins in the genome or transcriptome assemblies or the corresponding protein sequence predictions. The identified sequences were BLASTed against our local database of annotated Ras superfamily GTPases to discriminate genuine Rabs from other subgroups of the superfamily. As mentioned above, the delimitation of the Rab family is not completely settled; to be consistent with our previous study [6] we included in the analysis two Rab-like paralogs (RTW and IFT27) and the GTPase RAN. The latter is a universal eukaryotic gene found in every species investigated so far, so it provided an internal control of the completeness of the genome or transcriptome resource for a given species. Genome sequences were checked by tblastn for the possible presence of genes missing in the respective predicted protein sequence sets. Transcriptome data for the same species (EST databases or TSA – transcript shotgun assemblies) were also checked to identify possible genes missing in the draft genome assemblies. If needed, partial gene sequences due to gaps in the genome assembly were combined with the corresponding transcript sequence to obtain a complete coding sequence of the gene. Some Rab genes in the red alga *C. crispus* and in *Oryza sativa* could be identified in contigs or scaffolds not included in the most recent genome release, but their authenticity was indisputable. Existing protein sequence predictions were carefully checked by inspecting alignments to related sequences and all suspicious cases were reevaluated by investigating the respective nucleotide sequence, in many cases leading to a revision of the gene model and the resulting protein sequence. In species with only TSA and no genome sequence available, some Rab genes were represented by incomplete

transcript sequences, but for many of them a complete or at least longer coding sequence could be obtained by iterative addition of matching raw Illumina reads in the Short sequence archive (<http://www.ncbi.nlm.nih.gov/sra/>). Two transcripts from the red alga *Rhodella maculata* remained too incomplete to be included in phylogenetic analyses, but their identity as Rab2 and Rab7 was indisputable from BLAST comparisons with Rabs from other rhodophytes. A list of all sequences analyzed in this study, together with the corresponding accession numbers or sequence identifiers, is available in [Tab. S1](#). Revised or newly predicted protein sequences are provided in a separate supplementary file.

Alignment and phylogenetic analyses

The newly identified Rab protein sequences were divided into groups each representing a different ancestral paralog (the assignment of individual sequences to the paralogs was in virtually all cases straightforward based on BLAST comparisons) and for each group multiple alignment was created using MAFFT (version 7, default parameters; <http://mafft.cbrc.jp/alignment/server/> [32]). Each aligned group was then added manually to a large master alignment built for our previous study [6], using previously aligned sequences of the same paralog as a guide. From the expanded master alignment subsets of sequences were selected to create desired smaller alignments for phylogenetic analyses. We applied the same mask as before [6] to remove columns where the alignment was too uncertain. Phylogenetic trees were inferred using the maximum likelihood (ML) method as implemented in the program RAXML-HPC BlackBox (8.0.24) [33] accessible at the CIPRES Science Gateway (<https://www.phylo.org/portal2> [34]). The substitution model employed was LG+I, branch support was assessed by the rapid bootstrapping algorithm that is an inherent part of the best tree search strategy of RAXML. To test the robustness of the tree topologies we also employed ML inference using the program PhyML-CAT and the empirical profile mixture model C20 [35] with gamma correction (four categories) of the among-site rate heterogeneity; Chi2-based parametric branch support was calculated using the approximate likelihood ratio test implemented in PhyML (-b -2 option). Trees were visualized using iTOL (<http://itol.embl.de/> [36]) and rendered for publication using a graphical editor.

Results and discussion

Virtually all Rab genes in Archaeplastida can be readily assigned to known ancestral Rab paralogs

We relied on complete or high-quality draft genome sequences and/or deeply sequenced transcriptomes to build a manually curated set of Rab family protein sequences from 22 species of the Archaeplastida supergroup: two representatives of Glaucophyta, eight members of Rhodophyta, and twelve members of Chloroplastida ([Tab. 1](#)). While the sampling for glaucophytes, here represented only by one species with a draft genome sequence and one species with a deeply sequenced transcriptome, remains rather limited, the phylogenetic diversity of Rhodophyta and Chloroplastida is covered more comprehensively, which gives us an

Tab. 1 Rab GTPases in Archaeplastida.

	Rab1	Rab2	Rab5	Rab6	Rab7	Rab8	Rab11	Rab14	Rab18	Rab21	Rab23	Rab24	Rab28	Rab34	IFT27	RTW
Glaucophyta	<i>Cyanophora paradoxa</i>	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	<i>Cyanoptylche gloeocystis</i> (T)	5	1	?	1	1	4	?	?	1(2?)	?	?	1(2?)	1	?	1
Rhodophyta	<i>Galdieria sulphuraria</i>	1	1	1	1	1	2									
	<i>Cyanidioschyzon merolae</i>	1	1	1	1	1	1	1	1							
	<i>Chondrus crispus</i>	1	1	1	1	1	2									
	<i>Pyropia yezoensis</i>	1	1	1	1	1	1	1	1							
	<i>Rhodella maculata</i> (T)	1	1	1	1	1	1	?								
	<i>Porphyridium purpureum</i>	2	1	1	1	1	1	1	1							
	<i>Rhodosorus marinus</i> (T)	2	1	1	1	1	2	1	1							
	<i>Compsopogon coeruleus</i> (T)	2	1	1	1	1	1	1	1							
Chloroplastida	<i>Glulamnydomonas reinhardtii</i>	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1
	<i>Volvox carteri</i>	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1
	<i>Coccomyxa subellipsoidea</i>	1	1	2 (1xF1)	1	1	1	1	1	1	1	1	1	1	1	1
	<i>Chlorella variabilis</i>	1	1	2 (1xF1)	1	2	2	1	1	1	1	1	1	1	1	1
	<i>Ostreococcus lucimarinus</i>	1	1	2 (1xF1)	1	1	1	1	1	1	1	1	1	1	1	1
	<i>Micromonas pusilla</i> CCMP1545	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	<i>Micromonas pusilla</i> NOUM17	1	1	2 (1xF1)	1	1	1	1	1	1	1	1	1	1	1	1
	<i>Klebsormidium flaccidum</i>	2	1	2 (1xF1)	1	1	1	1	1	1	1	1	1	1	1	1
	<i>Physcomitrella patens</i>	3	4(+1)	5 (3xF1)	7	4(+1)	5	12	5(+1)	3(+2)	1	1	1	1	1	1
	<i>Selaginella moellendorffii</i>	2	1	2 (1xF1)	1	1	3	7	1	1	1	1	1	1	1	1
	<i>Arabidopsis thaliana</i>	4	3	3 (1xF1)	5	8	5	25(+1)	3	3	1	1	1	1	1	1
	<i>Oryza sativa</i>	5	3(+2)	5 (2xF1)	2	4	4	17	3	3	1	1	1	1	1	1

Distribution of Rab GTPases, including the Rab-like proteins IFT27 (=Rab12) and RTW (=Rab14) in 22 selected archaeplastidial species. For species marked with "(T)" only a deeply sequenced transcriptome (no genome sequence) is available. For each ancestral paralog the number of in-paralogs in each species is given. The problematic Rab gene from *Cyanoptylche gloeocystis*, Rab1L (see the text) is here counted as one of the Rab1 in-paralogs. For Rab23 and Rab28 in *C. gloeocystis*, two highly similar variants are recorded in the transcriptome and it is unclear whether they represent two different genes or allelic variants. In *Physcomitrella patens*, *Arabidopsis thaliana*, and *Oryza sativa* apparent pseudogenes are present; their number for the respective ancestral paralog is indicated in parentheses. In species represented only by a transcriptome assembly question mark indicates a suspicious absence of a gene present in related species. For the Rab5 paralog in Chloroplastida the table indicates the total number of in-paralogs and in parentheses how many of them correspond to the RabF1 form.

opportunity to infer conclusions with a potentially general validity for the respective archaeplastidial lineages.

Each Rab sequence was initially assigned to one of the ancestral paralogs (as defined by Elias et al. [6]) based on BLAST similarity scores to previously annotated sequences in our in-house database of Ras superfamily GTPases. The assignment was unambiguous for virtually all sequences, except one gene from the glaucophyte *C. gloeocystis* (eventually labelled Rab1L; Tab. S1, Fig. S1), which gave similar scores to members of the closely related Rab1 and Rab8 paralogs. One additional sequence from *C. variabilis* (Rab7L; Tab. S1) was most similar in BLAST-based comparisons to various Rab7 proteins, but was so divergent (maximal identity was only around 30%) that it could not be included in the phylogenetic analysis of Rab sequences. However, the fact that the protein is so far specific for *C. variabilis* and is consistently most similar to Rab7 indicates that it is most likely an extremely divergent lineage-specific offshoot of the Rab7 paralog (*C. variabilis* additionally harbours a canonical Rab7 gene; Tab. 1). The ability to readily annotate on the basis of sequence similarity most of the archaeplastidial Rab sequences indicates that most of them have been evolving rather slowly. This contrasts with the situation in some other eukaryotic groups, e.g. Amoebozoa or Ciliata, which tend to accumulate large numbers of highly divergent Rab paralogs that are often difficult to classify even by using phylogenetic analyses [6].

To corroborate the initial annotation we performed a phylogenetic analysis of a multiple sequence alignment comprising not only the archaeplastidial sequences, but also Rab sequences from other eukaryotic groups representing ancestral eukaryotic paralogs that were not found in Archaeplastida in our previous study [6]. The ML tree inferred using the program RAXML is displayed in Fig. S1; a tree obtained using the program PhyML and a different substitution model (see “Material and methods”) was topologically different in many regions, but agreed with the RAXML tree in all branches relevant for defining the main ancestral Rab paralogs (data not shown). As observed in previous phylogenetic reconstructions of the Rab family, the topology of the tree has many poorly supported branches at all levels of the phylogenetic depth and is not free from obvious topological artefacts, including the paraphyly of the Rab1 paralog due to the Rab8 paralog nested within (encountered also in a previous analysis [6]) and the paraphyly of the Rab2 paralog due to a clade of Rab4 and Rab14 nested within. Nevertheless, the assignment of virtually all archaeplastidial Rab sequences to different ancestral paralogs can be easily deduced from the tree and is compatible with the BLAST-based assignment. The only somewhat problematic sequence is the *C. gloeocystis* Rab1L gene mentioned above. In our tree its affiliation to the group comprising Rab1 and Rab8 paralogs indicated by BLAST is confirmed, suggesting that it is most likely a lineage-specific divergent in-paralog of Rab1 and Rab8. We consider the origin from Rab1 as more probable, since Rab1 shows a general tendency to recurrently duplicate in Archaeplastida (see below) while no Rab8 duplications were found for algal taxa in Archaeplastida, and because a focused phylogenetic analysis of Rab1 and Rab8 sequences using PhyML-CAT and the empirical profile mixture model

C20 suggested that it may actually be an extremely divergent relative of some glaucophyte Rab1 sequences (see below).

The expanded sampling for the first time documents the presence of Rab14 and Rab34 in Archaeplastida

The main novel finding of the analyses described above is the revision of the number of Rab paralogs inferred to have been present in the last common ancestor (LCA) of Archaeplastida. This is primarily thanks to the improved representation of the glaucophyte gene complement, which revealed the presence of two paralogs so far unknown from any archaeplastidial species – Rab14 and Rab34. Hence, the set of Rab paralogs that were apparently present in the archaeplastidial LCA now includes the following items: Rab1, Rab2, Rab5, Rab6, Rab7, Rab8, Rab11, Rab14, Rab18, Rab21, Rab23, Rab24, Rab28, Rab34, IFT27 and RTW. Identification of Rab14 and Rab34 in Archaeplastida is significant also for the general understanding of the evolution of the Rab family in eukaryotes, since it strengthens the notion that these two paralogs were present already in the LECA. This was previously uncertain, since with the hypotheses on the position of the root of the eukaryote phylogeny placed between Archaeplastida and remaining eukaryotes (see [37]), it was still possible that the absence of Rab14 and Rab34 in Archaeplastida is a primitive state.

For the remaining Rab paralogs thought to be present in the LECA [6], i.e. Rab4, Rab20, Rab22, Rab32A, Rab32B, Rab50, and RabTitan, no candidate orthologs were found in any archaeplastidial species systematically analyzed here. The most parsimonious scenario concerning the fate of these paralogs is that they had been lost before the archaeplastidial LCA (Fig. 1).

However, caution must be taken when inferences on early gene losses are made from a limited sampling of extant taxa. As a test case we decided to probe a possible presence of Rab14 and Rab34 orthologs in archaeplastidial species not systematically analyzed in this study. Using BLAST we screen transcriptomic data at NCBI and in the MMETSP database with the glaucophyte sequences as queries. Obvious Rab14 orthologs were found in a few transcriptome assemblies (from the charophyte *Nitella hyalina* and some angiosperms), but the sequences most likely represent contaminations from metazoan sources (data not shown; contaminations in the *Nitella* transcriptome assembly released by Finet et al. [38] were noticed also by others [39]). On the other hand, apparently authentic sequences similar to Rab34 were encountered in several “prasinophytes” and in the basally branching streptophyte alga *Chlorokybus atmophyticus*, and their assignment as Rab34 orthologs was confirmed by a phylogenetic analysis (Fig. S2, Tab. S1). Hence, despite the fact that Chloroplastida are represented in our set of systematically analyzed species by the highest number of genome sequences, the sampling was still insufficient to capture the actual repertoire of ancestral Rab paralogs retained in Chloroplastida.

Returning to the question of the set of Rab paralogs retained in Archaeplastida as a whole, we thus cannot exclude the possibility that future sampling of their phylogenetic diversity eventually reveals that at least some of the paralogs still unreported from this group (Rab4,

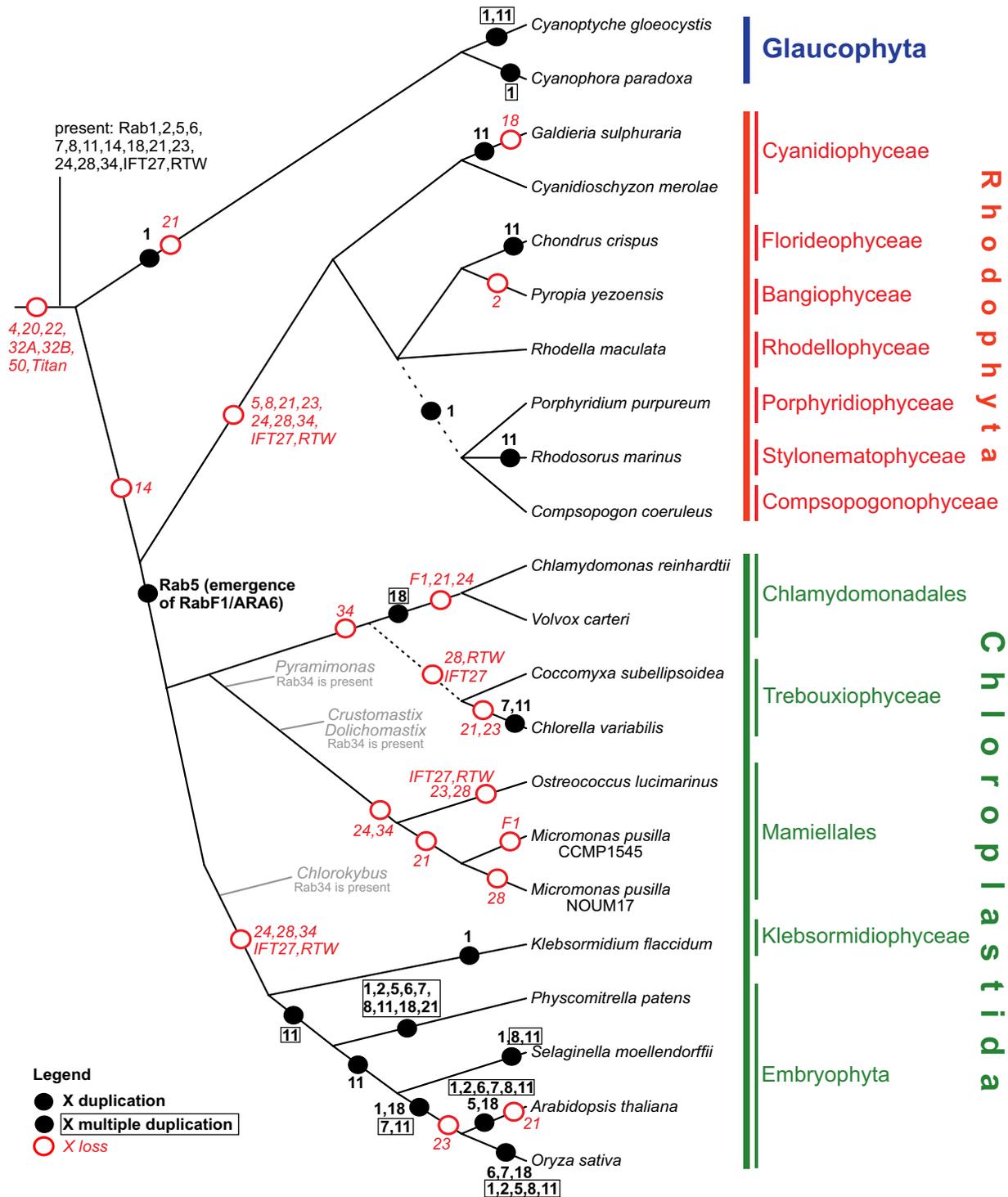


Fig. 1 Major events in the evolution of the Rab GTPase family in Archaeplastida. The events inferred from phylogenetic analyses of Rab sequences are mapped onto a schematic tree depicting phylogenetic relationships among the species analyzed in this study. The tree reflects the current understanding of the archaeplastidial phylogeny. Dashed lines indicate two branches that are uncertain. The clade comprising the red algal classes Porphyridiophyceae, Stylonematophyceae, and Compsopogonophyceae was suggested by some phylogenetic analyses [45], but statistical support was lacking. The monophyly of the green algal class Trebouxiophyceae, including *Coccomyxa subellipsoidea* and *Chlorella variabilis*, is generally accepted in the literature, but recent phylogenetic analyses based on plastid genome data cast doubt on this assumption [62].

Rab20, Rab22, Rab32A, Rab32B, Rab50, or RabTitan) do exist in some members. In addition, it was recently suggested that the group of Rab5-like sequences known from

some Chloroplastida and typified by the *A. thaliana* gene RabF1 (also called ARA6 [40]) may represent a paralog that originated before the divergence of Archaeplastida [41].

This group is characterized by the absence of a C-terminal geranylgeranylated tail, which is functionally replaced by an N-terminal extension modified by myristoylation and palmitoylation [40,41]. Hoepflinger et al. noticed the existence of Rab5-like proteins with the same modification in some members of the Alveolata group (e.g. the apicomplexan *Plasmodium falciparum*) [41]. However, their phylogenetic analysis failed to provide evidence for the common origin of the myristoylated/palmitoylated Rab5-like paralogs in Chloroplastida and Alveolata, and this negative result is consistent with our own, even broader phylogenies of the Rab5-related group (unpublished data). In fact, analyses of Rab sequences from diverse protist lineages indicate that such a replacement of the C-terminal geranylgeranylation by an N-terminal acylation has occurred convergently many times in the evolution of several different ancestral Rab paralogs (not only Rab5; unpublished data). Hence, we prefer a scenario in which the RabF1 (ARA6) group emerged by Rab5 duplication and modification in the stem lineage of Chloroplastida (Fig. 1).

Rab gene duplications have been very unevenly distributed across different Rab paralogs and organismal lineages in Archaeplastida

It is well established that gene duplication has been a very potent factor shaping the Rab family during the eukaryote evolution [5,6,12]. We used the expanded set of archaeplastidial Rab genes to make a fresh reconstruction of individual gene duplication events in the Rab family in this group (Fig. 1). Note that in case of organisms represented only by transcriptome data, we ignored highly similar variants of some Rab genes for which it is difficult to decide whether they represent allelic variants or recently duplicated genes.

The pattern of Rab duplications in Archaeplastida exhibits several notable features. The most conspicuous one is that duplication events have been very frequent in embryophytes, whereas their occurrence in other archaeplastidial taxa is much rarer. For some paralogs, including Rab2, Rab6, Rab8, and Rab21, no duplications have been recorded outside embryophytes. For Rab7, the only duplication outside embryophytes seem to be the one thought to give rise to the highly divergent Rab7-like gene in *C. variabilis* (see above). For Rab5, only the duplication at the base of Chloroplastida that resulted in the RabF1 group (see above) has been inferred outside embryophytes. Rab18, in addition to having multiplied in embryophytes, has triplicated in the lineage leading to chlamydomonadalean algae (*C. reinhardtii*, *V. carteri*; Fig. S1, Tab. 1). One of the chlamydomonadalean Rab18 in-paralogs (Rab18c; Fig. S1) is rather divergent and we speculate that it may have acquired a novel cellular function to regulate a membrane trafficking process unique for the respective group of algae.

The raison d'être for the expanded Rab gene sets in embryophytes may be the need of a multicellular body to finely regulate various, possibly tissue-specific, membrane trafficking pathways. At the same time it is clear that a substantial portion of the duplication events can be accounted for by the frequent occurrence of whole genome duplications (WGD) in the embryophyte evolution [42]. This is also obvious from the fact that the lycophyte *S. moellendorffii*, which does not seem to have an ancestor with a duplicated

genome [43], has a much smaller set of Rab genes than the moss *Physcomitrella patens*, *A. thaliana* and rice (Tab. 1), which all underwent WGD in their past (the angiosperms even multiple times [42]). However, it is beyond the scope of this paper to associate individual Rab gene duplications with the various inferred WGD events in plant evolution.

Two Rab paralogs – Rab1 and Rab11, stand out for their tendency to duplicate in a recurrent manner in Archaeplastida. To better understand the evolutionary history of these two paralogs, we conducted separate phylogenetic analyses of archaeplastidial Rab1 and Rab11 sequences using related paralogs (Rab8 and Rab2, respectively) as outgroups. Annotated ML trees are displayed as Fig. 2 and Fig. 3. Although the resolution of many relationships within the trees is poor, both trees enabled us to derive interesting conclusions.

In case of Rab1, independent duplications events appear to have occurred early in the evolution of three different archaeplastidial groups – glaucophytes, rhodophytes, and embryophytes, in all cases leading to an asymmetric evolution of the resulting in-paralogs (Fig. 2). While one in-paralog has stayed conservative in its sequence (note relatively short branches of the sequences denoted as “prototypical” in-paralogs in Fig. 2), the other is more divergent, suggesting that it has undergone neofunctionalization (“novel” in-paralogs in Fig. 2). In glaucophytes this duplication most likely occurred before the divergence of all known glaucophyte genera: although we included only sequences from *C. paradoxa* and *C. gloeocystis* in our trees, candidate orthologs of both the “prototypical” and the “novel” Rab1 in-paralogs are discernible in the transcriptome assembly of *G. wittrockiana* and among EST sequences from *G. nostochinearum* (data not shown). Interestingly, a PhyML-CAT tree inferred from the Rab1 and Rab8 sequences using the empirical profile mixture model C20 suggested that the somewhat problematic Rab1L gene from *C. gloeocystis* (see above) may actually be an extremely divergent additional member of the “novel” glaucophyte Rab1 in-paralog (data not shown).

In rhodophytes the Rab1 duplication is manifested in only three lineages, recently defined as classes separated from the traditionally circumscribed paraphyletic class Bangiophyceae [44]: Porphyridiophyceae, Stylonematophyceae, and Compsopogonophyceae (Fig. 2). The simplest explanation is that these three classes constitute a clade to the exclusion of other rhodophyte classes, and that the Rab1 duplication is exclusive for this putative clade. The interrelationships at the base of the rhodophyte phylogeny (except the firmly established basal position of Cyanidiophyceae and the sisterhood of Bangiophyceae sensu stricto and Florideophyceae) have not yet been resolved with confidence, but at least some phylogenies (e.g., [45]) do support the existence of the putative clade mentioned above.

Finally, in embryophytes an early Rab1 duplication (in addition to numerous ones specific for different terminal branches) can be traced back before the radiation of monocots and eudicots, but perhaps after the radiation of lycophytes (as there is no corresponding duplication in *S. moellendorffii*) and euphyllophytes (Fig. 2). Following the nomenclature of Rab genes in *A. thaliana* [46], the two in-paralogs are denoted RabD1 and RabD2, with the

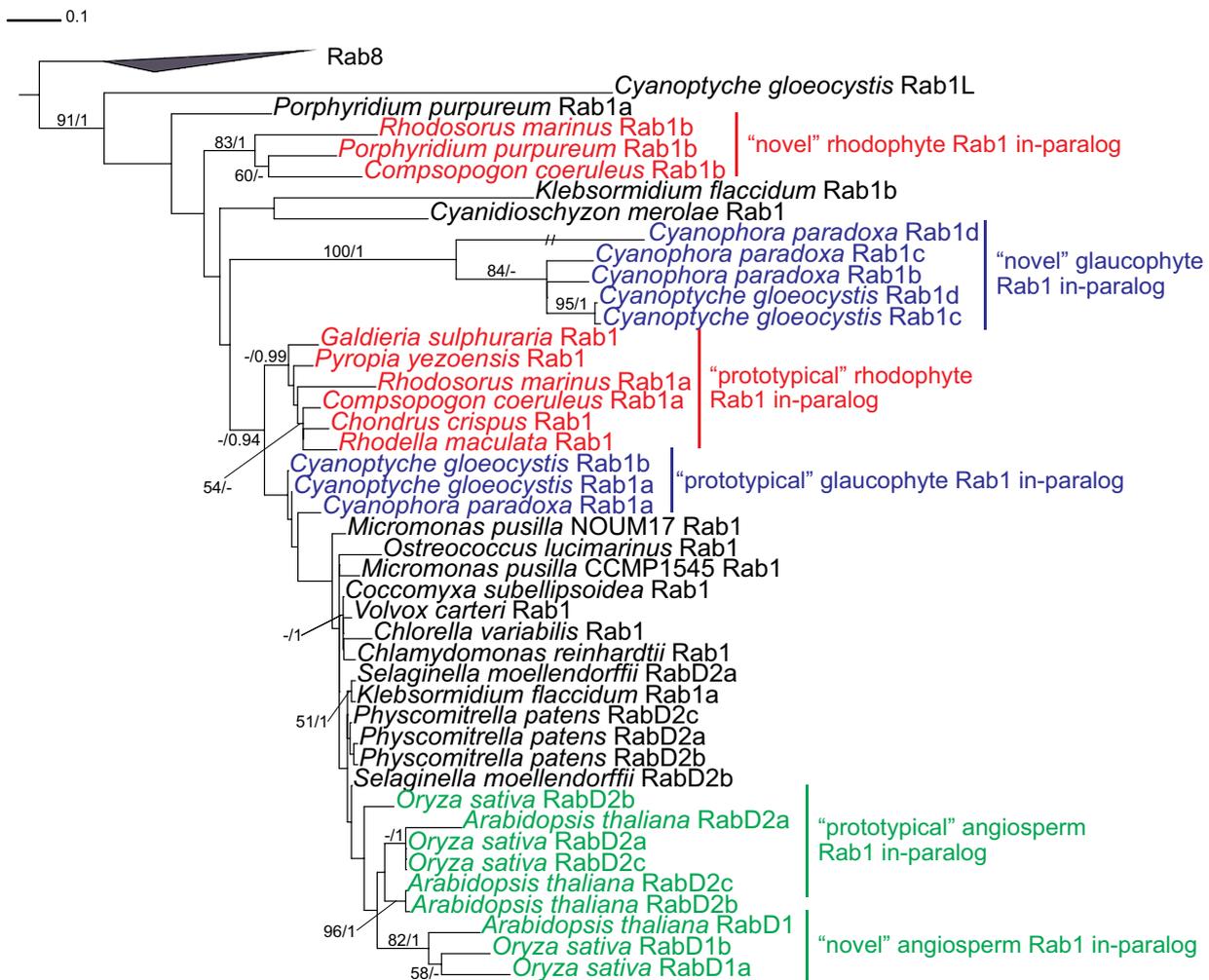


Fig. 2 Phylogenetic analysis of Rab1 genes in Archaeplastida. Portrayed is a maximum likelihood tree (RAXML, LG+Γ model) inferred from a multiple alignment of protein sequences representing the ancestral Rab1 paralog in 22 archaeplastidial species. Sequences representing the Rab8 paralog were used as an outgroup. Numbers at branches correspond to bootstrap support values (higher than 50) calculated using the rapid bootstrapping algorithm of the RAXML program. The bar on the top corresponds to the estimated number of substitutions per site. Groups of sequences of special interest are shown in color; their annotation is discussed in the text. Note that the sole Rab1 gene from *Cyanidioschyzon merolae* and the Rab1a gene from *Porphyridium purpureum* probably represent the “prototypical” rhodophyte Rab1 paralog and their failure to branch with the other rhodophyte sequences may be an artifact caused by their higher divergence. Sequences IDs are available in [Tab. S1](#).

former being the more divergent (“novel”). Although a more comprehensive analysis of Rab1 genes in embryophytes is needed to more precisely pinpoint the emergence of the two in-paralogs, a previously published phylogeny of the Rab1 subfamily including also data from the conifer *Pinus taeda* indicates that RabD1 and RabD2 split before the divergence of gymnosperms and angiosperms [47].

Similar to Rab1, duplications of Rab11 genes frequently exhibit a pattern suggesting neofunctionalization of one of the duplicated versions (Fig. 3). Thus, the glaucophyte *C. gloeocystis*, some rhodophytes (*G. sulphuraria*, *Rhodosorus marinus*, and *C. crispus*) and the green alga *C. variabilis* each harbour markedly divergent Rab11 in-paralogs in addition to more canonical ones. One of the *C. gloeocystis* paralogs is even so divergent that we denoted it Rab11L (Rab11-like). The more divergent Rab11 in-paralogs from *G. sulphuraria*

and *C. variabilis* (Rab11b in Fig. 3) cluster together, which might suggest that they arose from the same duplication event. However, statistical support for this cluster is low and the scenario mentioned above would be quite complex, as a number of independent losses of this in-paralog within rhodophytes and Chloroplastida would have to be assumed, so we prefer independent duplications as a more likely explanation.

Our analysis finally provided hints for an interpretation of the evolutionary history of Rab11 genes in embryophytes. A model compatible with the phylogenetic tree of Rab11 sequences and with the phylogenetic distribution of the different in-paralogs assumes the following successive events: (i) a triplication of Rab11 at the base of embryophytes (or at least before the divergence of mosses and vascular plants); (ii) an additional duplication at the base

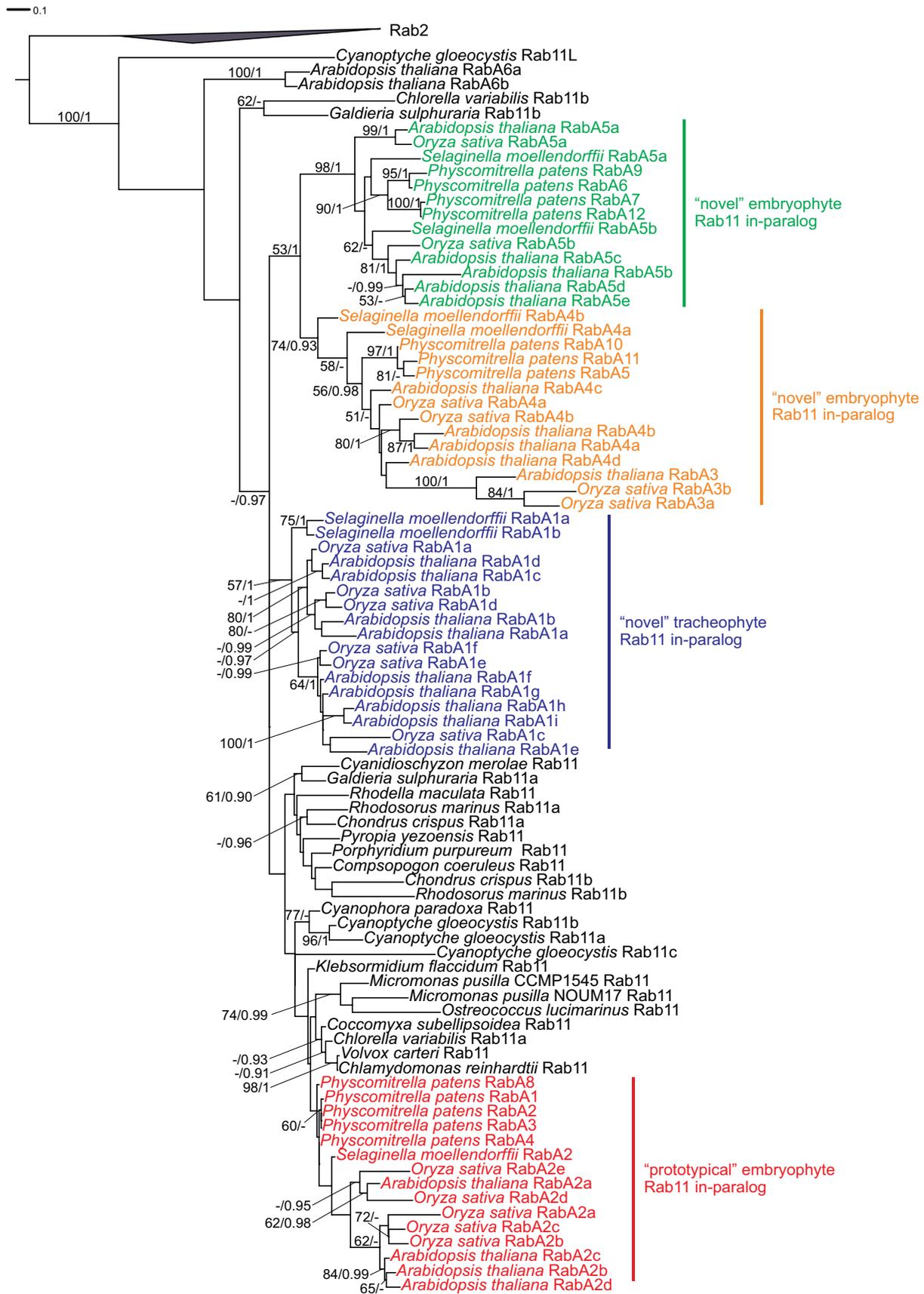


Fig. 3 Phylogenetic analysis of Rab11 genes in Archaeplastida. Portrayed is a maximum likelihood tree (RAxML, LG+ Γ model) inferred from a multiple alignment of protein sequences representing the ancestral Rab11 paralog in 22 archaeplastidial species. Sequences representing the Rab2 paralog were used as an outgroup. Numbers at branches correspond to bootstrap support values (higher than 50) calculated using the rapid bootstrapping algorithm of the RAxML program. The bar on the top corresponds to the estimated number of substitutions per site. Groups of sequences of special interest are shown in color; their annotation is discussed in the text. Sequences IDs are available in [Tab. S1](#).

of vascular plants (before the divergence of lycophytes and euphyllophytes); (iii) multiple additional duplications at the base of angiosperms and within the eudicot and monocot lineages (Fig. 1 and Fig. 3). From the perspective of Rab11 subgroups defined for *A. thaliana* (see [46]), the angiosperm RabA2 group represents the least derived (“prototypical”) in-paralog, the RabA1 group (apparently having specific orthologs in *S. moellendorffii* but not in *P. patens*), is a novelty of tracheophytes, whereas the RabA5 group and a combined RabA4 plus RabA3 group represent two separate in-paralogs that emerged from the “prototypical” paralog by duplications before the moss-tracheophyte divergence. The RabA6 group defined in *A. thaliana* probably represents a very divergent in-paralog specific for the *A. thaliana* lineage, but its exact origin and phylogenetic distribution need to be investigated further using a much broader sampling of angiosperm Rab11 diversity.

Gene loss has significantly sculpted the Rab family in Archaeplastida

Although gene duplications as a means of increasing the complexity of an organism may be viewed as the more interesting and significant events in the evolutionary history of a gene family, the relevance of an opposite process – simplification due to gene loss – may be even higher in particular cases. Concerning Rabs, there are ancestral paralogs that have been rarely duplicated, yet they were lost many times independently from different eukaryotic lineages. One such example is Rab24, whose scattered phylogenetic distribution in eukaryotes implies a high number of independent losses [48]. The same pattern is seen also within Archaeplastida, where Rab24 was lost at least four times independently, specifically in the rhodophyte stem lineage, in chlamydomonadalean green algae, in the Mamiellales lineage, and in streptophytes (Fig. 1).

The glaucophyte *C. paradoxa*, is an extant archaeplastidial species with the most complete set of ancestral Rab paralogs known so far. However, the genomic and transcriptomic data currently available for glaucophytes suggest that this lineage may have lost one Rab paralog certainly present in the last common archaeplastidial ancestor – Rab21 (note that a Rab21 sequence could be found in the MMETSP transcriptome assembly for *G. wittrockiana*, but it most likely represents a contamination, see “Materials and methods”). The loss of Rab21 from glaucophytes needs to be confirmed by further genome and transcriptome sampling, but it would not be at all unprecedented, as Rab21 appears to have been lost many times independently in eukaryotes [6]. Within Archaeplastida, Rab21 is missing not only from glaucophytes, but also from rhodophytes and some lineages of Chloroplastida, in our taxon sampling represented by chlamydomonadalean algae (both *C. reinhardtii* and *Volvox carteri*), *Chlorella variabilis*, both *Micromonas* strains, and *A. thaliana* (Tab. 1). Hence, at least six independent losses of Rab21 need to be assumed to explain the current distribution of Rab21 genes in Archaeplastida (Fig. 1).

One of the earliest reductions of the Rab complement within Archaeplastida was probably that affecting Rab14. In the most parsimonious scenario of archaeplastidial phylogeny, which assumes that glaucophytes are sister to a clade comprising Rhodophyta and Chloroplastida [49], a single

loss of the Rab14 gene in the exclusive ancestor of the latter two lineages would explain its phylogenetic distribution. However, the branching order of the three archaeplastidial lineages has not yet been resolved with confidence (see [50]), so independent losses of Rab14 in Rhodophyta and Chloroplastida cannot be excluded.

A branch with the highest number of inferred Rab losses within Archaeplastida is the stem lineage of Rhodophyta (Fig. 1). Specifically, Rab5, Rab8, Rab21, Rab23, Rab24, Rab28, Rab34, IFT27, and RTW appear to have been lost from rhodophytes before the radiation of modern forms. In combination with losses that occurred earlier in the archaeplastidial evolution, modern rhodophytes have retained only six out of the 23 paralogs of Rab and Rab-like proteins inferred as ancestral for eukaryotes [6]. Most rhodophytes thus possess genes representing Rab1, Rab2, Rab6, Rab7, Rab11, and Rab18 paralogs (Tab. 1), but some red algal species have reduced this set even further. *Galdieria sulphuraria* apparently lacks Rab18 (this paralog is also missing from a related species, *Galdieria phlegrea*; data not shown) and *Pyropia yezoensis* seems to lack Rab2 (no Rab2 could be identified also in the transcriptome data for a related species, *Pyropia haitanensis*). With only five Rab genes, *P. yezoensis*, a multicellular red alga, exhibits the minimal number of Rab paralogs ever recorded for a eukaryotic cell. According to our knowledge, the same number is found only in the extremely reduced and divergent parasitic group of Microsporidia (unpublished data).

The configuration of Rab paralogs currently inferred for the LCA of Chloroplastida was this one: Rab1, Rab2, Rab5 (the “prototypical” version, in *A. thaliana* called RabF2), RabF1 (=ARA6), Rab6, Rab7, Rab8, Rab11, Rab18, Rab21, Rab23, Rab24, Rab28, Rab34, IFT27 and RTW. While Rab1, Rab2, Rab5, Rab6, Rab7, Rab8, Rab11, and Rab18 are always represented by at least one gene in every chloroplastidial species studied so far, the remaining ancestrally present paralogs are completely missing from some extant taxa. Losses affecting Rab21 and Rab24 were already mentioned above. The unexpected finding of Rab34 in “prasinophytes” (*Dolichomastix*, *Crustomastix*, *Pyramimonas*) specifically related to the Mamiellales [51,52], which are devoid of Rab34 genes, and in *C. atmophyticus*, which represents a streptophyte lineage basal to Rab34-lacking streptophytes covered by this study (Klebsormidiophyceae, embryophytes), necessitates at least three independent losses of Rab34 in Chloroplastida (Fig. 1). The “green” lineage also seems to easily dispense with Rab28 (lost at least four times) and with Rab23, IFT27, and RTW (each lost at least three times).

Finally, we could infer at least two losses of the RabF1 paralog, specifically from *M. pusilla* CCMP1545 and from chlamydomonadalean algae (*C. reinhardtii* and *V. carteri*). It was previously claimed that RabF1 is missing also from *O. lucimarinus* [41], but we found in this species an obvious ortholog that also possess the characteristic N-terminal extension (Tab. S1 and Fig. S3). The failure of Hoepflinger et al. [41] to detect this ortholog may relate to the fact that the respective gene model is incomplete in the official genome annotation release for this species. We were also able to correct the gene model for the protein sequence EFN55859 from *Chlorella variabilis* that was noticed by Hoepflinger

et al. [41] as lacking the N-terminal region; the revised sequence has the N-terminal extension typical for all RabF1 proteins (Fig. S3).

Cell biological implications of the varying Rab complement in different archaeplastidal taxa

The pattern of Rab gene duplications and losses in Archaeplastida described above is important per se, but its main value is primarily in that it can serve as a framework for understanding the evolution of archaeplastidal cells and their functionalities. Below we discuss some implications that we consider most significant.

One of the dominant patterns seems to be the recurrent simplification of cellular complexity due to the loss of particular Rab paralogs. A massive series of losses appears to have occurred already during the emergence of Archaeplastida as a group (Fig. 1), which may imply substantial reduction of the complexity of membrane trafficking processes of the original eukaryotic host cell concomitant with the evolution of a plastid from a cyanobacterial endosymbiont. The loss of Rab14, inferred to have happened in an exclusive ancestor of Rhodophyta and Chloroplastida, may perhaps be viewed as a continuation of this trend, although it is not at all clear why it has been preserved in glaucophytes, as little specific information on membrane trafficking and the endomembrane system in general is available for this group. Indeed, no obvious functional explanation is available to account for the evolutionary younger loss events that affected many other paralogs. This is not only due to very poor knowledge of the cell biology of most archaeplastidal lineages, but also because for many Rab paralogs (e.g. Rab21, Rab24, Rab28, or Rab34) only restricted functional information is available in general.

The lack of a significant correlation between differences or similarities in the configuration of Rab genes and differences or similarities at the organismal level in different organisms can be documented by several notable examples. One is provided by the green algae *C. reinhardtii* and *V. carteri*. The former is a single-cell organism with isogamous mating, while the latter is a multicellular organism with differentiated somatic and germ lines and with oogamous sexual reproduction [53]. However, their sets of Rab genes are exactly the same (Tab. 1), suggesting that expansions of the Rab family is not a sine qua non for achieving a higher morphological complexity. On the other hand, Rab21 has been kept in monocots and some basal “dicots”, but is missing from *A. thaliana* and perhaps all other “core” eudicots (Tab. 1 and data not shown). What is so fundamentally different between cells of the different angiosperms that this Rab paralog, and presumably a specific cellular process regulated by Rab21 (most likely in the endocytic pathway [54]), could have been lost from the *A. thaliana* lineage? An even more extreme case is represented by the pair of organisms nominally representing the same species, *M. pusilla* CCMP1545 and *M. pusilla* NOUM17 (sometimes referred to as *Micromonas* sp. RCC299, e.g. in [41]). Global analyses of the gene complements of these two morphologically indistinguishable strains revealed hundreds of genes specific for one or the other strain [55], and this difference pertains also to the Rab family, as RabF1 is missing from *M. pusilla*

CCMP1545 and Rab28 is missing from *M. pusilla* NOUM17 (Tab. 1).

However, for a group of paralogs there is an emerging cellular correlate for their presence/absence pattern. At least three Rab and Rab-like proteins – Rab23, IFT27, and RTW (Rab12) – are functionally associated with the flagellar apparatus, so they are generally missing from species unable to make flagella or cilia [9,10,56]. Hence, their absence from red algae, angiosperms, *O. lucimarinus*, and *C. variabilis*, which all lack flagella (or where a typical flagellum is at least unknown, see [57]), is not at all surprising. However, IFT27 and RTW may get lost independently of a flagellum loss, since both paralogs are absent from the streptophytes *Klebsormidium flaccidum*, *P. patens*, and *S. moellendorffii*, although they do exhibit flagellated reproductive cells [58]. It is possible that since these flagellated stages are only transient, some original flagellum-associated functions may have been lost in these groups, which is consistent with the fact that many other conserved flagellar proteins are missing in land plants with flagellated stages [59]. The trebouxiophyte alga *C. subellipsoidea* is notable for possessing Rab23 (but not IFT27 and RTW; Tab. 1), although to our knowledge, no flagellated stages have been described for the genus *Coccoyomyxa*. One possible explanation is that a cryptic, rarely occurring stage with a (possibly reduced) flagellum does exist in the life cycle of *C. subellipsoidea*, but has not been documented yet (as seems to be the case for another trebouxiophyte, *C. variabilis*, see [57]). Alternatively, Rab23 might have kept or evolved a new, flagellum-independent, function in *Coccoyomyxa*.

In light of the previous discussion, the massive reduction of the Rab gene complement in the stem rhodophyte lineage (Fig. 1) is partially explained as a consequence of the flagellum loss in red algae. However, most of the remaining paralogs lost in rhodophytes perhaps lack a specific functional connection to the flagellum, so their absence from red algal cells is indicative of a general simplification of their endomembrane system compared to other eukaryotes. Quite unusual is the absence of Rab8, which occurs in most eukaryotes, perhaps functioning as a general exocytotic factor [6]. But even more striking is the lack of Rab5, an essential component of the endocytic machinery associated with early (recycling) endosomes [60]. Rab5 is nearly universal in eukaryotes and the only known group outside rhodophytes devoid of an apparent Rab5 ortholog are the extremely divergent diplomonads (*Giardia intestinalis*, *Spironucleus salmonicida*; [6] and unpublished data). Interestingly, other components of the canonical endocytic machinery are missing from at least some rhodophytes. For example, *C. crispus* (a multicellular red alga) lacks the AP-2 adaptor complex and endocytic Qc-SNARE proteins [22]. Future systematic comparative analyses of rhodophyte genomes combined with cell biological experiments will hopefully clarify whether and how endocytosis takes place in red algal cells.

Conclusions

Thanks to a considerably improved sampling of the phylogenetic diversity of Archaeplastida we were able to draw

a more comprehensive picture of the intricate evolutionary history of the Rab family in this significant segment of the eukaryotic tree of life. However, a recurrent theme of the previous discussion was that the data from a much higher number of archaeplastidial lineages are needed to provide more robust answers to many important questions, such as the actual composition of Rab gene complements at different nodes of the archaeplastidial phylogeny. Some

lineages may be especially informative and should become the prime targets for next genome sequencing projects. For example, some “prasinophytes” have retained the ability to phagocytose prey [61] and it is possible that their membrane trafficking machinery exhibits some primitive features lost in other Archaeplastida. We are sure that there still are exciting unexpected aspects of the Rab family (not only) in Archaeplastida to be revealed.

Acknowledgments

We thank Aleš Horák and Vladimír Klimeš for their technical help with phylogenetic analyses and Vyacheslav Yurchenko for critical comments on the manuscript. This study was supported by the Czech Science Foundation (grant No. 13-24983S), the project CZ.1.05/2.1.00/03.0100 (IET) financed by the structural funds of the EU, the project LO1208 of the National Feasibility Programme I of the Czech Republic, and the internal student grant SGS26/PfF/2014-2015.

Authors' contributions

The following declarations about authors' contributions to the research have been made: carried out the bioinformatic analyses, prepared figures and contributed to the final text: RP; conceived the study, contributed to the bioinformatic analyses and drafted the manuscript: ME.

Competing interests

No competing interests have been declared.

Supplementary material

The following supplementary material for this article is available online at <http://pbsociety.org.pl/journals/index.php/asbp/rt/suppFiles/asbp.2014.052/0>:

1. Tab. S1: list of sequences analyzed.
2. Figs. S1–S3.
3. A text file with revised or newly predicted protein sequences.

References

1. Stenmark H. Rab GTPases as coordinators of vesicle traffic. *Nat Rev Mol Cell Biol.* 2009;10(8):513–525. <http://dx.doi.org/10.1038/nrm2728>
2. Woollard AA, Moore I. The functions of Rab GTPases in plant membrane traffic. *Curr Opin Plant Biol.* 2008;11(6):610–619. <http://dx.doi.org/10.1016/j.pbi.2008.09.010>
3. Brighouse A, Dacks JB, Field MC. Rab protein evolution and the history of the eukaryotic endomembrane system. *Cell Mol Life Sci.* 2010;67(20):3449–3465. <http://dx.doi.org/10.1007/s00018-010-0436-1>
4. Briguglio JS, Turkewitz AP. *Tetrahymena thermophila*: a divergent perspective on membrane traffic. *J Exp Zool B Mol Dev Evol.* 2014;322(7):500–516. <http://dx.doi.org/10.1002/jez.b.22564>
5. Klöpffer TH, Kienle N, Fasshauer D, Munro S. Untangling the evolution of Rab G proteins: implications of a comprehensive genomic analysis. *BMC Biol.* 2012;10:71. <http://dx.doi.org/10.1186/1741-7007-10-71>
6. Elias M, Brighouse A, Gabernet-Castello C, Field MC, Dacks JB. Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. *J Cell Sci.* 2012;125(pt 10):2500–2508. <http://dx.doi.org/10.1242/jcs.101378>
7. Koonin EV. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* 2010;11(5):209. <http://dx.doi.org/10.1186/gb-2010-11-5-209>
8. Koumandou VL, Wickstead B, Ginger ML, van der Giezen M, Dacks JB, Field MC. Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit Rev Biochem Mol Biol.* 2013;48(4):373–396. <http://dx.doi.org/10.3109/10409238.2013.821444>
9. Lo JC, Jamsai D, O'Connor AE, Borg C, Clark BJ, Whisstock JC, et al. RAB-like 2 has an essential role in male fertility, sperm intra-flagellar transport, and tail assembly. *PLoS Genet.* 2012;8(10):e1002969. <http://dx.doi.org/10.1371/journal.pgen.1002969>
10. Huet D, Blisnick T, Perrot S, Bastin P. The GTPase IFT27 is involved

in both anterograde and retrograde intraflagellar transport. *Elife.* 2014;3:e02419. <http://dx.doi.org/10.7554/eLife.02419>

11. Burki F. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb Perspect Biol.* 2014;6(5):a016147. <http://dx.doi.org/10.1101/cshperspect.a016147>
12. Diekmann Y, Seixas E, Gouw M, Tavares-Cadete F, Seabra MC, Pereira-Leal JB. Thousands of rab GTPases for the cell biologist. *PLoS Comput Biol.* 2011;7(10):e1002217. <http://dx.doi.org/10.1371/journal.pcbi.1002217>
13. del Campo J, Sieracki ME, Molestina R, Keeling P, Massana R, Ruiz-Trillo I. The others: our biased perspective of eukaryotic genomes. *Trends Ecol Evol.* 2014;29(5):252–259. <http://dx.doi.org/10.1016/j.tree.2014.03.006>
14. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 2014;12(6):e1001889. <http://dx.doi.org/10.1371/journal.pbio.1001889>
15. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol.* 2005;52(5):399–451. <http://dx.doi.org/10.1111/j.1550-7408.2005.00053.x>
16. Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, et al. The revised classification of eukaryotes. *J Eukaryot Microbiol.* 2012;59(5):429–493. <http://dx.doi.org/10.1111/j.1550-7408.2012.00644.x>
17. de Clerck O, Bogaert KA, Leliaert F. Diversity and evolution of algae: primary endosymbiosis. *Adv Bot Res.* 2012;64:55–86. <http://dx.doi.org/10.1016/B978-0-12-391499-6.00002-5>
18. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups”. *Proc Natl Acad Sci USA.* 2009;106(10):3859–3864. <http://dx.doi.org/10.1073/pnas.0807880106>
19. Yabuki A, Kamikawa R, Ishikawa SA, Kolisko M, Kim E, Tanabe AS, et al. *Palpitomonas bilix* represents a basal cryptist lineage: insight into the character evolution in Cryptista. *Sci Rep.* 2014;4:4641. <http://dx.doi.org/10.1038/srep04641>
20. Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, et al. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature.* 2004;428(6983):653–657. <http://dx.doi.org/10.1038/nature02398>
21. Bhattacharya D, Price DC, Chan CX, Gross J, Steiner JM, Löffelhardt. Analysis of the genome of *Cyanophora paradoxa*: an algal model for understanding primary endosymbiosis. In: Löffelhardt W, editor. *Endosymbiosis*. Vienna: Springer; 2014. p. 135–148. http://dx.doi.org/10.1007/978-3-7091-1303-5_7
22. Collén J, Porcel B, Carré W, Ball SG, Chaparro C, Tonon T, et al. Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc Natl Acad Sci USA.* 2013;110(13):5247–5252. <http://dx.doi.org/10.1073/pnas.1221259110>
23. Bhattacharya D, Price DC, Chan CX, Qiu H, Rose N, Ball S, et al. Genome of the red alga *Porphyridium purpureum*. *Nat Commun.* 2013;4:1941. <http://dx.doi.org/10.1038/ncomms2931>
24. Schönknecht G, Chen WH, Ternes CM, Barbier GG, Shrestha RP, Stanke M, et al. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science.* 2013;339(6124):1207–1210. <http://dx.doi.org/10.1126/science.1231707>

25. Nakamura Y, Sasaki N, Kobayashi M, Ojima N, Yasuike M, Shigenobu Y, et al. The first symbiont-free genome sequence of marine red alga, *Susabi-nori* (*Pyropia yezoensis*). PLoS One. 2013;8(3):e57122. <http://dx.doi.org/10.1371/journal.pone.0057122>
26. Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber AP, et al. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science. 2012;335(6070):843–847. <http://dx.doi.org/10.1126/science.1213561>
27. Saito C, Ueda T. Functions of RAB and SNARE proteins in plant life. Int Rev Cell Mol Biol. 2009;274:183–233. [http://dx.doi.org/10.1016/S1937-6448\(08\)02004-2](http://dx.doi.org/10.1016/S1937-6448(08)02004-2)
28. Zhang Z, Hill DR, Sylvester AW. Diversification of the RAB guanosine triphosphatase family in dicots and monocots. J Integr Plant Biol. 2007;49(8):1129–1141. <http://dx.doi.org/10.1111/j.1672-9072.2007.00520.x>
29. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, et al. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science. 2008;319(5859):64–69. <http://dx.doi.org/10.1126/science.1150646>
30. Huber H, Beyser K, Fabry S. Small G proteins of two green algae are localized to exocytic compartments and to flagella. Plant Mol Biol. 1996;31(2):279–293. <http://dx.doi.org/10.1007/BF00021790>
31. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>
32. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–780. <http://dx.doi.org/10.1093/molbev/mst010>
33. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–1313. <http://dx.doi.org/10.1093/bioinformatics/btu033>
34. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: Proceedings of the gateway computing environments workshop (GCE). New Orleans, LA; 2010. p. 1–8. <http://dx.doi.org/10.1109/GCE.2010.5676129>
35. Quang LS, Gascuel O, Lartillot N. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics. 2008;24(20):2317–2323. <http://dx.doi.org/10.1093/bioinformatics/btn445>
36. Letunic I, Bork P. Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 2011;39(web server issue):W475–W478. <http://dx.doi.org/10.1093/nar/gkr201>
37. Rogozin IB, Basu MK, Csürös M, Koonin EV. Analysis of rare genomic changes does not support the unikont-bikont phylogeny and suggests cyanobacterial symbiosis as the point of primary radiation of eukaryotes. Genome Biol Evol. 2009;1:99–113. <http://dx.doi.org/10.1093/gbe/evp011>
38. Finet C, Timme RE, Delwiche CF, Marlétaz F. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. Curr Biol. 2010;20(24):2217–2222. <http://dx.doi.org/10.1016/j.cub.2010.11.035>
39. Laurin-Lemay S, Brinkmann H, Philippe H. Origin of land plants revisited in the light of sequence contamination and missing data. Curr Biol. 2012;22(15):R593–R594. <http://dx.doi.org/10.1016/j.cub.2012.06.013>
40. Ueda T, Yamaguchi M, Uchimiya H, Nakano A. Ara6, a plant-unique novel type Rab GTPase, functions in the endocytic pathway of *Arabidopsis thaliana*. EMBO J. 2001;20(17):4730–4741. <http://dx.doi.org/10.1093/emboj/20.17.4730>
41. Hoepflinger MC, Geretschlaeger A, Sommer A, Hoefftberger M, Nishiyama T, Sakayama H, et al. Molecular and biochemical analysis of the first ARA6 homologue, a RAB5 GTPase, from green algae. J Exp Bot. 2013;64(18):5553–5568. <http://dx.doi.org/10.1093/jxb/ert322>
42. Vanneste K, Maere S, van de Peer Y. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philos Trans R Soc Lond B Biol Sci. 2014;369(1648). <http://dx.doi.org/10.1098/rstb.2013.0353>
43. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, et al. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science. 2011;332(6032):960–963. <http://dx.doi.org/10.1126/science.1203810>
44. Yoon HS, Müller KM, Sheath RG, Ott FD, Bhattacharya D. Defining the major lineages of red algae (Rhodophyta). J Phycol. 2006;42(2):482–492. <http://dx.doi.org/10.1111/j.1529-8817.2006.00210.x>
45. Verbruggen H, Maggs CA, Saunders GW, Le Gall L, Yoon HS, de Clerck O. Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. BMC Evol Biol. 2010;10:16. <http://dx.doi.org/10.1186/1471-2148-10-16>
46. Rutherford S, Moore I. The *Arabidopsis* Rab GTPase family: another enigma variation. Curr Opin Plant Biol. 2002;5(6):518–528. [http://dx.doi.org/10.1016/S1369-5266\(02\)00307-2](http://dx.doi.org/10.1016/S1369-5266(02)00307-2)
47. Elias M, Patron NJ, Keeling PJ. The RAB family GTPase Rab1A from *Plasmodium falciparum* defines a unique paralog shared by Chromalveolates and Rhizaria. J Eukaryot Microbiol. 2009;56(4):348–356. <http://dx.doi.org/10.1111/j.1550-7408.2009.00408.x>
48. Elias M. Patterns and processes in the evolution of the eukaryotic endomembrane system. Mol Membr Biol. 2010;27(8):469–489. <http://dx.doi.org/10.3109/09687688.2010.521201>
49. Reyes-Prieto A, Bhattacharya D. Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within Plantae. Mol Biol Evol. 2007;24(11):2358–2361. <http://dx.doi.org/10.1093/molbev/msm186>
50. Deschamps P, Moreira D. Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. Mol Biol Evol. 2009;26(12):2745–2753. <http://dx.doi.org/10.1093/molbev/msp189>
51. Marin B, Melkonian M. Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. Protist. 2010;161(2):304–336. <http://dx.doi.org/10.1016/j.protis.2009.10.002>
52. Lemieux C, Otis C, Turmel M. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. BMC Genomics. 2014;15:857. <http://dx.doi.org/10.1186/1471-2164-15-857>
53. Umen JG. Green algae and the origins of multicellularity in the plant kingdom. Cold Spring Harb Perspect Biol. 2014;6(11). <http://dx.doi.org/10.1101/cshperspect.a016170>
54. Ali M, Leung KF, Field MC. The ancient small GTPase Rab21 functions in intermediate endocytic steps in trypanosomes. Eukaryot Cell. 2014;13(2):304–319. <http://dx.doi.org/10.1128/EC.00269-13>
55. Worden AZ, Lee JH, Mock T, Rouzé P, Simmons MP, Aerts AL, et al. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. Science. 2009;324(5924):268–272. <http://dx.doi.org/10.1126/science.1167222>
56. Lumb JH, Field MC. Rab23 is a flagellar protein in *Trypanosoma brucei*. BMC Res Notes. 2011;4:190. <http://dx.doi.org/10.1186/1756-0500-4-190>
57. Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, et al. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell. 2010;22(9):2943–2955. <http://dx.doi.org/10.1105/tpc.110.076406>
58. Lewis LA, McCourt RM. Green algae and the origin of land plants. Am J Bot. 2004;91(10):1535–1556. <http://dx.doi.org/10.3732/ajb.91.10.1535>
59. Hodges ME, Wickstead B, Gull K, Langdale JA. The evolution of land plant cilia. New Phytol. 2012;195(3):526–540. <http://dx.doi.org/10.1111/j.1469-8137.2012.04197.x>
60. Woodman PG. Biogenesis of the sorting endosome: the role of Rab5. Traffic. 2000;1(9):695–701. <http://dx.doi.org/10.1034/j.1600-0854.2000.010902.x>
61. Maruyama S, Kim E. A modern descendant of early green algal phagotrophs. Curr Biol. 2013;23(12):1081–1084. <http://dx.doi.org/10.1016/j.cub.2013.04.063>
62. Lemieux C, Otis C, Turmel M. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. BMC Evol Biol. 2014;14:211. <http://dx.doi.org/10.1186/s12862-014-0211-2>