Phylogenomics of neglected flagellated protists supports a revised eukaryotic tree of life

Guifré Torruella[1,2]*, Luis Javier Galindo[1,3]*, David Moreira[1] and Purificación López-García[1]

[1] Ecologie Systématique Evolution, CNRS, Université Paris-Saclay, AgroParisTech, Gif-sur-Yvette, France
[2] Institut de Biologia Evolutiva, UPF-CSIC, Barcelona, Catalonia, Spain.
[3] Department of Biology, University of Oxford, Oxford, United Kingdom.

* These authors equally contributed to this work.

For correspondence: guifftc@gmail.com, puri.lopez@universite-paris-saclay.fr

Eukaryotes radiated from their last common ancestor, diversifying into several supergroups with unresolved deep evolutionary connections. Heterotrophic flagellates, often branching deeply in phylogenetic trees, are arguably the most diverse eukaryotes. However, many of them remain undersampled and/or *incertae sedis*. Here, we conducted comprehensive phylogenomics analyses with an expanded taxon sampling of early-branching protists including 22 newly sequenced transcriptomes (apusomonads, ancyromonads, *Meteora*). They support the monophyly of Opimoda, one of the largest eukaryotic supergroups, with CRuMs being sister to the Amorphea (amoebozoans, breviates, apusomonads, and opisthokonts – including animals and fungi–), and the ancyromonads+malawimonads clade. By mapping traits onto this phylogenetic framework, we infer a biflagellate opimodan ancestor with an excavate-like feeding groove. Breviates and apusomonads retained the ancestral biflagellate state. Other Amorphea lost one or both flagella, enabling the evolution of amoeboid shapes, novel feeding modes, and palintomic cell division resulting in multinucleated cells, which likely facilitated the subsequent evolution of fungal and metazoan multicellularity.

Keywords:
ancyromonad; apusomonad; free-living flagellate; *Meteora*; phylogenomics; transcriptomics; RNAseq

Eukaryotes evolved from their prokaryotic ancestors in the early Proterozoic[1,2] and rapidly diversified into a multitude of lineages from an already complex last eukaryotic common ancestor (LECA) that possessed all typical traits of extant eukaryotes[3]. The majority of eukaryotes comprise hugely diverse unicellular, mostly flagellated and heterotrophic, protists[4,5], although much of this diversity remains undescribed, as suggested by environmental studies[6,7]. Several eukaryotic lineages developed complex multicellularity, such as animals, fungi, plants and brown algae, and/or acquired photosynthesis though the endosymbiosis of a cyanobacterium or their algal derivatives[4,5]. Heterotrophic protists are generally phagotrophic, preying on bacteria or other protists[4]. Other heterotrophic eukaryotes such as fungi and oomycetes became osmotrophic, feeding from the absorption of extracellularly digested organic molecules; several of them evolved into parasites[4,8]. With their diversity of trophic modes, protists play crucial roles in ecosystem's networks and the carbon cycle[9].

Eukaryotic lineages are currently known to form several large supergroups but their deep phylogenetic relationships remain unresolved[5], as is the root of the eukaryotic tree[10]. These uncertainties can be explained by an inherent phylogenetic signal limitation due to a rapid diversification in a short time span, methodological artefacts linked, among others, to heterogeneous evolutionary rates, hidden paralogy and horizontal gene transfer, as well as to patchy sampling across the eukaryotic diversity[11,12]. Nonetheless, significant progress has been achieved in recent years owing to the improvement of phylogenetic methods and to the generation of genomic and/or transcriptomic data from poorly studied or newly identified eukaryotes[5,13]. The addition of this diversity has increasingly enlarged, reshaped and consolidated eukaryotic supergroups. For instance, the long supported Opisthokonta clade, comprising metazoans, fungi and their unicellular relatives (e.g. choanoflagellates, ichthyosporeans, nucleariids, aphelids, rozellids)[14,15] was found to form a robust supergroup, the Amorphea, with classical amoeba (Amoebozoa) plus two groups of flagellated protists previously considered *incertae sedis*, the amoeboflagellated Breviatea and the biflagellated Apusomonadida[16]. Opisthokonta and Amoebozoa were previously thought to share a uniflagellated ancestor[17], and the root of the eukaryotic tree to lie between this Unikonta clade and the rest of eukaryotes (Bikonta)[18]. The latter clustered several lineages sharing a biflagellate ancestor, including Archaeplastida (glaucophytes, red algae, green algae and plants) and SAR, composed of Stramenopiles (e.g. oomycetes, diatoms, brown algae), Alveolata (e.g. dinoflagellates, ciliates) and Rhizaria (e.g. radiolarians, cercozoans)[18]. However, the nested position of apusomonads (represented at the time by the genome of a single species) within unikonts rejected the idea of a uniflagellated ancestor, and the new names Opimoda and Diphoda were proposed for those redefined major splits[19]. Archaeplastida and SAR, together with other lineages of unicellular protists, many photosynthetic (cryptophytes, haptophytes), and their allies (e.g. telonemids, centrohelids, *Palpitomonas*) often cluster in a large supergroup called Diaphoretickes[20,21]. At the same time, another group of poorly known protists called CRuMs (collodictyonids, rigifilids, and *Mantamonas*) appears to branch as sister to the Amorphea[5,22]. To complete the global picture, Excavata, once thought to be a major eukaryotic supergroup characterized by the shared phenotypic feature of a ventral feeding groove, lacks molecular phylogenetic support and is no longer considered monophyletic, being split into Discoba (e.g. euglenids, jakobids), Malawimonadida[20] and Metamonada (e.g. trichomonads, *Trimastix*, *Carpediomonas*)[5,19].

If the inclusion of phylogenomic data for many newly described flagellates has confirmed some eukaryotic supergroups, many lineages of heterotrophic flagellates are yet to find a home in the eukaryotic tree and remain *incertae sedis*. This is the case of ancyromonads[23] or malawimonads[20], so far not clearly related to any eukaryotic supergroup. Other newly described phylum-level lineages such as Provora[24] (including *Ancoracysta*[25]) Hemimastigophora[21] and *Meteora*[26,27] also branch deeply in the eukaryotic tree, although showing some affinity with the Diaphoretickes. In general, many of these difficult-to-place lineages are represented by one or few representative species available in culture. However, they are most likely undersampled and encompass a wider within-group diversity. This can be illustrated by apusomonads[28], for which the use of specific 18S rRNA gene primers revealed a broad diversity of these gliding flagellates in freshwater and marine benthos across the globe[29]. Getting these tiny predators in culture for further study is challenging for several reasons. They depend on specific bacterial or eukaryotic prey which, in turn, are not necessarily easy to identify and maintain. Also, being higher in the trophic chain, they are in lower abundances in their native ecosystems and, hence, underrepresented in metagenomic data. Likewise, metagenomic studies rarely target sediment samples, where many of these organisms roam. Sediment-dwelling heterotrophic flagellates therefore remain largely understudied despite their long appreciated ecological role as grazers[30], and their incorporation to global phylogenetic analyses can help resolving the eukaryotic tree[21,24,27]. Here, we generated almost complete transcriptomes of diverse members of the Apusomonadida (14 species) and Ancyromonadida (7 species) that we recently cultured from different marine and freshwater environments[31,32], as well as the type species of *Meteora sporadica*, which we cultured from Mediterranean samples[26]. Our phylogenomic analyses with an expanded dataset of early-branching heterotrophic flagellates resolve the internal evolutionary relationships within Apusomonadida and Ancyromonadida and shed new light onto the eukaryotic tree, reinforcing the Opimoda monophyly. The distribution of complex phenotypic traits in this phylogenomic framework, notably the presence and number of flagella, the occurrence of pseudopodia or the karyokinesis-cytokinesis coupling, helps inferring ancestral states and crucial steps that conditioned lifestyles and subsequent major evolutionary trends along the natural history of eukaryotes.

Results and Discussion

High-quality curated transcriptomes for phylogenomic analysis

We considerably enriched the number of available transcriptomic datasets for members of the Apusomonadida[28] Karpov & Mylnikov 1989, and Ancyromonadida[23] Atkins et al. 2000 by sequencing polyA-containing transcripts of recently cultured species spanning the diversity of these two clades[31,32]. We also generated transcriptome data for the Mediterranean type strain of *Meteora sporadica*[26] (Supplementary Table 1). Since these eukaryotic species were in co-culture with diverse bacteria and sometimes other protists, we established a pipeline (Supplementary Fig.1) to eliminate contaminant sequences and retained only targeted protein-coding genes for subsequent phylogenomic analyses (see Methods). After the strict manual

curation of these datasets, we obtained a set of highly complete transcriptomes, as assessed by the presence of universal single-copy genes using BUSCO[33]. Most of our transcriptomes were more than 80% and up to ~93% complete (Supplementary Table 1). Regarding the presence of the selected 303 phylogenetic markers for phylogenomic analyses (Methods), they contained relatively few missing markers (~6% and 8%, on average, for apusomonads and ancyromonads, respectively; 8.5% for *Meteora*) and total amino acid gaps in the final multimarker alignment (7.9% and 11.7% , on average, for apusomonads and ancyromonads, respectively; 11.5% for *Meteora*). These represented comparable or even more complete datasets than those for the few previously 454-sequenced transcriptomes for the same clades[34], as well as the only available apusomonad genome, that of *Thecamonas trahens*[35] (13% missing data), and the transcriptomes of the apusomonad *Striomonas* (formerly *Nutomonas*) *longa*[36] (15.9% gaps) and the ancyromonads[22] *Ancyromonas kenti* (42.83% gaps) and *Fabomonas tropica* (24.2% gaps) (Supplementary Table 2).

The less complete transcriptomes in our newly generated datasets were those of *Apusomonas proboscidea* and *Nutomonas limna terrestris* (29.2% and 24.4% gaps, respectively). *A. proboscidea* was co-cultured with a stramenopile contaminant. In general, the less complete transcriptomes corresponded to those that suffered most cross-contamination; other examples were the ancyromonad *Nyramonas silfraensis* (14.9% gaps) and the apusomonad *Chelonemonas dolani* (14.52% gaps) (Supplementary Table 2). Although we could easily remove cross-contaminant sequences when analysing single marker trees, we were very strict not to include any potential contamination from prey or co-cultured microorganisms in order to retain only high-quality data (see Methods; Supplementary Fig.1. and Supplementary Tables 3-4). In addition, since de novo transcriptomes are prone to show artificially duplicated sequences in comparative genomics analyses, we tested the inferred oligopeptide redundancy by clustering sequences with CD-HIT at 90% identity. This procedure removed few such oligopeptide sequences for most species (6.5% on average), except for the highly duplicated *Mylnikovia oxoniensis* (~42%), as well as for *Multimonas media* (20.5%), *Apusomonas australiensis* (15.5%) and *Cavaliersmithia chaoae* (9.6%). However, the removal of this redundancy did not affect the BUSCO completeness (Supplementary Table 3) and the information available for the set of conserved proteins used in phylogenomic analyses.

Phylogenomic analyses of an expanded dataset of heterotrophic flagellates

To infer the evolutionary relationships among major eukaryotic supergroups, we included data from our transcriptomes and from a large representation of heterotrophic flagellates. Our final, manually curated dataset contained 303 concatenated markers, corresponding to 97,171 amino acid positions, for a total of 101 selected taxa (Supplementary Table 2). We applied state-of-the-art complex mixture models of sequence evolution in both maximum likelihood (ML) and Bayesian inference (BI) methods to alleviate putative homoplasy and long-branch attraction artefacts. The ML tree was reconstructed using the PMSF approximation of the model LG+C60+F+G (Supplementary Fig.2), and the BI trees, with the CAT-GTR (Supplementary Fig.3) and CAT-Poisson (Supplementary Fig.4) models of sequence evolution, respectively. BI and ML phylogenomic analyses yielded congruent tree topologies for major eukaryotic groups, with only minor changes in the position of some branches (Fig.1 and Supplementary Figs.1-3).

All our analyses retrieved the monophyly of Apusomonadida and Ancyromonadida with maximal support. The internal relationships within each of these groups were overall stable, except for very minor uncertainties (Fig.1). The Apusomonadida comprised two major clades, A and B, both grouping a mix of marine and freshwater species showing an overall internal topology congruent with that based on 18S rRNA gene phylogenies[31]. Clade A grouped species characterized by an elongated *Amastigomonas*-type morphology and relative large cell size (~8 µm; genera *Multimonas*, *Podomonas*, *Cavaliersmithia*, *Mylnikovia* and *Catacumbia*). Clade B comprised the Thecamonadinae, also grouping species with the elongated *Amastigomonas*-type morphology but of smaller size (~5 µm cell diameter; genera *Singekia*, *Karpovia*, *Chelonemonas* and *Thecamonas*)[31], and the Apusomonadinae, which grouped the genera *Apusomonas*, characterized by rounded cells, and *Manchomonas*, exhibiting elongated and larger cells (9.5 µm)[37]. The only internal uncertainty within the apusomonads concerned the placement of *Manchomonas*, which branched sister to the Thecamonadinae, instead of *Apusomonas*, in BI trees using the CAT-GTR model, albeit with modest support (0.74 posterior probability –PP–; Supplementary Fig.3). These topological differences may be a consequence of the high percentage of missing data from *Manchomonas bermudensis* (only 62 out of the 303 selected markers proteins, 90.49% gaps), with only EST data available for this species (F. Lang, unpublished; Supplementary Table 2). Apusomonadida formed a fully supported clade with Opisthokonta, which was sister to the Breviatea (Obazoa). In turn, this Obazoa clade was sister to Amoebozoa with full support (Amorphea clade) and, the Amorphea to the CRuMs with very good to full support, forming the supergroup Podiata[38] (Fig.1).

Similarly, we retrieved the monophyly of the two described ancyromonad families, Ancyromonadidae and Planomonadidae[39], and the recovered internal topology for the clade was again congruent with that based on 18S rRNA genes[32]. Within Ancyromonadidae, we retrieved a clade of marine representatives (*Caraotamonas* and *Ancyromonas* species) and another of freshwater members (*Striomonas*, *Nutomonas* and *Nyramonas*) (Fig.1). However, we only retrieved the monophyly of Planomonadidae, including *Planomonas* and *Fabomonas*, with ML; BI analyses placed *Fabomonas* as the earliest-branching ancyromonad, sister to all other ancyromonad genera (0.8 PP for CAT-GTR, and 0.99 PP for CAT-Poisson; Supplementary Figs.3-4). Interestingly, malawimonads[20], traditionally classified within Excavata[40-42], branched sister to the ancyromonads in our ML tree (Fig.1, Supplementary Fig.2), in line with some previous analyses[22,34]. Ancyromonadida and Malawimonadida formed a monophyletic supergroup with Amorphea and CRuMs, the Opimoda, fully supported in the ML tree and well supported in BI with CAT-GTR (0.99 PP). However, in the BI analyses, Ancyromonadida appeared as the earliest-branching lineage within Opimoda (0.99 PP and 0.75 PP under CAT-GTR and CAT-Poisson models, respectively), Malawimonadida being sister to the Podiata clade (0.98 PP and 0.73 PP under CAT-GTR and CAT-Poisson models respectively) (Fig.1; Supplementary Figs.2-4).

To test whether the aforementioned uncertainties in our phylogenomic tree, i.e. the position of *Manchomonas* within Apusomonadida, that of *Fabomonas* within Ancyromonadida and the position of Malawimonadida, could result from marked differences in evolutionary rate, we analysed the stability of the respective branches in the tree against alternative topologies (Fig.2a-c) after progressive removal of the fastest-evolving sites in the alignment (5% at a time). As can be seen in Fig.2d, the monophyly of *Manchomonas* and *Apusomonas* was

strongly supported until only 15% of the data remained. Likewise, the monophyly of *Planomonas* and *Fabomonas* was fully supported until less than 25% of sites remained. In the case of malawimonads, we monitored the statistical support for, in addition to the two observed topologies in our analyses, the monophyly of malawimonads and Discoba. We followed, as control, the monophyly of Opisthokonta, which were fully supported until 15% of sites remained (Fig.2d). We recovered general low support for the two observed alternatives, albeit the statistical support for the monophyly of malawimonads and ancyromonads was always much higher than malawimonads as sister of the Podiata. By contrast, the monophyly of malawimonads and Discoba was never observed (Fig. 2d).

Although our eukaryotic tree was not rooted with an external outgroup, we included the widest possible diversity of free-living protists also on the Diphoda[19] side, including heterotrophic flagellates when possible. We arbitrarily rooted our trees using the excavate clade Metamonada, with the inclusion of the relatively short-branching *Trimastix* species (Fig.1). As expected, Metamonada appeared paraphyletic with respect to Discoba[43,44]. Also, we recovered the monophyly of Diaphoretickes with full support (Fig.1). Within Diaphoretickes, the SAR supergroup also received full support. However, the monophyly of SAR and telonemids found in some analyses (the TSAR group[45]) was not observed. Telonemids branched within or sister to Haptista, albeit with moderate-to-low support (Fig.1). The Archaeplastida clade was recovered with moderate (96%; ML) to full support (1 PP; BI CAT-GTR) and included the Picozoa as sister to Rhodelphida + Rhodophyta, in agreement with recent observations[46], albeit with full support only from BI analyses using CAT-GTR. We also recovered with full support the widely accepted monophyly of Archaeplastida and Cryptista[21,46]. Finally, we retrieved the monophyly of *Ancoracysta* (Provora)[25], Hemimastigophora[21] and the *Meteora sporadica* type strain CRO19MET[26] with full ML support, as has been recently observed with another strain of *M. sporadica*[27]. *Ancoracysta* and other Provora members have also been suggested to form a monophyletic group with Hemimastigophora[24]. Our results thus provide additional support for this new supergroup of morphologically diverse predatory protists.

Evolution of major phenotypic traits during the early eukaryotic radiation

Our results suggest that, although resolving the phylogenetic tree of eukaryotes is challenging, the topology of the tree can be stabilized with the incorporation of a more balanced taxon sampling, including a wider representation of deep-branching free-living protists, combined with the use of appropriate phylogenetic reconstruction approaches. Our phylogenomic analyses including an expanded sampling of apusomonads and ancyromonads converge to a rather stable topology for some major clades, notably in the Opimoda supergroup and its increasingly nested phylogenetic clades (Podiata, Amorphea, Obazoa), which appear sister to a clade containing the ancyromonads and, most likely, the malawimonads (Fig.1). This phylogenetic framework allows to comparatively assess the distribution of complex phenotypic features, infer the ancestral states for the different clades and propose plausible evolutionary scenarios for trait evolution. Obviously, the availability of morphological and structural data is still limited, many morphological features may not be homologous, and more biodiversity alongside ultrastructural and phylogenetic analyses with, notably, the inference of a rooted tree of eukaryotes (e.g. using mitochondrial or Asgard archaeal-derived markers) will be needed to validate and/or complete this emerging evolutionary scheme (Fig.3).

It is interesting to note that some large clades, notably the ancyromonads, show apparent phenotypic stasis. Indeed, ancyromonads have comparable levels of sequence divergence to those of other morphologically and structurally diverse lineages such as Amoebozoa or Opisthokonta, but have retained a constrained morphotype for hundreds of million years, suggesting an efficient adaptation to their predatory lifestyle in benthic and soil ecosystems. Unlike apusomonads, there seems to be no obvious morphological differences between ancyromonad family members in terms of cell size, shape or any other particular feature observable under light microscopy[32]. Ecologically, ancyromonads include marine and freshwater species, with the topology of the tree for the current taxon sampling suggesting a single transition from marine ancestors to a clade of freshwater genera (*Nyramonas, Nutomonas, Striomonas*; Fig.1). Accordingly, the last common ancestor of ancyromonads was likely marine and resembled extant species, having bean-shaped flattened biflagellate cells with a short anterior flagellum and a rostrum with extrusomes[32]. Although ancyromonads have a dorsal pellicle, a complex cytoskeleton and a ventral groove for feeding, they do not bear the bona fide excavate groove or actin-based pseudopods similar to those of diverse Obazoa[32] (Fig.3). By contrast, apusomonads, another diverse clade with high internal evolutionary divergence, exhibit some apparent morphological differences and ecological variation, with marine-freshwater transitions having occurred multiple times during their evolution[29,31] (Fig.1). Within clade A, members of the four described genera exhibit a short proboscis sleeve as compared to the cell body length, as well as more-prominent pseudopodia than other apusomonads[31]. Additionally, *Podomonas* has a tusk, a character also present in Thecamonadinae[47]. *Multimonas* also displays a potential tusk, posterior extrusomes[48] and division by binary and multiple fission. *Podomonas* and *Mylnikovia* possess refractile granules running in parallel to the posterior flagellum[49]. All these characteristics can be found in other apusomonads[31]. In clade B, if *Karpovia* or *Singekia* present a tusk, not clearly observed yet under optical microscopy, they would also fit the description of the subfamily Thecamonadinae[47], alongside trailing pseudopodia[31]. Also within clade B, *Manchomonas* and *Apusomonas* share some morphological similarities, such as few pseudopods, a hidden posterior flagellum, the absence of tusks, and some ultrastructural features[50]. *Manchomonas* has the largest observed sleeve compared to other apusomonads of elongated shapes, while *Apusomonas* exhibits a unique structure called mastigophore. Given the widespread elongated, *Amastigomonas*-type, cell shape in apusomonads, this seems the ancestral phenotype for the group, with the rounded *Apusomonas* cells being derived. Although there is considerable morphological variation within the group and many environmental species remain to be described[29], the available information suggests that the last apusomonad common ancestor had a typical elongated cell type, with dorsal pellicle, ventral feeding groove, actin-based pseudopodia, and proboscis (likely with a short sleeve), probably with a tusk and able to divide by multiple fission (Fig.3).

In our phylogenomic tree, malawimonads and ancyromonads were sister groups (Fig.1). If confirmed, this would imply that this clade is one of the earliest branching lineages after the Opimoda-Diphoda split. Alternatively, malawimonads could be sister to the Podiata. Whatever the actual topology, since both clades encompass small bacterivorous heterotrophic biflagellates with a, likely homologous[51], ventral feeding groove[20,32,52], the last common ancestor of Opimoda most likely shared this excavate-like phenotype (Fig.3). Furthermore,

since excavates are likely paraphyletic[22,24,35], with Metamonada and Discoba branching deeply in the eukaryotic tree (Fig.1), it might be argued that the ancestral LECA phenotype also corresponded to that of an excavate-like biflagellate[53], irrespective of the specific position of the root[10] (Fig.3). This seems further supported by the strong conservation of the microtubular cytoskeleton and the flagellar apparatus[54] across eukaryotes[55]. From such an ancestor, ancyromonads developed a dorsal pellicle and a rostrum with extrusomes, losing the flagellar vanes and shortening the anterior flagellum[20]. It is possible that the opimodan ancestor was a marine planktonic biflagellate that lost free swimming capabilities and adapted to glide on benthic substrates. Without the excavate physical constrains for feeding that are intertwined with the flagellar motility[51], the ancyromonad morphotype might have easily evolved (Fig. 3). Bearing a dorsal pellicle, ancyromonads maintained ventral feeding, keeping the posterior flagellum attached to the surface, which is likely at the origin of the "twitch- yanking" movement for feeding[32].

Our phylogenomic tree supported the monophyly of Amorphea (Amoebozoa, Breviatea, Apusomonadida and Opisthokonta) and CRuMs (Fig.1). This clade had already been proposed to occur based on the shared capability to produce pseudopodia across their members and, accordingly, named Podiata[38]. The last common ancestor of Amorphea was clearly able to produce pseudopodia (Fig.3). However, the available cell biology descriptions for CRuMs are still limited[22] and the presence of some features, such as a pellicle homologous to that of apusomonads and ancyromonads[56], is unclear. Likewise, the presence of true pseudopodia in some species (e.g., *Micronuclearia* and *Rigifila*) needs further confirmation. Nonetheless, at least some *Mantamonas* species (e.g. *M. plastica*) do bear pseudopodia[57] and, accordingly, the ancestor of the group. Mantamonadida have a stable phylogenetic position[58] and might have retained some of the ancestral features of CRuMs; indeed, *Mantamonas* share some similarities with ancyromonads, notably the possession of small flattened cells with a short anterior flagellum and benthic/soil-associated lifestyles. However, the other CRuM linages (Diphyllatea and Hilomonadea) include larger, freshwater planktonic species, also possessing a ventral groove[56,59]. Populating the CRuMs' branch with new described members should help to ascertain trait evolution within this clade.

The Amorphea contain the most diversified and studied lineages of the opimodan side of the eukaryotic tree, Amoebozoa and Opisthokonta. These clades exhibit more diverse and derived morphoplans than the two other amorphean lineages, Apusomonadida and Breviatea. Amoebozoa and Opisthokonta lost one or both flagella[60,61]. Biflagellate protists have complex microtubular cytoskeletons that impose severe structural constrains on their cell shape, such that flagellar loss, freeing those constrains, likely facilitated the evolution of more diverse morphoplans. This increased morphological evolvability linked to concomitant changes in selective pressures allowed the exploration of novel cell shapes (e.g. amoeboid), feeding modes (e.g., osmotrophy) and cell-cell interactions (e.g., multicellularity), as currently observed across this clades[34]. By contrast, both Breviatea and Apusomonadida, being paraphyletic within the Obazoa, can be inferred to have retained the ancestral biflagellate state alongside other features likely present in the obazoan ancestor (Fig.3). Both lineages encompass small bacterivorous amoeboflagellates that phagocytize prey using pseudopodia (unlike ancyromonads or excavates)[49]. Breviates are anaerobic and possess gliding and swimming forms, lack a pellicle and display more pronounced amoeboid shapes than apusomonads[16].

Apusomonads comprise gliding, elongated, semi-rigid amoeboflagellates with an anterior proboscis and a dorsal pellicle allowing ventral feeding similarly to ancyromonads[31,47]. Interestingly, both breviates and apusomonads can divide palintomically, i.e. their karyokinesis is uncoupled from cellular division leading, at least transiently, to multinucleated cells. The ability to generate multinucleated cells is also widespread across amoebozoans and opisthokonts, including several unicellular relatives of animals (e.g. corallochytrids)[62] and fungi (e.g. aphelids)[14]. It seems increasingly clear that the occurrence of cenocytic (i.e., multinucleated) stages was important along the road to metazoan multicellularity[15]. Therefore, the ancestral Amorphea capability to produce multinucleated cells appears intriguingly crucial for the subsequent evolution of plasmodial growth, and metazoan and hyphal-based fungal multicellularity.

Methods

Protist culture, RNA extraction and transcriptome sequencing

The 22 species of heterotrophic flagellate protists from which we generated transcriptome data were previously isolated or enriched from benthos of marine or freshwater ecosystems or soil, and included apusomonads[31], ancyromonads[32] and the type strain of *Meteora sporadica*[26] (Supplementary Table 1). Soil and freshwater flagellates were grown in Volvic water with 1% yeast tryptone (YT), and marine flagellates in 0.2 micron-filtered seawater with 1% YT medium or Cerophyl medium. Flagellates were grown in flat cell culture flasks with ~10 ml of medium, just to cover the bottom surface (75 cm^2). The gliding protist cells from high-density cultures were collected by gently scratching the bottom of the flask with a cell scraper and pooled in 50 ml Falcon tubes. Cells were pelleted by centrifugation at 10ºC for 15 minutes at 15,000g. Total RNA for each species was extracted with the RNeasy mini Kit (Qiagen) following the manufacturer protocol, including DNAse treatment. Purified RNA was quantified using a Qubit fluorometer (ThermoFisher Scientific). For each species, cDNA Illumina libraries were constructed after polyA mRNA selection, tagged and paired-end (2 × 150 bp) sequenced with Illumina NovaSeq 6000 S2 (Eurofins Genomics, Germany) in three different sequencing runs (NG-22350, NG-24277, NG-25209; Supplementary Fig.1 and Supplementary Table 4). Sequence statistics and accession numbers are provided in Supplementary Table 1.

Transcriptome assembly and decontamination

The quality of Illumina sequences was checked with FastQC[63] v0.11.8. High-quality reads were retained and used for transcriptome *de novo* assembly using Spades[64] v3.13.1 66 with -rna mode and default parameters. Cross-contamination of transcripts among multiplexed cDNA libraries within the same sequencing run were detected and removed using CroCo[65] v1.2 (Supplementary Fig.1, Supplementary Tables 3-4). We translated the remaining transcripts into oligopeptides using TransDecoder v5.5 (https://github.com/TransDecoder/) with the script LongOrfs and clustered them with CD-HIT[66] v4.8.1 at 100% identity. Predicted oligopeptides of non-eukaryotic origin were removed using similarity search against a custom protein

database (BAUVdb: bacteria, archaea, eUkaryotes and viruses). BAUVdb included 32 reference genomes of eukaryotic species spanning all supergroups (Supplementary Table 5) as well as, respectively, 62,291 and 3,412 bacterial and archaeal genomes from the Genome Taxonomy DataBase, GTDB[67] release 207, 361,930 viral genomes from the reference viral database (RVDB)[68] and RVDB proteins[69] clustered using CD-HIT at 90% identity and coverage. To discriminate target eukaryotic sequences from those of alien origin, we then applied Diamond[70] v2.0.14.152 in ultra-sensitive mode setting an e-value threshold of 1e-3 for a maximum of 100 hits. From the tabular output, eukaryotic oligopeptides were retained if *Thecamonas trahens* or five eukaryotes were the first hits. Eukaryotic oligopeptides were automatically annotated using eggNOG-mapper[71] using all orthologs and all annotations as evidence. The number of identified single-copy markers from BUSCO[33] v5.2.2 were used as a proxy for gene set completeness using the eukaryote database db10. All transcripts and oligopeptides generated in this study are available in the figshare repository (10.6084/m9.figshare.22148027). Since *A. proboscidea* MPSANABRIA15 was co-cultured with a difficult-to-eliminate stramenopile contaminant (belonging to the genus *Paraphysomonas*, as inferred from its 18S rRNA gene sharing ~98% pairwise identity with members of this genus), we also sequenced the transcriptome of the latter from a monoeukaryotic culture (NCBI biosample SRR23610779). To decontaminate the apusomonad set of proteins from stramenopile sequences, eukaryotic proteins of *A. proboscidea* MPSANABRIA15 were used as queries in BLASTP (minimum e-value of 1e-25) against a protein database with the closest related taxa (PROMEX), our own stramenopile contaminant, plus three publicly available *Paraphysomonas* transcriptomes: MMETSP1103, MMETSP0103 and MMETSP1107. Eukaryotic oligopeptides were considered contaminant when they had only stramenopile hits or, in the case of hits in both lineages, when the pairwise identity, weighted by coverage, was higher in stramenopiles than in PROMEX.

Phylogenomic analyses

Our phylogenomic dataset was updated from a previous paneukaryotic study with 104 taxa and 351 conserved markers[21]. To that dataset, we added the corresponding identified markers from recently sequenced genomes/transcriptomes of early-branching eukaryotes[45,72-75] and from our 22 flagellate brut transcriptomes (Supplementary Table 2). To identify the selected phylogenetic markers in these transcriptomes, we queried these datasets with the 351 marker sequences from *Homo*, *Saprolegnia*, *Spizellomyces* and *Diphylleia* with BLASTp and retrieved all possible homologs. Each marker from the 230 taxa representing all eukaryotic supergroups, was aligned with MAFFT[76] v7.427 77 (L-INS-i with 1000 iterations), and trimmed using Trimal[77] v1.4.rev22 in automated mode. Approximate-maximum likelihood single marker trees were inferred from each trimmed alignment using FastTree[78] v2.1.11 with default parameters, and examined manually with Figtree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Alignments were inspected with Aliview[79] v1.26 80, and pruned from contaminants, paralogs or spurious sequences. Markers with complex histories were removed from the dataset, resulting in a selection of 303 protein-coding genes. For final multi-marker phylogenetic analyses and to limit computational load, the taxon sampling was reduced to 101 eukaryotes, keeping at least five representatives for each known supergroup, mostly free-living heterotrophic flagellates. Finally, the markers were realigned, trimmed and concatenated with

alvert.py from the barrel-o-monkeys (http://rogerlab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm#Monkeybarrel), generating a concatenated matrix with 97,171 amino acidic sites that was used to infer phylogenies (10.6084/m9.figshare.22148027). The Maximum likelihood (ML) phylogenetic tree was inferred using IQ-TREE[80] v1.6 using the PMSF approximation[81] and a guide tree inferred with the LG+C60+F+G mixture model. Statistical support was generated with 1,000 ultrafast bootstraps. The Bayesian inference (BI) analyses were conducted with PhyloBayes-MPI[82] v1.5a 81 with both CAT-Poisson and CAT-GTR models[83], with two and four MCMC chains respectively run for 10,000 generations, saving one every 10 trees. Analyses were stopped once convergence thresholds were reached (i.e. maximum discrepancy <0.1 and minimum effective size >100, calculated using bpcomp) and consensus trees constructed after a burn-in of 25%. Additionally, to minimize possible systematic errors, the fastest-evolving sites were progressively removed at 5% of sites at a time. For that, among-site substitution rates were inferred using IQ-TREE under the -wsr option and the best-fitting model, generating a total of 19 new data subsets (10.6084/m9.figshare.22148027). Each of them was used to infer a phylogeny using IQTREE with the LG+C60+F+G model and obtain the bootstrap supports for each split. We used CONSENSE from the PHYLIP v3.695 package (https://phylipweb.github.io/phylip/) to interrogate the UFBOOT files using a Python script (Nick Irwin, pers. comm.).

Data availability

Raw read sequences have been submitted to Sequence Read Archive under BioProject PRJNA907040; the specific accession numbers for each protist transcriptome are given in Supplementary Table 1. Transcripts and oligopeptides generated in this study are available in the figshare repository (10.6084/m9.figshare.22148027).

References

1 Eme, L., Sharpe, S. C., Brown, M. W. & Roger, A. J. On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* **6**, a016139 (2014).
2 Brocks, J. J. *et al.* Lost world of complex life and the late rise of the eukaryotic crown. *Nature* **618**, 767-773 (2023).
3 Koonin, E. V. The incredible expanding ancestor of eukaryotes. *Cell* **140**, 606-608 (2010).
4 Adl, S. M. *et al.* Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4-119 (2019).
5 Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The New Tree of Eukaryotes. *Trends Ecol. Evol.* **35**, 43-55 (2020).
6 Moreira, D. & López-García, P. Molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol.* **10**, 31-38 (2002).
7 Burki, F., Sandin, M. M. & Jamy, M. Diversity and ecology of protists revealed by metabarcoding. *Curr. Biol.* **31**, R1267-r1280 (2021).
8 Torruella, G. *et al.* Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. *Communications biology* **1**, 231 (2018).
9 Worden, A. Z. *et al.* Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**, 1257594 (2015).

10 Al Jewari, C. & Baldauf, S. L. Conflict over the Eukaryote Root Resides in Strong Outliers, Mosaics and Missing Data Sensitivity of Site-Specific (CAT) Mixture Models. *Syst. Biol.* **72**, 1-16 (2023).
11 Roger, A. J. & Hug, L. A. The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 1039-1054 (2006).
12 Eme, L. & Tamarit, D. Microbial Diversity and Open Questions about the Deep Tree of Life. *Genome Biol. Evol.* **16**, evae053 (2024).
13 del Campo, J. *et al.* The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* **29**, 252-259 (2014).
14 Galindo, L. J. *et al.* Phylogenomics Supports the Monophyly of Aphelids and Fungi and Identifies New Molecular Synapomorphies. *Syst. Biol.* **72**, 505-515 (2023).
15 Ruiz-Trillo, I., Kin, K. & Casacuberta, E. The Origin of Metazoan Multicellularity: A Potential Microbial Black Swan Event. *Annu. Rev. Microbiol.* **77**, 499-516 (2023).
16 Brown, M. W. *et al.* Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc. Biol. Sci.* **280**, 20131755 (2013).
17 Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297-354. (2002).
18 Richards, T. A. & Cavalier-Smith, T. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* **436**, 1113-1118 (2005).
19 Derelle, R. *et al.* Bacterial proteins pinpoint a single eukaryotic root. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E693-699 (2015).
20 Heiss, A. A. *et al.* Combined morphological and phylogenomic re-examination of malawimonads, a critical taxon for inferring the evolutionary history of eukaryotes. *Royal Society open science* **5**, 171707 (2018).
21 Lax, G. *et al.* Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature* **564**, 410-414 (2018).
22 Brown, M. W. *et al.* Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biol. Evol.* **10**, 427-433 (2018).
23 Atkins, M. S., McArthur, A. G. & Teske, A. P. Ancyromonadida: A new phylogenetic lineage among the protozoa closely related to the common ancestor of metazoans, fungi, and choanoflagellates (Opisthokonta). *J. Mol. Evol.* **51**, 278-285 (2000).
24 Tikhonenkov, D. V. *et al.* Microbial predators form a new supergroup of eukaryotes. *Nature* **612**, 714-719 (2022).
25 Janouškovec, J. *et al.* A New Lineage of Eukaryotes Illuminates Early Mitochondrial Genome Reduction. *Curr. Biol.* **27**, 3717-3724.e3715 (2017).
26 Galindo, L. J., López-García, P. & Moreira, D. First Molecular Characterization of the Elusive Marine Protist *Meteora sporadica*. *Protist* **173**, 125896 (2022).
27 Eglit, Y. *et al.* Meteora sporadica, a protist with incredible cell architecture, is related to Hemimastigophora. *Curr. Biol.* **34**, 451-459.e456 (2024).
28 Karpov, S. A. & Mylnikov, A. P. Biology and ultrastructure of colourless flagellates Apusomonadida ord. n. . *Zool Zh* **68**, 5-17 (in Russian) (1989).
29 Torruella, G., Moreira, D. & Lopez-Garcia, P. Phylogenetic and ecological diversity of apusomonads, a lineage of deep-branching eukaryotes. *Environ. Microbiol. Rep.* **9**, 113-119 (2017).
30 Fenchel, T. in *Adv. Microb. Ecol.* (ed K. C. Marshall) 57-97 (Springer US, 1986).
31 Torruella, G. *et al.* Expanding the molecular and morphological diversity of Apusomonadida, a deep-branching group of gliding bacterivorous protists. *J. Eukaryot. Microbiol.* **70**, e12956 (2023).

32 Yubuki, N. *et al.* Molecular and morphological characterization of four new ancyromonad genera and proposal for an updated taxonomy of the Ancyromonadida. *J. Eukaryot. Microbiol.* **70**, e12997 (2023).

33 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).

34 Cavalier-Smith, T. *et al.* Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. *Mol. Phylogenet. Evol.* **81**, 71-85 (2014).

35 Derelle, R. & Lang, B. F. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. *Mol. Biol. Evol.* **29**, 1277-1289 (2012).

36 Torruella, G. *et al.* Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr. Biol.* **25**, 2404-2410 (2015).

37 Molina, F. I. & Nerad, T. A. Ultrastructure of Amastigomonas bermudensis ATCC 50234 sp. nov.: A new heterotrophic marine flagellate. *European journal of protistology* **27**, 386-396 (1991).

38 Cavalier-Smith, T. Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. *European journal of protistology* **49**, 115-178 (2013).

39 Glücksman, E., Snell, E. A. & Cavalier-Smith, T. Phylogeny and evolution of Planomonadida (Sulcozoa): eight new species and new genera Fabomonas and Nutomonas. *European journal of protistology* **49**, 179-200 (2013).

40 Simpson, A. G., Inagaki, Y. & Roger, A. J. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Mol. Biol. Evol.* **23**, 615-625 (2006).

41 Parfrey, L. W. *et al.* Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst. Biol.* **59**, 518-533 (2010).

42 Cavalier-Smith, T. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma* **255**, 297-357 (2018).

43 Galindo, L. J., Prokina, K., Torruella, G., López-García, P. & Moreira, D. Maturases and Group II Introns in the Mitochondrial Genomes of the Deepest Jakobid Branch. *Genome Biol. Evol.* **15** (2023).

44 Stairs, C. W. *et al.* Anaeramoebae are a divergent lineage of eukaryotes that shed light on the transition from anaerobic mitochondria to hydrogenosomes. *Curr. Biol.* **31**, 5605-5612.e5605 (2021).

45 Strassert, J. F. H., Jamy, M., Mylnikov, A. P., Tikhonenkov, D. V. & Burki, F. New Phylogenomic Analysis of the Enigmatic Phylum Telonemia Further Resolves the Eukaryote Tree of Life. *Mol. Biol. Evol.* **36**, 757-765 (2019).

46 Schön, M. E. *et al.* Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. *Nat Commun* **12**, 6651 (2021).

47 Heiss, A. A., Walker, G. & Simpson, A. G. The microtubular cytoskeleton of the apusomonad *Thecamonas*, a sister lineage to the opisthokonts. *Protist* **164**, 598-621 (2013).

48 Jeong, D. H., Lee, H. B., Heiss, A. A., Cho, B. C. & Park, J. S. Morphological, phylogenetic and biogeographic characterizations of three heterotrophic nanoflagellates isolated from coastal areas of Korea. *Mar. Biol. Res.* **19**, 407-418 (2023).

49 Heiss, A. A., Brown, M. W. & Simpson, A. G. B. in *Handbook of the Protists* (eds John M. Archibald *et al.*) 1-27 (Springer International Publishing, 2016).

50 Cavalier-Smith, T. & Chao, E. E. Phylogeny and evolution of apusomonadida (protozoa: apusozoa): new genera and species. *Protist* **161**, 549-576 (2010).

51 Suzuki-Tellier, S., Kiørboe, T. & Simpson, A. G. B. The function of the feeding groove of 'typical excavate' flagellates. *J. Eukaryot. Microbiol.* **71**, e13016 (2024).

52 Heiss, A. A. *et al.* Description of *Imasa heleensis*, gen. nov., sp. nov. (Imasidae, fam. nov.), a Deep-Branching Marine Malawimonad and Possible Key Taxon in Understanding Early Eukaryotic Evolution. *J. Eukaryot. Microbiol.* **68**, e12837 (2021).

53 Cavalier-Smith, T. Ciliary transition zone evolution and the root of the eukaryote tree: implications for opisthokont origin and classification of kingdoms Protozoa, Plantae, and Fungi. *Protoplasma* **259**, 487-593 (2022).

54 Yubuki, N. & Leander, B. S. Evolution of microtubule organizing centers across the tree of eukaryotes. *Plant J.* **75**, 230-244 (2013).

55 Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J. B. & Bettencourt-Dias, M. Evolution: Tracing the origins of centrioles, cilia, and flagella. *J. Cell Biol.* **194**, 165-175 (2011).

56 Yabuki, A., Ishida, K. & Cavalier-Smith, T. *Rigifila ramosa* n. gen., n. sp., a filose apusozoan with a distinctive pellicle, is related to Micronuclearia. *Protist* **164**, 75-88 (2013).

57 Glucksman, E. *et al.* The novel marine gliding zooflagellate genus Mantamonas (Mantamonadida ord. n.: Apusozoa). *Protist* **162**, 207-221 (2011).

58 Blaz, J. *et al.* One high quality genome and two transcriptome datasets for new species of Mantamonas, a deep-branching eukaryote clade. *Scientific data* **10**, 603 (2023).

59 Cavalier-Smith, T. & Chao, E. E. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *J. Mol. Evol.* **56**, 540-563. (2003).

60 Ocaña-Pallarès, E. *et al.* Divergent genomic trajectories predate the origin of animals and fungi. *Nature* **609**, 747-753 (2022).

61 Tekle, Y. I., Wang, F., Wood, F. C., Anderson, O. R. & Smirnov, A. New insights on the evolutionary relationships between the major lineages of Amoebozoa. *Sci. Rep.* **12**, 11173 (2022).

62 McCartney, B. & Dudin, O. Cellularization across eukaryotes: Conserved mechanisms and novel strategies. *Curr. Opin. Cell Biol.* **80**, 102157 (2023).

63 Andrews, S. Vol. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Cambridge, United Kingdom, 2010).

64 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).

65 Simion, P. *et al.* A software tool 'CroCo' detects pervasive cross-species contamination in next generation sequencing data. *BMC Biol.* **16**, 28 (2018).

66 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).

67 Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785-d794 (2022).

68 Goodacre, N., Aljanahi, A., Nandakumar, S., Mikailov, M. & Khan Arifa, S. A Reference Viral Database (RVDB) To Enhance Bioinformatics Analysis of High-Throughput Sequencing for Novel Virus Detection. *mSphere* **3**, 10.1128/mspheredirect.00069-00018 (2018).

69 Bigot, T., Temmam, S., Pérot, P. & Eloit, M. RVDB-prot, a reference viral protein database and its HMM profiles. *F1000Research* **8**, 530 (2019).

70 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **12**, 59-60 (2015).

71 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825-5829 (2021).
72 Leger, M. M. *et al.* Organelles that illuminate the origins of Trichomonas hydrogenosomes and Giardia mitosomes. *Nature ecology & evolution* **1**, 0092 (2017).
73 Burki, F. *et al.* Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. Biol. Sci.* **283**, 20152802 (2016).
74 Gawryluk, R. M. R. *et al.* Non-photosynthetic predators are sister to red algae. *Nature* **572**, 240-243 (2019).
75 Tikhonenkov, D. V. *et al.* New Lineage of Microbial Predators Adds Complexity to Reconstructing the Evolutionary Origin of Animals. *Curr. Biol.* **30**, 4500-4509.e4505 (2020).
76 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
77 Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
78 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
79 Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276-3278 (2014).
80 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2015).
81 Wang, H. C., Minh, B. Q., Susko, E. & Roger, A. J. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* **67**, 216–235 (2018).
82 Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611-615 (2013).
83 Rodrigue, N. & Lartillot, N. Site-heterogeneous mutation-selection models within the PhyloBayes-MPI package. *Bioinformatics* **30**, 1020-1021 (2014).

Acknowledgements

We are grateful to Tom Cavalier-Smith and Emma Chao for their contribution to evolutionary protistology and their friendly availability to discuss ideas and share protist strains. We thank Ares Rocañín-Arjó, Saioa Manzano, Moisès Gil, Xavier Grau-Bové and Nick Irwin for bioinformatic insights. We thank the UNICELL platform (http://www.deemteam.fr/en/unicell) for help in transcriptome production. This work was supported by the European Research Council (ERC) Advanced Grants 'Protistworld' and 'Plast-Evol' (322669 and 787904, respectively) and the Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN project SINGEK (http://www.singek.eu/; grant agreement no. H2020-MSCA-ITN-2015-675752). G.T. was supported by 2019 BP 00208, Beatriu de Pinós-3 Postdoctoral Programme (BP3), Grant agreement ID: 801370. LG was funded by the Horizon 2020 research and innovation programme under the European Marie Skłodowska-Curie Individual Fellowship H2020-MSCA-IF-2020 (grant agreement no. 101022101−FungEye).

678 **Figure legends**

679

680 **Figure 1. Phylogenomic tree of eukaryotes including an expanded diversity of**
681 **Apusomonadida and Ancyromonadida.** The tree was inferred using Maximum Likelihood
682 from 97,171 amino acid positions and 101 taxa (model of sequence evolution: LG+C60+F+G
683 – PMSF). The root of the tree has been arbitrarily placed between Metamonada and the rest of
684 eukaryotes. Numbers at nodes represent ultrafast bootstrap approximation percentages (1,000
685 replicates) followed by Bayesian posterior probabilities under CAT-GTR and CAT-Poisson
686 models, respectively. Black dots denote maximum support with all methods. The scale bar
687 indicates the number of expected substitutions per unit branch length. The asterisks indicate
688 taxa of typical excavates.

689

690 **Figure 2. Progressive removal of the fastest evolving sites in 5% increments to evaluate**
691 **alternative topologies within Opimoda.** The alternative positions tested were as follows. **a**,
692 Within the Apusomonadida: *Manchomonas* being sister to *Apusomonas* or to the
693 Thecamonadinae (dashed line). **b**, Within the Ancyromonadida: the monophyly of *Fabomonas*
694 and *Planomonas* (the Planomonadidae clade) versus their paraphyly (*Fabomonas* branching
695 deeply, dashed line). **c**, Regarding the placement of Malawimonadida and Ancyromonadida:
696 Malawimonadida and Ancyromonadida being monophyletic versus Malawimonadida
697 branching as sister to Podiata (and, accordingly, ancyromonads as the earliest-branching
698 Opimoda lineage) or Malawimonadida sister to Discoba (dashed lines). **d**, Plot showing the
699 bootstrap support in ML phylogenetic trees under the LG+C60+F+G model of sequence
700 evolution as sites are progressively removed. The monophyly of Opisthokonta (yellow) is used
701 as a control to indicate when the phylogenetic signal is too low to retrieve well-known robust
702 monophyletic clades.

703

704 **Figure 3. Early trait evolution across Opimoda lineages.** The distribution of five key
705 morphological traits is parsimoniously inferred for each lineage ancestor (upper panel) based
706 on available descriptions and, to their respective last common ancestors (LCA), based on the
707 inferred phylogenetic backbone (lower panel). Numbers in black circles refer to the number of
708 flagella. The small character drawings in the cladogram represent the innovations at the onset
709 of each lineage.

710

711

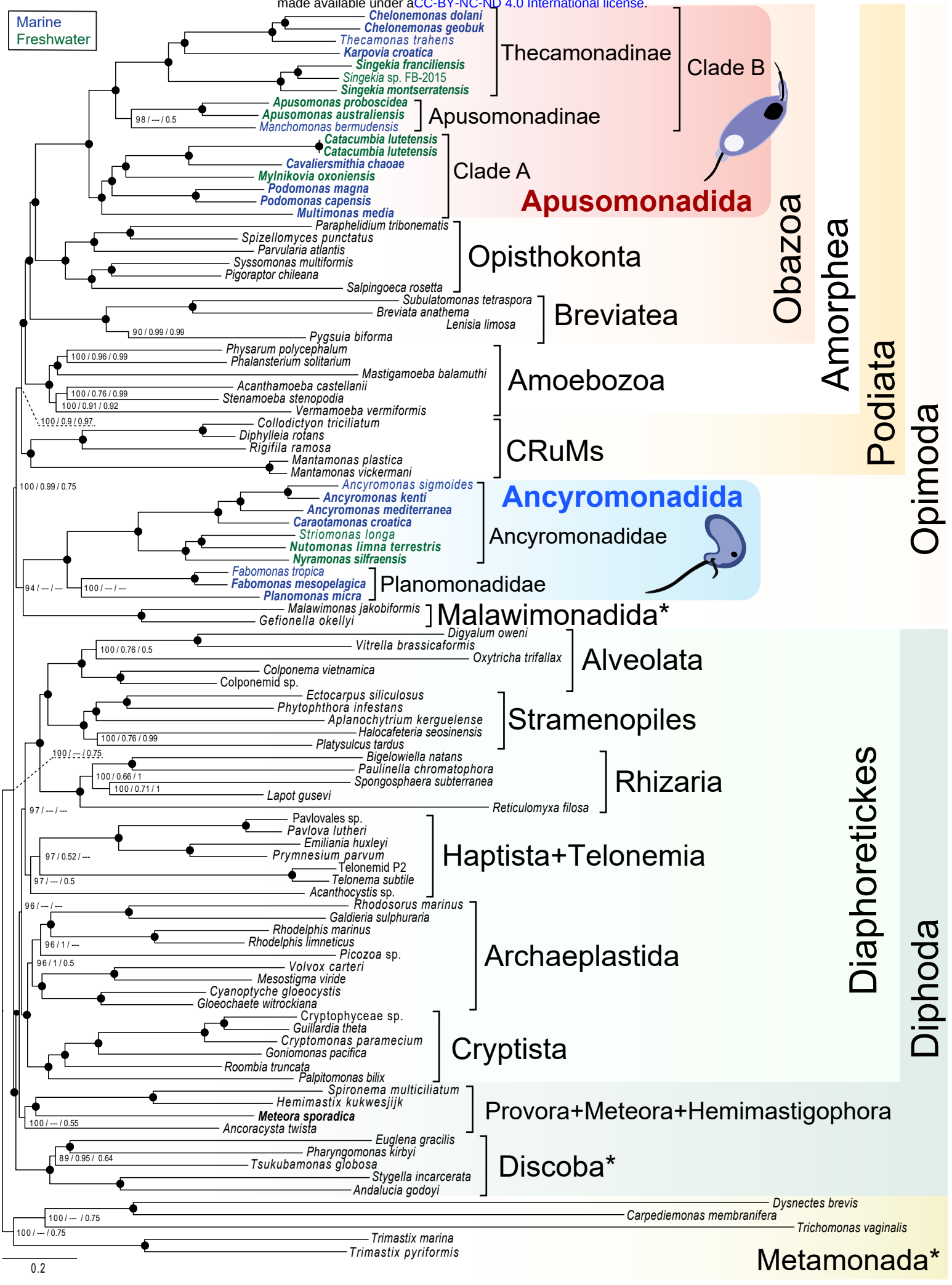

Figure 1

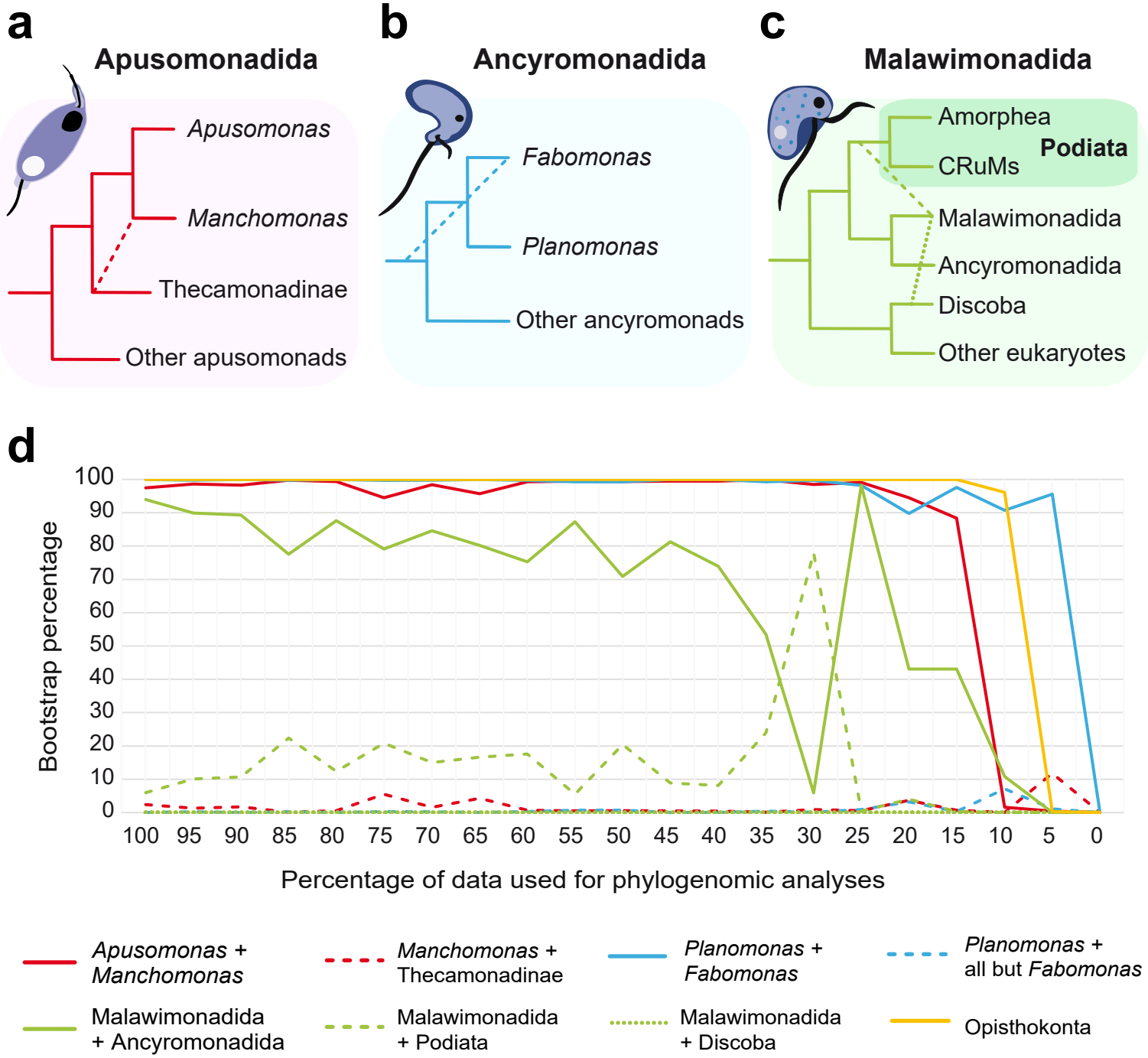

Figure 2

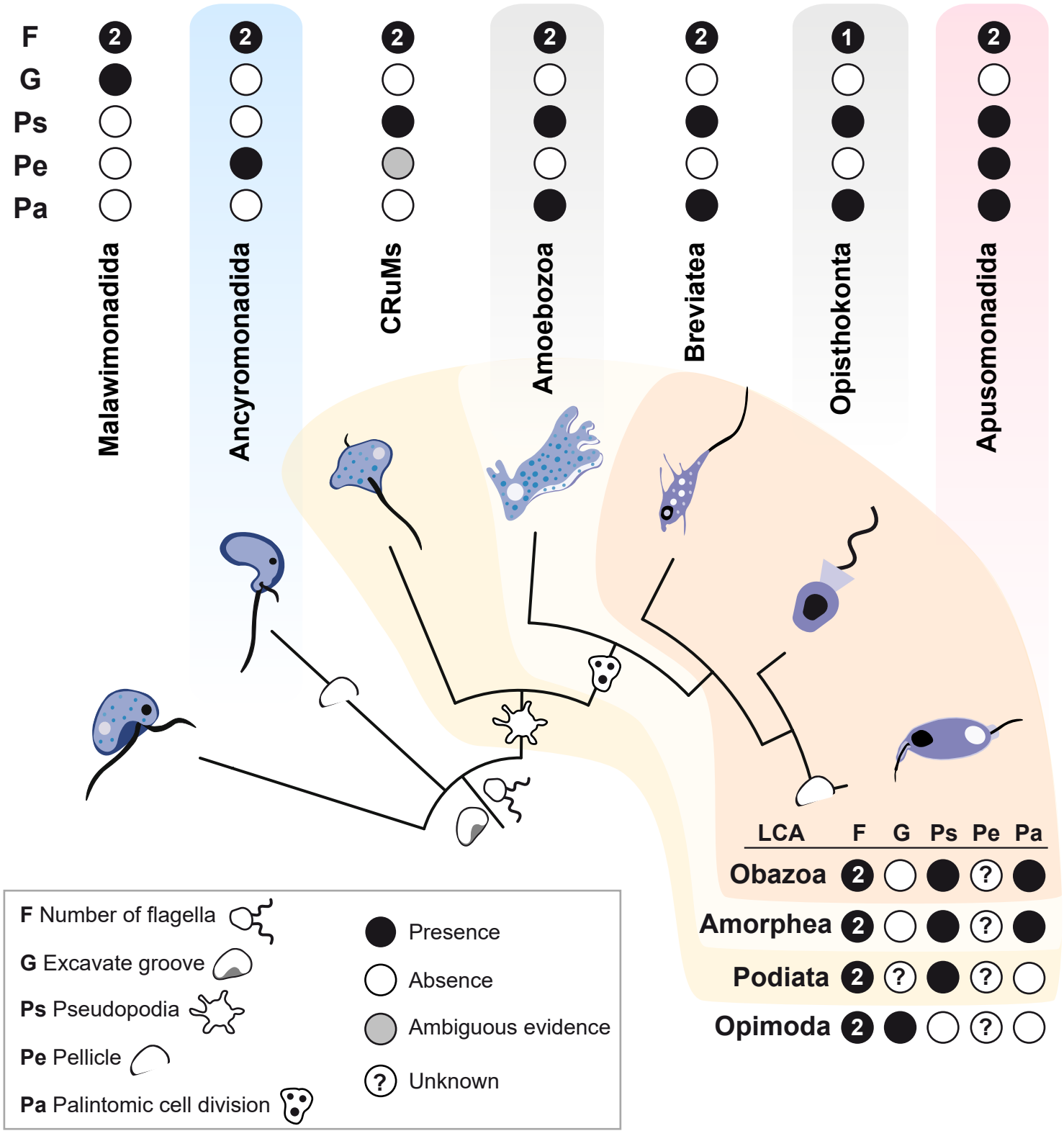

Figure 3