Article

Current Biology

Phylogenomics and Morphological Reconstruction of Arcellinida Testate Amoebae Highlight Diversity of Microbial Eukaryotes in the Neoproterozoic

Highlights

- Arcellinida testate amoebae are comprised of seven major lineages
- Reconstructed hypothetical ancestral states are congruent with Tonian fossils
- Combined analysis of phylogenies and fossils suggest divergence as early as 730 mya

Authors

Daniel J.G. Lahr, Anush Kosakyan, Enrique Lara, ..., Tomáš Pánek, Seungho Kang, Matthew W. Brown

Correspondence

dlahr@ib.usp.br (D.J.G.L.), mbrown@biology.msstate.edu (M.W.B.)

In Brief

Arcellinid testate amoebae are some of the earliest eukaryotic microbes. Lahr et al. show that these amoebae have seven major deep phylogenetic lineages, probably diverged as early as 730 mya. Diversification of these microbes in the mid-Neoproterozoic was probably happening alongside the gradual oxygenation of oceans.

Lahr et al., 2019, Current Biology *29*, 991–1001
March 18, 2019
https://doi.org/10.1016/j.cub.2019.01.078

CellPress

Current Biology
Article

Phylogenomics and Morphological Reconstruction of Arcellinida Testate Amoebae Highlight Diversity of Microbial Eukaryotes in the Neoproterozoic

Daniel J.G. Lahr,[1,10,*] Anush Kosakyan,[1,2] Enrique Lara,[3,4] Edward A.D. Mitchell,[4,5] Luana Morais,[6] Alfredo L. Porfirio-Sousa,[1] Giulia M. Ribeiro,[1] Alexander K. Tice,[7,8] Tomáš Pánek,[7,9] Seungho Kang,[7,8] and Matthew W. Brown[7,8,*]

[1]Department of Zoology, Institute of Biosciences, University of São Paulo, Brazil
[2]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 1160/31, 37005, České Budějovice, Czech Republic
[3]Real Jardín Botánico, CSIC, Plaza Murillo 2, ES 28014 Madrid, Spain
[4]Laboratory of Soil Biodiversity, University of Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland
[5]Botanical Garden of Neuchâtel, Pertuis-du-Sault 58, 2000 Neuchâtel, Switzerland
[6]Department of Geophysics, Institute of Astronomy, Geophysics and Atmospheric Sciences, University of São Paulo, Brazil
[7]Department of Biological Sciences, Mississippi State University, Starkville, MS, USA
[8]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Starkville, MS, USA
[9]Department of Biology and Ecology, Faculty of Science, University of Ostrava, Czech Republic
[10]Lead Contact
*Correspondence: dlahr@ib.usp.br (D.J.G.L.), mbrown@biology.msstate.edu (M.W.B.)
https://doi.org/10.1016/j.cub.2019.01.078

SUMMARY

Life was microbial for the majority of Earth's history, but as very few microbial lineages leave a fossil record, the Precambrian evolution of life remains shrouded in mystery. Shelled (testate) amoebae stand out as an exception with rich documented diversity in the Neoproterozoic as vase-shaped microfossils (VSMs). While there is general consensus that most of these can be attributed to the Arcellinida lineage in Amoebozoa, it is still unclear whether they can be used as key fossils for interpretation of early eukaryotic evolution. Here, we present a well-resolved phylogenomic reconstruction based on 250 genes, obtained using single-cell transcriptomic techniques from a representative selection of 19 Arcellinid testate amoeba taxa. The robust phylogenetic framework enables deeper interpretations of evolution in this lineage and demanded an updated classification of the group. Additionally, we performed reconstruction of ancestral morphologies, yielding hypothetical ancestors remarkably similar to existing Neoproterozoic VSMs. We demonstrate that major lineages of testate amoebae were already diversified before the Sturtian glaciation (720 mya), supporting the hypothesis that massive eukaryotic diversification took place in the early Neoproterozoic and congruent with the interpretation that VSM are arcellinid testate amoebae.

INTRODUCTION

The detailed phylogenetic history of microbial eukaryotes is a matter that has received serious attention only in recent decades. Because morphological characters are limited, reliable reconstructions were possible only after molecular methods were applied to a wide diversity of microbes [1]. Conversely, reconciling molecular evolution insights with morphological ones remains challenging [2, 3]. Importantly, the majority of single-celled organisms leave no fossil record, with the valuable exception of microbes with hard parts, which can come in the form of cysts as in green algae, endoskeletons as in Radiolaria, or shells as in testate amoebae, the focus of the present paper [4]. Many microbial lineages across the tree of life are able to produce shells. These include Amphitremida and Bacillariophyta (diatoms) in Stramenopiles; Foraminifera and Euglyphida in Rhizaria; arcellinids in Amoebozoa; as well as a few other minor lineages [5]. Where they occur (marine and fresh water, humid soil and terrestrial mosses), shelled microbes tend to be highly abundant and diverse. The biological purpose of these microscopic shells is still a matter of debate, although a favored interpretation is defense against predators [4]. These ornamented and intricate structures provide valuable taxonomic information. Additionally, durable shells facilitate manipulation and identification for scientific areas that depend on counting individuals and species such as microbial community ecology, paleoenvironmental reconstruction, and even forensic science [6–8]. The presence of a hard part also increases possibilities of fossilization: examples of fossil shells range from Foraminifera, widely used as index fossils in the Phanerozoic, to the lesser known but older vase-shaped microfossils (VSMs) found as far back as the Tonian period in the Neoproterozoic (1,000–720 million years ago) [9]. Recent studies have demonstrated that Neoproterozoic microfossils are more common, abundant, and diverse than previously thought [10–12], thus indicating that the planet was supporting eukaryotic life before and during Snowball Earth events (Cryogenian period) [13]. However, interpreting the phylogenetic affinities of Neoproterozoic microfossils is still an ongoing exercise. Based on morphological data, these fossils have generally been attributed to either Arcellinida or

Euglyphida testate amoebae—though the euglyphid interpretation has recently fallen out of favor, as they appear to be a much younger lineage [14, 15]. Given the importance and applicability of microscopic shells, it is paramount to generate solid phylogenetic reconstructions upon which other interpretations can be made.

Even though shells provide a suite of additional characters compared to non-shelled eukaryotes, these are still insufficient to reconstruct phylogenetic history and quite often even the taxonomy of modern organisms [5]. Molecular reconstructions of extant taxa have completely shifted views on foraminiferan taxonomy, the largest and most well-documented group of shelled amoebae. Even in such a well-studied group, reconciliation with morphological features is still open to debate [16]. Similarly, phylogenetics of diatoms has also undergone a period of turbulence with incongruity not only between morphological and molecular reconstructions, but also between distinct molecular datasets originating from nuclear and chloroplast genomes [17, 18]. This leaves the third most diverse shelled group, the Arcellinida, a species-rich group of shelled amoebae in the Amoebozoa clade. The Amoebozoa is a supergroup of eukaryotes sister to Obazoa, the supergroup including Animals and Fungi [19]. The Arcellinida are home to an estimated 800–2,000 morphospecies, which is high for microbial eukaryote standards [5], and molecular eDNA and DNA barcoding studies show that their true diversity is likely much higher [20, 21]. They occur exclusively in fresh-water and soil habitats, where they are abundant and diverse [19]. Recent molecular studies are challenging the taxonomy and systematics of Arcellinida, which were historically based on morphology. While most wide-scope reconstructions focusing on Amoebozoa generally recognize the Arcellinida as a stable Tubulinea lineage within the Amoebozoa [22–24], reconstructions that focus on deeper taxonomic sampling of Arcellinida show the group to be non-monophyletic or poorly supported [25, 26]. Reconstructions based on the gene for the small sub-unit ribosomal RNA (18S), along with a small suite of protein-coding genes, have been helpful in indicating major lineages while demonstrating that the classical key morphological character, shell composition, cannot be relied upon [26]. Additionally, these reconstructions are not suitable for interpretation of deep lineages: due to very deep divergence times, the backbone of trees based on few genes is not strongly supported [5]. The majority of Neoproterozoic VSMs have been attributed to arcellinids [9], which makes this group at least 730 million years old, in accordance with most recent dating for the Chuar Group fossils [27], and potentially as old as 800 million years ago (mya) such as the VSMs from the Chichkan Formation (South Kazakhstan) [28]. As such, the lineage is a cornerstone in interpreting the evolution of microbial eukaryotes because this is the only group with good fossilization potential in the entirety of the Amoebozoa clade.

Here, we present a comprehensive phylogenetic reconstruction of the Arcellinida based on an unprecedented 250 genes dataset obtained from 19 newly sequenced transcriptomes. Samples were obtained mostly using single-cell transcriptomic methods, which enabled us to sample the major lineages, including those that cannot be cultured or are difficult to obtain (Figure 1, see also Figure S1). The resulting tree is robustly supported; corroborating the monophyly and revealing for the first time the deepest lineages of the group (Figure 2, see also Figures S2 and S3). Using this tree, we have performed a reconstruction of ancestral states to elucidate the possible morphology of ancient shells and enable direct comparison with microfossils (Figure 2). These novel results create a solid phylogenetic backbone for the Arcellinida, and spur a re-interpretation of the fossil record associated to the group, potentially illuminating key events in the early evolution of eukaryotes.

RESULTS

The phylogenetic tree presented here shows a monophyletic Arcellinida, with full support for all except three nodes that received high support (Figure 2, see also Figure S3). While Arcellinida have been treated as monophyletic in the majority of morphology-based works (e.g., [29]), and has been shown as monophyletic in wider-scope Amoebozoa molecular reconstructions [2, 22–24], most previous reconstructions focusing on the Arcellinida point either to a non-monophyletic or poorly supported clade while additionally showing unsupported branching of deep lineages within the group [26, 30]. The current tree presents 8 well-defined and robustly supported major lineages in the Arcellinida, of which only three had been previously identified (Sphaerothecina, Hyalospheniidae, and Phryganellina [5]). Based on the groups determined in the phylogenetic tree, we have scored 7 shell characteristics that describe shell shape and are generally recognizable both in microfossils and modern organisms (Table 1). We performed ancestral state reconstruction of these characters on a sample of the best population of 1,000 maximum likelihood phylogenomic trees [31]. Ancestral reconstruction of characters is done individually for each character, and thus the quantity of characters does not interfere with the robustness of analysis. Conversely, because all ancestral reconstructions are based on the same sampling of trees and derived rates of evolution, potential biases will be correlated in all reconstructions. In general, the characters with only 2 possible states (aperture outline, presence of neck, shell outline in apertural view) have generated robust results for hypothetical ancestors (i.e., a very high probability for a character state when compared to other possible character states, Table 2). Two of the characters with 4 possible states (shell composition, shell outline in lateral view) have generated robust results for most nodes except the deepest ones, while the two remaining characters (the morphometric ratios, ratio length/breadth and ratio aperture/breadth) have only generated robust results for shallow nodes. We interpret that the two morphometric ratios analyzed are highly variable and thus are not informative at this phylogenetic depth. For other results, we considered robust any reconstruction in which a character that showed probability above 50% and at least 10% difference from the second possible character state. For the last common ancestor of all Arcellinida, it is only possible to infer that the aperture was circular (p = 97%) and that a neck was absent (p = 99%). All other characters are dubious except shell outline in lateral view with a marginally significant chance of being hemispheric (p = 50%). The last common ancestor of all Arcellinida except Phryganellina had a circular aperture (p = 90%), no neck (p = 99%), was circular in apertural view (79%), had an agglutinated shell (p = 75%), and was hemisphaeric in lateral view (p = 52%, 12% more likely than oval). All other nodes that were less inclusive (shallower in the phylogeny) presented even more robust results (Table 2).

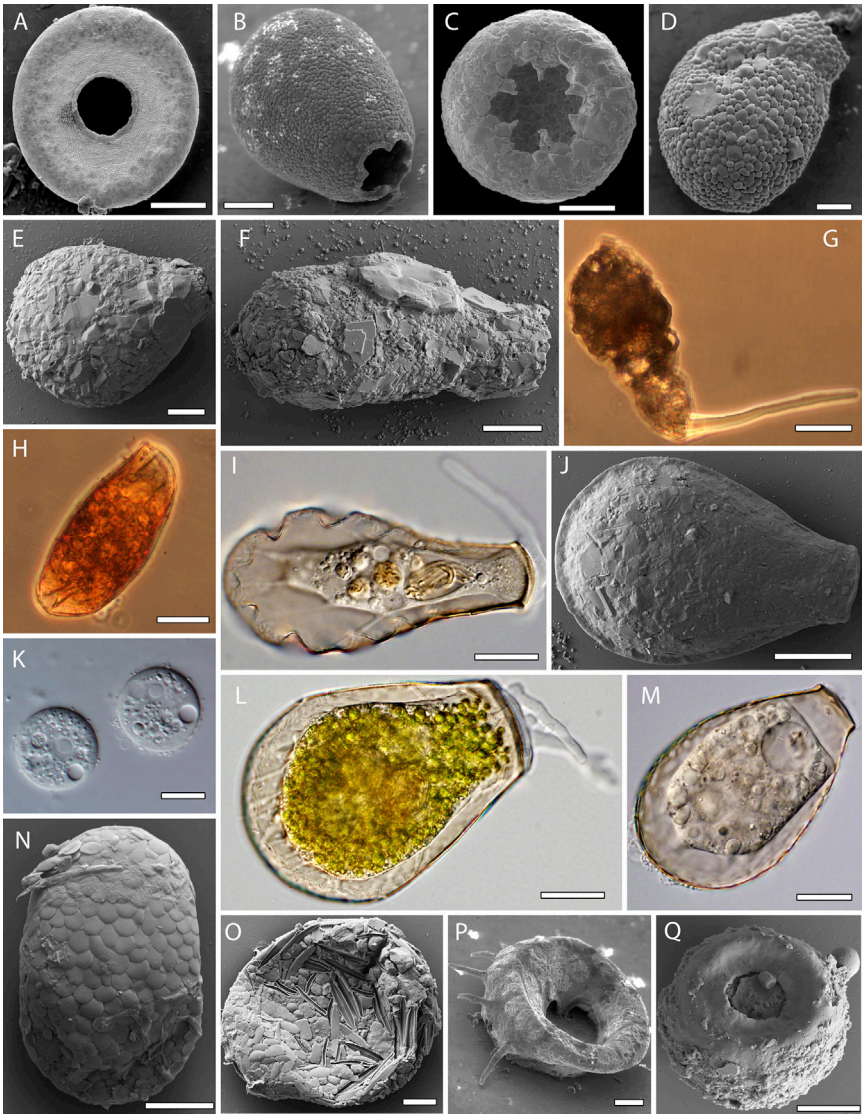

Figure 1. Representative Images of Sampled Testate Amoebae—Arcellinida

(A) Scanning Electron Micrograph (SEM) of *Arcella intermedia*, apertural view, scale bar 15 μm.
(B) *Netzelia sp.* (SEM), aperturo-lateral view, scale bar 20 μm.
(C) *Cyclopyxis lobostoma* (SEM), apertural view, scale bar 80 μm.
(D) *Lesquereusia mimetica* (SEM), lateral view, scale bar 10 μm.
(E) *Difflugia compressa* (SEM), lateral view, scale bar 40 μm.
(F) *Difflugia sp.* (SEM), lateral view, scale bar 40 μm.
(G) *Difflugia bryophila*, light micrograph (LM), lateral view of live individual with pseudopod protruding from the aperture, this is a voucher of the exact individual that was sequenced, scale bar 20 μm.
(H) *Microchlamys patella* (LM), dorsal view of a curled up live individual, this is the exact individual that was sequenced, scale bar 20 μm.
(I) *Hyalosphenia elegans* (LM), lateral view of live individual with pseudopod protruding from the aperture, scale bar 15 μm.
(J) *Planocarina carinata* (SEM), lateral view, scale bar 30 μm.
(K) *Pyxidicula operculata* (LM), dorsal view of two live individuals, with visible nuclei, scale bar 10 μm.
(L) *Hyalosphenia papilio* (LM), lateral view of live individual with protruding pseudopods from the aperture and visible endosymbiotic green algae, scale bar 20 μm.
(M) *Nebela tincta* (LM), lateral view of live individual, scale bar 20 μm.
(N) *Heleopera sylvatica* (SEM), lateral view with aperture towards the top, the many scales visible are scavenged from hyalosphenids and euglyphid testate amoebae, scale bar 10 μm.
(O) *Heleopera sphagni* (SEM), lateral view (aperture is on bottom left), the many scales seen are scavenged from euglyphids, hyalosphenids and diatoms, scale bar 10 μm.
(P) *Centropyxis aculeata* (SEM), aperturo-lateral view, scale bar 20 μm.
(Q) *Centropyxis sp.* (SEM), apertural view, scale bar 20 μm.
See also Figure S1.

Our approach enables a more detailed interpretation of the evolution of the arcellinid shell. Reconstructed hypothetical ancestors can be directly compared to real microfossils from the Neoproterozoic. It is important to note that these reconstructions are hypothetical, and represent the general characteristics that should be present in the common ancestor and all descendants of a given lineage. When congruent characteristics are identified between the reconstructed hypothetical ancestor and a real fossil, we then conclude that the fossil may represent a member of that lineage, as opposed to the actual ancestral taxon. The Glutinoconcha+Organoconcha hypothetical ancestral presents similar characteristics to *Taruma rata*, described from the Urucum Formation [11]. *Taruma rata* has a cylindrical test, with one flat end and may have a siliceous wall with an organic lining; alternatively, *Melanocyrillium hexodiadema* (described in [32]) also presents morphological characteristics similar to the reconstructed ancestral states of G+O hypothetical ancestor, except for the hexagonal aperture. An intrinsic limitation to the method is that character states need to be present in the terminal taxa to be reconstructed (the method cannot "create" intermediate states). As such, the hexagonal outline in apertural view is not included in our reconstruction as a possibility, since none of the terminal taxa possess such an aperture. However, there are modern taxa such as *Arcella mitrata var. spectabilis* that can present hexagonal apertures, but this case is very likely a convergent feature since *A. mitrata* is nested in the derived genus *Arcella* [26]. The Organoconcha ancestral is comparable to the genus *Paleoarcella* from the Chuar Group [32], in general shape and form, while additionally showing a hypothetical organic composition, as has been previously speculated for *Paleoarcella* and other VSMs [9, 33]. The Glutinoconcha reconstructed ancestral shares many similarities with the genus *Cycliocyrillium*, described originally for the Chuar Group in the Grand Canyon, USA (730 mya [9, 27, 32], later reported for both the Urucum Formation in Central Brazil (889–706 mya [11]) and the Callison Lake Formation in Canada (756–740 mya [10]). These are oval shaped in lateral view,

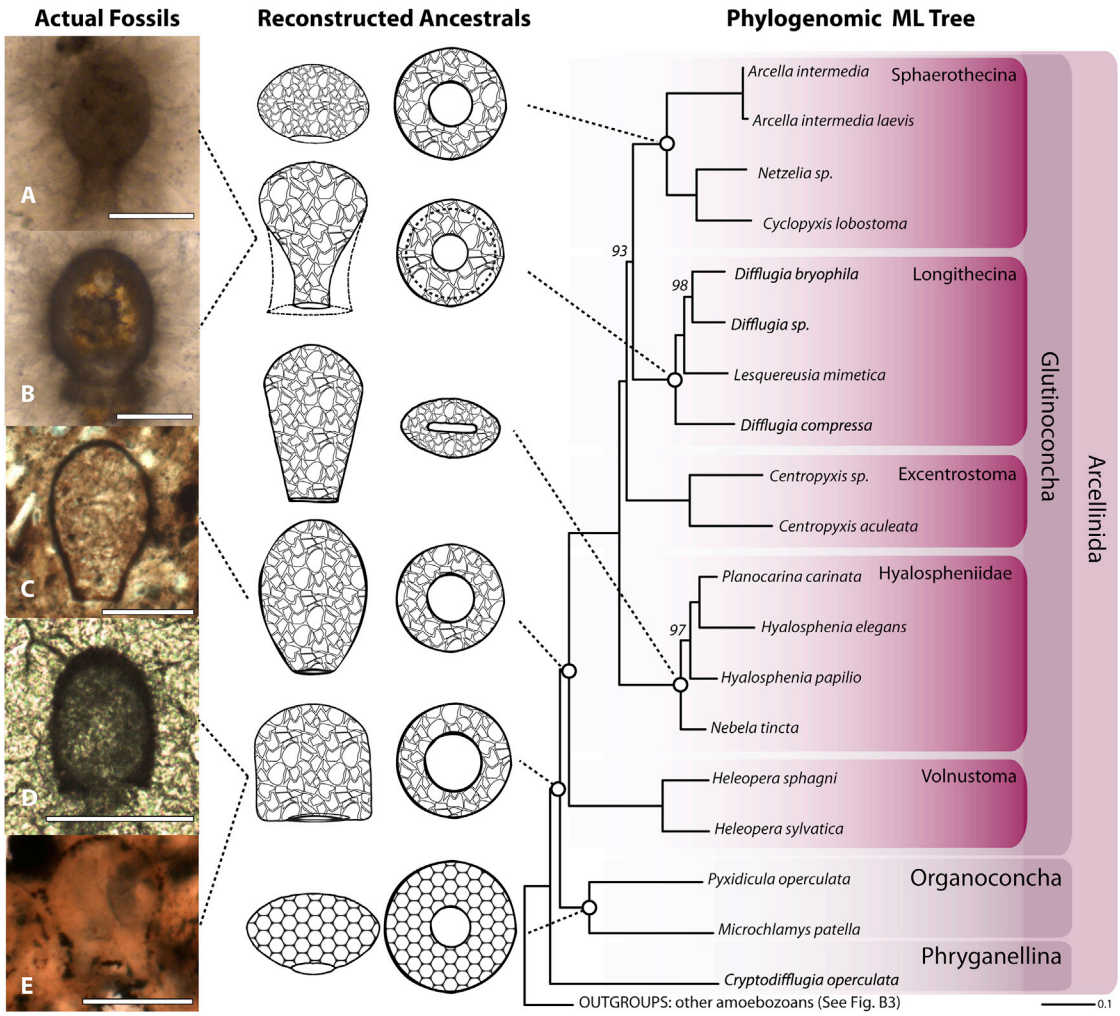

Figure 2. Phylogenomics, Ancestral State Reconstruction, and Neoproterozoic Fossils of Arcellinida

To the left are depicted five Neoproterozoic fossils thought to belong to shelled testate amoebae, newly photographed in petrographic slides for this work: (A) *Limeta lageniformis*; (B) *Palaeoamphora urucumense*; (C) *Cycliocyrillium torquata*; (D) *Taruma rata*. (A), (B), and (D) are from Jacadigo Group, Urucum Formation, Brazil (706–889 mya) – specimens in the Scientific Palaeontological collection at the Institute of Geosciences; and (E) *Mellanocyrillum hexodiadema* and (C) are from the Kawagunt Formation, Chuar Group, USA (730–750 mya) – specimens are in the collection at the Earth Science Department, University of California Santa Barbara. Scale bars are 50μm. The middle column shows cartoon representations reconstructed ancestrals in lateral and apertural view, the hexagonal pattern represents an organic shell, while the irregular pattern represents an agglutinated shell. The right column shows a Maximum Likelihood (ML) phylogenetic tree built in IQ-Tree under the LG+G4+C60+PMSF model of evolu- tion obtained from the analysis of 250 genes of the arcellinid species in terminals. Branches are drawn to scale, all nodes have a Bayesian posterior probability of 1 in the two converged PhyloBayes-MPI chains inferred under the CAT-GTR model, and full ML-Bootstrap support unless otherwise noted. See also Figure S3.

with a circular aperture, for which there are diverging interpretations about the original composition: some authors arguing for an agglutinated shell [11], while others argue for an originally organic shell [32]; alternatively, the genus *Trigonocyrillium*, also described for the Chuar Group [9, 32], has very similar overall shape. *Trigonocyrillium* has a triangular aperture, and since our reconstructions have not included the possibility of a triangular aperture (again, a limitation of the method), it remains a viable possible interpretation for the morphology of the hypothetical Glutinoconcha ancestral. The general shape and composition of the Longithecina reconstructed ancestral is comparable to either *Palaeoamphora urucumense* [11], if the larger aperture possibility is considered, or to *Limeta lageniformis* [11] if a smaller aperture is considered. The reconstructed hypothetical ancestors of Hyalosphenidae and Volnustoma (not illustrated, see Table 2) are almost identical, with the exception of the character "Shell Outline in Lateral View," which is likely oval for Volnustoma and likely pyriform for Hyalospheniidae. Combined with a slit-like aperture, absence of a neck, an ellipsoid shell outline in apertural view, the reconstructed hypothetical ancestor for Volnustoma is similar to the egg-shaped *Pakupaku kabin*, recently discovered in the Tonian Togari Group, Tasmania [12]. Finally, the reconstructed ancestral of Sphaerothecina presents a hemisphaeric outline in lateral view, and an agglutinated shell, a combination of characteristics that has not yet been described in the fossil record.

Table 1. Scoring of Morphological Characteristics Used for Ancestral State Reconstruction

	0	1	2	3
Aperture	circular	slit		
Neck	absent	present		
Shell outline (AV)	circular	ellipsoid		
L/B ratio	$x \leq 0.5$	$0.5 < x \leq 1$	$x > 1$	$x > 2$
A/B ratio	$x \leq 0.3$	$0.3 < x \leq 0.5$	$0.5 < x \leq 0.7$	$x > 0.7$
Composition	organic	xenosomes	idiosomes	calcareous
Shell outline (LV)	hemisphaeric	circular	oval	pyriform

AV = apertural view; L/B = length to breadth; A/B = aperture to breadth; LV = lateral view

DISCUSSION

Comparative Systematics of Arcellinida in Light of Recent Results: The Deepest Relationships within Arcellinida

The phylogenetic tree presented here, while not sampling the entirety of Arcellinida, represents well the breadth of morphological diversity within the group. The resulting relationships demand a departure from current interpretation of the evolutionary history and classification of Arcellinida (Figure 2, see also Figure S3 and Data S1). To stabilize and organize the current state of knowledge, we propose a renewed higher-level classification of the Arcellinida (Table 3, see also Data S1 for extensive comment on the taxonomic actions taken here). The deepest split shows a separation of the Phrygannelina, including *Cryptodifflugia*, from the rest of Arcellinida. While the monophyly of Phrygannelina has been demonstrated using SSU-rDNA analysis [34], the basal position of the whole group had not been identified previously in phylogenies. It is, however, justified by the shape of the pseudopodia (thin and conical), which differs from all other Arcellinida (cylindric and blunt). Based on this morphological feature, Bovee had created the taxon Phrygannelina of uncertain affinities [35], later placed within Arcellinida by Meisterfeld [29].

In the larger, also monophyletic lineage sister to *Cryptodifflugia*, the first group to split out is the Organoconcha, including both *Pyxidicula* and *Microchlamys*, which were long-branched lineages without a robust position in previous SSU-rDNA reconstructions (see [26, 34]). Because of their saucer-shaped, organic test, both genera were placed previously close to *Arcella* [29]. While it is true that test outline can be considered as a trait on considerable taxonomic importance in Arcellinida [36], both *Pyxidicula* and *Microchlamys* have a very wide aperture without or almost without edges. A third genus that shares these morphological features, *Spumochlamys*, is tentatively placed in Organoconcha pending phylogenomic data to confirm its position.

Glutinoconcha Includes the Majority of Sampled Arcellinida Diversity

The second large group to split is the Glutinoconcha, including the majority of known Arcellinida diversity. The first group within Glutinoconcha is the Volnustoma, which in previous reconstructions based on SSU-rDNA consistently fell out of Arcellinida, even showing non-monophyly of the genus *Heleopera* [25, 26], or was spuriously grouped due to long-branch attraction with unrelated taxa [34]. The group is now deeply nested within Arcellinida with full support. The four remaining fully supported groups are the Hyalospheniidae, consistently recovered previously and formalized based on Cox1 reconstructions [5, 37]; the Excentrostoma (genus *Centropyxis*) which had not been properly sampled until this work; and two other large groups which include a number of *Difflugia* related lineages. The first one includes *Arcella* with organic shells and other genera that form quasi-spherical, agglutinated shells, such as *Netzelia* and *Cyclopyxis*. This group has been previously revealed by molecular reconstruction based on SSU-rDNA and named Sphaerothecina [5]. The sister-group (Longithecina) is composed of pyriform *Diffflugia* and *Lesquereusia* and has not been previously identified.

Conflicts and Similarities between Phylogenomic and Single-Gene Approaches

In very general terms, the SSU-rDNA tree obtained here and elsewhere shares many similarities with the transcriptome-based tree (Figure S2). The main divergences concern deep-branching nodes, which are weakly supported in SSU-rDNA trees or not at all. This concerns mostly nodes branching near the root of the tree (Organoconcha, Volnustoma, and Phryganellina are intermixed but with low supports). Similarly, the monophyly of the Longithecina is not supported. Most of these artifacts are likely caused by long-branch attraction due to the presence of particularly divergent sequences (such as, for instance, *Spumochlamys* spp.). In the case of the Longithecina, a group of species characterized by elongated shells (*Difflugia acuminata*, *D. lanceolata*, *D. oblonga*, *D. hiraethogii*, *D. parva*, and *D. bacillariarum*) does not branch with other *Difflugia* or with *Lesquereusia* but forms a separate, strongly supported group backed with a long branch. Whether these species, characterized by an elongated, cylindrical shell, form a group with *Lesquereusia* and other (pyriform-shaped) *Difflugia* such as *D. bryophila* still remains to be determined, pending transcriptomic data.

A tenth lineage, the recently described Corycidia [2], comprised of amoebae with an outer leathery cover, groups outside of Arcellinida, with the enigmatic *Trichosphaerium* in a fully supported clade that is sister to the Echinamoebidae, in Tubulinea, in accordance with another recent Amoebozoa level phylogenomic reconstruction (Figure S3 [2]). These organisms have previously been included in the Arcellinida, but the highly distinct characteristics of the shell combined with the molecular reconstruction support exclusion of the group from Arcellinida.

Implications for Interpretation of the Fossil Record

The reconstructed ancestral shells are strikingly congruent with some representatives in the fossil record (Figure 2). The framework of analysis presented here uses the powerful backbone of phylogenomic historical reconstruction based on molecular data while also yielding results in terms of morphological characters, which are comparable to fossil morphologies. The reconstruction of ancestral states uses rates of molecular evolution obtained from the phylogenetic analysis to calculate the most likely states of morphological characters in a given node,

Table 2. Averaged Results from Maximum Likelihood-Based Reconstruction of Ancestral States for All Clades of Arcellinida, with SD

	Aperture Outline	(SD)	Neck	(SD)	Shell Outline AV	(SD)	A/B	(SD)	L/B	(SD)	Composition	(SD)	Shell Outline LV	(SD)
Sphaerothecina P(0)	0,99*	0,00	0,98*	0,00	0,97*	0,00	0,33	0,00	0,33	0,00	0,33	0,01	0,74	0,01
Sphaerothecina P(1)	0,01	0,00	0,02	0,00	0,03	0,00	0,36	0,01	0,33	0,00	0,66*	0,01	0,00	0,00
Sphaerothecina P(2)							0,24	0,00	0,12	0,00	0,00	0,00	0,24	0,01
Sphaerothecina P(3)							0,07	0,00	0,23	0,00	0,00	0,00	0,02	0,00
Longithecina P(0)	1,00*	0,00	0,02	0,00	0,96*	0,00	0,32	0,00	0,02	0,00	0,02	0,00	0,01	0,00
Longithecina P(1)	0,00	0,00	0,98*	0,00	0,04	0,00	0,14	0,01	0,01	0,00	0,98*	0,00	0,00	0,00
Longithecina P(2)							0,37	0,00	0,66*	0,01	0,00	0,00	0,34	0,02
Longithecina P(3)							0,17	0,01	0,32	0,01	0,00	0,00	0,64*	0,03
Excentrostoma P(0)	0,99*	0,00	0,96*	0,00	0,93*	0,00	0,19	0,01	0,38	0,01	0,11	0,01	0,81*	0,01
Excentrostoma P(1)	0,01	0,00	0,04	0,00	0,07	0,00	0,08	0,01	0,47	0,01	0,88*	0,01	0,00	0,00
Excentrostoma P(2)							0,26	0,00	0,04	0,00	0,00	0,00	0,18	0,01
Excentrostoma P(3)							0,46	0,01	0,12	0,00	0,00	0,00	0,01	0,00
Hyalosphenidae P(0)	0,00	0,00	1,00*	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,21	0,03	0,00	0,00
Hyalosphenidae P(1)	1,00*	0,00	0,00	0,00	1,00*	0,00	0,97*	0,00	0,00	0,00	0,79*	0,03	0,00	0,00
Hyalosphenidae P(2)							0,01	0,00	0,91*	0,00	0,00	0,00	0,01	0,00
Hyalosphenidae P(3)							0,00	0,00	0,09	0,00	0,00	0,00	0,99*	0,00
Volnustoma P(0)	0,00	0,00	0,99*	0,00	0,03	0,00	0,25	0,01	0,01	0,00	0,05	0,00	0,13	0,01
Volnustoma P(1)	1,00*	0,00	0,01	0,00	0,97*	0,00	0,45	0,01	0,00	0,00	0,95*	0,00	0,00	0,00
Volnustoma P(2)							0,24	0,00	0,77*	0,01	0,00	0,00	0,78*	0,01
Volnustoma P(3)							0,06	0,00	0,22	0,00	0,00	0,00	0,09	0,01
Organoconcha P(0)	0,97*	0,00	0,92*	0,00	0,87*	0,00	0,28	0,00	0,35	0,00	0,66*	0,00	0,70*	0,01
Organoconcha P(1)	0,03	0,00	0,08	0,00	0,13	0,00	0,23	0,00	0,39	0,00	0,26	0,00	0,00	0,00
Organoconcha P(2)							0,26	0,00	0,09	0,00	0,00	0,00	0,26	0,01
Organoconcha P(3)							0,22	0,00	0,17	0,00	0,08	0,00	0,05	0,00
Organo+Glutino P(0)	0,90*	0,01	0,99*	0,00	0,79*	0,01	0,27	0,00	0,18	0,00	0,24	0,00	0,52*	0,01
Organo+Glutino P(1)	0,10	0,01	0,01	0,00	0,21	0,01	0,37	0,00	0,14	0,00	0,75*	0,00	0,00	0,00
Organo+Glutino P(2)							0,23	0,00	0,34	0,00	0,00	0,00	0,40	0,01
Organo+Glutino P(3)							0,13	0,00	0,33	0,00	0,00	0,00	0,08	0,01
Glutinoconcha P(0)	0,41	0,02	0,96*	0,00	0,49	0,01	0,26	0,00	0,09	0,00	0,12	0,00	0,26	0,01
Glutinoconcha P(1)	0,59*	0,02	0,04	0,00	0,51	0,01	0,40	0,00	0,06	0,00	0,88*	0,00	0,00	0,00
Glutinoconcha P(2)							0,22	0,00	0,52*	0,01	0,00	0,00	0,45	0,01
Glutinoconcha P(3)							0,12	0,00	0,32	0,00	0,00	0,00	0,28	0,02
Hyalo+Excentro+Longi +Sphaero P(0)	1,00*	0,00	0,86*	0,02	0,99*	0,00	0,27	0,00	0,10	0,00	0,05	0,00	0,19	0,01
Hyalo+Excentro+Longi +Sphaero P(1)	0,00	0,00	0,14	0,02	0,01	0,00	0,44	0,01	0,06	0,00	0,95*	0,00	0,00	0,00
Hyalo+Excentro+Longi +Sphaero P(2)							0,22	0,00	0,46	0,01	0,00	0,00	0,52*	0,01

(*Continued on next page*)

Table 2. *Continued*

	Aperture Outline	(SD)	Neck	(SD)	Shell Outline AV	(SD)	A/B	(SD)	L/B	(SD)	Composition	(SD)	Shell Outline LV	(SD)
Hyalo+Excentro+Longi +Sphaero P(3)							0,07	0,00	0,37	0,00	0,00	0,00	0,30	0,02
Excent+Longi+Sphaer P(0)	1,00*	0,00	0,86*	0,02	0,99*	0,00	0,31	0,00	0,25	0,01	0,06	0,00	0,54*	0,01
Excent+Longi+Sphaer P(1)	0,00	0,00	0,14	0,02	0,01	0,00	0,20	0,00	0,20	0,01	0,94*	0,00	0,00	0,00
Excent+Longi+Sphaer P(2)							0,29	0,00	0,23	0,01	0,00	0,00	0,40	0,01
Excent+Longi+Sphaer P(3)							0,21	0,01	0,32	0,01	0,00	0,00	0,06	0,01
Longi+Sphaero P(0)	1,00*	0,00	0,57*	0,06	0,97*	0,00	0,32	0,00	0,16	0,02	0,11	0,02	0,34	0,05
Longi+Sphaero P(1)	0,00	0,00	0,43	0,06	0,03	0,00	0,27	0,02	0,12	0,02	0,89*	0,02	0,00	0,00
Longi+Sphaero P(2)							0,27	0,01	0,37	0,04	0,00	0,00	0,47	0,02
Longi+Sphaero P(3)							0,14	0,02	0,35	0,00	0,00	0,00	0,18	0,04
Arcellinida P(0)	0,97*	0,00	0,99*	0,00	0,64*	0,01	0,27	0,00	0,15	0,00	0,45	0,00	0,50	0,01
Arcellinida P(1)	0,03	0,00	0,01	0,00	0,36	0,01	0,30	0,00	0,11	0,00	0,51	0,01	0,00	0,00
Arcellinida P(2)							0,25	0,00	0,41	0,00	0,00	0,00	0,40	0,01
Arcellinida P(3)							0,18	0,00	0,33	0,00	0,04	0,00	0,09	0,01

Numbers marked by asterisks (*) represent probabilities that are above 50% and more than 10% likely than the next higher probability. AV = apertural view; L/B = length to breadth; A/B = aperture to breadth; LV = lateral view

carrying the assumption that at this scale of analysis, rates of molecular and morphological evolution are reasonably comparable [38]. We have then chosen to reconstruct the morphology of hypothetical ancestors for key nodes leading to the major lineages revealed by our phylogenomic analysis (Table 2). In this manner, the comparison with fossil morphologies becomes more objective because we can explicitly compare hypothetical ancestors in well-defined nodes with this approach. For instance, the two main lineages defined here are done so based on the reconstructed ancestral shell composition: Glutinoconcha has a hypothetical ancestor with an agglutinated shell, and Organoconcha has a hypothetical ancestor with an organic shell. However, we were unable to determine the shell composition of the hypothetical Arcellinida's most recent common ancestor: analyses have shown a 51% chance for agglutinated shell and a 45% chance for an organic shell. This result accommodates both the prevailing interpretation that Tonian VSMs were originally bearing organic shells [9, 39] and also the diverging data demonstrating possible originally mineral shells [11, 13]. Most likely, distinct Tonian VSMs were belonging to distinct, already established major lineages of VSMs. Our analyses indicate that by 730 mya, at least five of the major lineages of Arcellinida had already been established (Figure 2). This realization carries an implicit departure from current interpretation of Neoproterozoic VSMs: because these morphological and genetic lineages had already diversified and were already established before the Cryogenian, we must understand these as diverging lineages that could possibly have distinct ecological niches rather than a single lineage (VSMs) encompassing distinct species with similar ecological requirements. We argue that this view potentially obfuscates the wide diversity that these fossils may represent. If the dozens of described VSMs in the Neoproterozoic can be attributed to distinct, established lineages of eukaryotes as demonstrated here for six fossils, instead of being lumped into a single category that is often interpreted as having simplified life-histories and ecology (VSMs), this may imply that the major arcellinid lineages had already been established long before the Cryogenian – though a molecular clock analysis is still necessary to determine just how long before the Cryogenian had these arcellinid lineages been established. Current molecular clock analysis places them between 1 billion years ago (bya) and 730 mya [40, 41]. Such an interpretation carries important implications, because Arcellinida are a derived lineage within Tubulinea, which is itself a derived crown group in Amoebozoa [2]. Each of the lineages recovered here represent, in modern organisms, distinct evolutionary strategies, with distinct ecologies and feeding modes: while some Sphaerothecina are bacterivores, such as smaller species in the genus *Arcella*, other Sphaerothecina (e.g., *Netzelia*), their sister group Longithecina are composed mostly of eukaryotic predators such as *Difflugia compressa*. Mixotrophy, the capacity of obtaining carbon through both predation and photosymbiosis, has emerged at least four times: *Cucurbitella mespiliformis* and *Netzelia gramen* in Sphaerothecina, *Difflugia nodosa* in Longithecina, *Hyalosphenia papilio* in the Hyalospheniidae, and *Heleopera sphagni* in Volnustoma [42]. Because the Arcellinida are composed of deeply divergent biodiversity, and the results shown here demonstrate that the main lineages were already established by the Cryogenian, Tonian VSMs must be re-interpreted under this light. A crucial

Table 3. Summarized Classification of the Arcellinida

Suborder	Infraorder	Family, Included Genera
Glutinoconcha *Subord. Nov.*	Sphaerothecina Kosakyan et al., 2016	Family Arcellidae Ehrenberg 1843, Genera: *Antarcella** Deflandre 1928, *Arcella Ehrenberg 1832*
		Family Netzeliidae Kosakyan, Lara and Lahr 2016, Genera: *Cyclopyxis* Deflandre 1929, *Netzelia* Ogden 1979
		Incertae Sedis, Genera: *Cornuapyxis** Couteaux and Chardez 1981, *Cucurbitella** Penard 1902, *Distomatopyxis** Bonnet 1964, *Ellipsopyxella** Bonnet 1975, *Ellipsopyxis** Bonnet 1965, *Geopyxella** Bonnet & Thomas 1955, *Lamptopyxis** Bonnet 1974, *Protocucurbitella** Gauthier-Lievre & Thomas 1960, *Suiadifflugia** Green 1975, *Trigonopyxis** Penard 1912
	Longithecina *Infraord. Nov.*	Family Difflugiidae Wallich 1864, Genera: *Difflugia* Leclerc 1815, *Pseudonebela** Gauthier-Lievre 1953
		Family Lesquereusiidae Jung 1942, Genera: *Lesquereusia* Schlumberger 1845, *Pomoriella** Golemansky 1970, *Paraquadrula** Deflandre 1932, *Microquadrula** Golemansky 1968
	Excentrostoma *Infraord. Nov.*	Family Centropyxidae Jung 1942, Genera: *Centropyxis* Stein 1857, *Proplagiopyxis** Schonborn 1964
		Family Plagiopyxidae Bonnet & Thomas 1960, Genera: *Bullinularia** Deflandre 1953, *Geoplagiopyxis** Chardez 1961, *Protoplagiopyxis** Bonnet 1962, *Paracentropyxis** Bonnet 1960, *Plagiopyxis** Penard 1910, *Hoogenraadia** Gauthier-Lievre &Thomas 1958, *Planhoogenraadia** Bonnet 1977
		Incertae Sedis Excentrostoma, Genera: *Conicocassis** Nasser & Patterson 2015, *Oopyxis** Jung 1942
	Hyalospheniformes *Infraord. Nov.*	Family Hyalospheniidae Schulze 1977, Genera: *Alabasta* Duckert et al., 2018, *Alocodera* Jung 1942, *Apodera* Loeblich & Tappan 1961, *Certesella* Loeblich & Tappan 1961, *Cornutheca* Kosakyan et al., 2016, *Gibbocarina* Kosakyan et al., 2016, *Hyalosphenia* Stein 1859, *Longinebela* Kosakyan et al., 2016, *Mrabella* Kosakyan et al., 2016, *Nebela* Leidy 1874, *Padaungiella* Lara et Todorov 2012, *Planocarina* Kosakyan et al., 2016, *Porosia* Jung 1942, *Quadrulella* Cockerell 1909
	Volnustoma *Infraord. Nov.*	Family Heleoperidae Jung 1942, Genus *Heleopera* Leidy 1879
Organoconcha *Subord. Nov.*		Family Microchlamyiidae Ogden 1985, Genera: *Microchlamys* Cockerell 1911, *Spumochlamys* Kudryavtsev & Hausmann 2007, *Pyxidicula* Ehrenberg 1838
Phryganellina Bovee 1985		Family Phryganellidae Jung 1942, Genus *Phryganella* Penard 1902
		Family Cryptodiflugiidae Jung 1942, Genera: *Cryptodifflugia* Penard 1890, *Meisterfeldia* Bobrov 2016, *Wailesella* Deflandre 1928
Incertae Sedis Arcellinida		Genera: *Argynnia** Vucetich 1974, *Awerintzewia** Schouteden 1906, *Geamphorella** Bonnet 1959, *Jungia** Loeblich and Tappan 1961, *Lagenodifflugia** Medioli & Scott 1983, *Lamtoquadrula** Bonnet 1974, *Leptochlamys** West 1901, *Maghrebia** Gauthier-Lievre & Thomas 1960, *Pentagonia** Gauthier-Lievre & Thomas 1960, *Physochila** Jung 1942, *Pontigulasia** Rhumbler 1896, *Pseudawerintzewia** Bonnet 1959, *Schoenbornia** Decloitre 1964, *Schwabia** Jung 1942, *Sexangularia** Awerintzew 1906, *Zivkovicia** Ogden 1987

Each Linnean rank is demonstrated here with included taxa. Taxa marked with an asterisk have not yet been sampled for molecular data. See also Data S1.

question that is now evident is whether these vastly distinct ecological modes found in the modern representatives of Arcellinida lineages were already present in the Neoproterozoic representatives or emerged only later.

A hypothesis previously put forth claims that the deeper oxygenation of the oceans in the mid-Neoproterozoic followed by emergence of predatory habits (including predation by animals) generated a burst of eukaryotic diversification in the late-Neoproterozoic [4, 43–46]. While accumulating geochemical evidence indicates that the deeper oxygenation was not as steep as previously thought [47] and may have taken up to 100 million years to occur [48] even minute changes in redox conditions of the atmosphere are likely to have generated the same result. Whether providing protection from predation, or conversely a more efficient predation tool (many modern arcellinids feed on other eukaryotes, exhibiting complex behavior such as pack-hunting [49], or using the shell as a type of "lever" to open up prey [34]), shells have played a significant role in the diversification of microbial eukaryotes, and here we have highlighted some details of this important morphological feature for one of the main shelled groups. Our analysis may be able to add texture to the hypothesis of an "earlier-than-Tonian" diversification of eukaryotes (whatever the geochemical driver may have been, an interesting hypothesis based on the capabilities of eukaryotes to live in anoxic environments was recently presented in [50]), and puts it within a novel framework that enables testable taxonomic placement of microfossils.

We conclude that the crown-group of arcellinids must have been established before the Tonian. Just how long before the Tonian record reveals them, remains an open question. We infer that the diversification of crown Arcellinida was contemporary to the oxygenation of deep parts of oceans in the mid-Neoproterozoic (even if slight [47, 48]), and may potentially have been driven by it. Consequently, the divergences that led to the establishment of the major lineages of eukaryotes must have happened before the oxygenation of oceans, as previously suggested [44, 51]. Our conclusions offer support to the idea that diversification of the major lineages of eukaryotes happened long before the Neoproterozoic/Cambrian boundary, and thus the actual geochemical drivers (if any) for the earliest eukaryotic diversification that gave rise to the major groups (Amoebozoa, Archaeplastida, Excavata, Opisthokonta, and SAR) remain unknown.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Biological Samples
- METHOD DETAILS
  - RNA extraction and Sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatics pipeline and Phylogenetic Reconstructions
  - Ancestral State Reconstructions
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://doi.org/10.1016/j.cub.2019.01.078.

ACKNOWLEDGMENTS

We are thankful to three anonymous reviewers for their excellent critiques during peer-review, as well as Andy Knoll for comments on an early version of this manuscript; Daniel M. Alcântara for technical assistance in the early stages of NextGen sequencing; the Core Facility for Scientific Research—University of São Paulo (CEFAP-USP/GENIAL) for sequencing in the MiSeq and NextSeq platforms; and Mississippi State University's High Performance Computing Collaboratory for computational resources. Access to fossil specimens in the Scientific Palaontological collection at the Institute of Geosciences was kindly granted by Dr. Juliana M.L. Basso, and access to fossil specimens in the collection at the Earth Science Department, University of California Santa Barbara was kindly granted by Dr. Susannah Porter. This project was supported by FAPESP grant 2013/04585-3 awarded to D.J.G.L., 2013/12852-1 to L.M., 2013/25729-3 to A.L.P.-S., and 2015/02689-1 to G.M.R.; National Science Foundation (NSF) Division of Environmental Biology (DEB) grant 1456054 and Initiation Grant from the Henry Family Research Fund at the College of Arts & Sciences at Mississippi State University awarded to M.W.B.; Comunidad de Madrid grant 2017-T1/AMB-5210 and Contrato Intramural para Doctores, CSIC (201730E063) to E.L.; Swiss NSF early postdoc mobility grant P2NEP3-155421 to A.K.; and J.W. Fulbright Commission of Czech Republic and European Report on Development Funds (OPVVV 16-019/0000759) to T.P.

AUTHOR CONTRIBUTIONS

Conceptualization, Supervision, Project Administration and Funding Acquisition, Visualization, Methodology, Writing – Original Draft, D.J.G.L. and M.W.B.; Resources, D.J.G.L., E.A.D.M., E.L. and M.W.B.; Data Generation, D.J.G.L., M.W.B., A.K., L.M., A.L.P.-S., and G.M.R.; Formal Analysis and Investigation, D.J.G.L., M.W.B., A.K.T., S.K., T.P., E.A.D.M., and E.L.; Writing – Review Editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 21, 2018
Revised: October 26, 2018
Accepted: January 30, 2019
Published: February 28, 2019

REFERENCES

1. Simpson, A., Slamovits, C.H., and Archibald, J.M. (2017). Protist diversity and Eukaryote phylogeny (Springer).
2. Kang, S., Tice, A.K., Spiegel, F.W., Silberman, J.D., Pánek, T., Čepička, I., Kostka, M., Kosakyan, A., Alcântara, D.M.C., Roger, A.J., et al. (2017). Between a pod and a hard test: the deep evolution of amoebae. Mol. Biol. Evol. *34*, 2258–2270.
3. Krabberød, A.K., Orr, R.J.S., Bråte, J., Kristensen, T., Bjørklund, K.R., and Shalchian-Tabrizi, K. (2017). Single cell transcriptomics, mega-phylogeny, and the genetic basis of morphological innovations in Rhizaria. Mol. Biol. Evol. *34*, 1557–1573.
4. Porter, S.M. (2016). Tiny vampires in ancient seas: Evidence for predation via perforation in fossils from the 780–740 million-year-old Chuar Group, Grand Canyon, USA. Proc R Soc B, 283(1831), 20160221.
5. Kosakyan, A., Gomaa, F., Lara, E., and Lahr, D.J.G. (2016). Current and future perspectives on the systematics, taxonomy and nomenclature of testate amoebae. Eur. J. Protistol. *55* (Pt B), 105–117.
6. Kajukało, K., Fiałkiewicz-Kozieł, B., Gałka, M., Kołaczek, P., and Lamentowicz, M. (2016). Abrupt ecological changes in the last 800 years inferred from a mountainous bog using testate amoebae traits and multi-proxy data. Eur. J. Protistol. *55* (Pt B), 165–180.
7. Payne, R.J., Babeshko, K.V., Van Bellen, S., Blackford, J.J., Booth, R.K., Charman, D.J., Ellershaw, M.R., Gilbert, D., Hughes, P.D.M., Jassey, V.E.J., et al. (2016). Significance testing testate amoeba water table reconstructions. Quat. Sci. Rev. *138*, 131–135.
8. Seppey, C.V., Fournier, B., Szelecz, I., Singer, D., Mitchell, E.A., and Lara, E. (2016). Response of forest soil euglyphid testate amoebae (Rhizaria: Cercozoa) to pig cadavers assessed by high-throughput sequencing. Int. J. Legal Med. *130*, 551–562.
9. Porter, S.M., and Knoll, A.H. (2000). Testate amoebae in the neoproterozoic era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. Paleobiology *26*, 360–385.
10. Cohen, P.A., Irvine, S.W., and Strauss, J.V. (2017). Vase-shaped microfossils from the Tonian Callison Lake Formation of Yukon, Canada: taxonomy, taphonomy and stratigraphic palaeobiology. Palaeontology *60*, 683–701.
11. Morais, L., Fairchild, T.R., Lahr, D.J.G., Rudnitzki, I.D., Schopf, J.W., Garcia, A.K., Kudryavtsev, A.B., and Romero, R.R. (2017). Carbonaceous and siliceous Neoproterozoic vase-shaped microfossils (Urucum Formation, Brazil) and the question of early protistan biomineralization. J. Paleontol. *91*, 393–406.
12. Riedman, L.A., Porter, S.M., and Calver, C.R. (2017). Vase-shaped microfossil biostratigraphy with new data from Tasmania, Svalbard, Greenland, Sweden and the Yukon. Precambrian Res. *319*, 19–36.
13. Bosak, T., Lahr, D., Pruss, S., Macdonald, F., Dalton, L., and Matys, E. (2011). Agglutinated tests in post-sturtian cap carbonates of Namibia and Mongolia. Earth Planet. Sci. Lett. *308*, 29–40.
14. Barber, A., Siver, P.A., and Karis, W. (2013). Euglyphid testate amoebae (rhizaria: euglyphida) from an arctic eocene waterbody: evidence of evolutionary stasis in plate morphology for over 40 million years. Protist *164*, 541–555.
15. Delaye, L., Valadez-Cano, C., and Pérez-Zamorano, B. (2016). How really ancient is *Paulinella chromatophora?* PLoS Curr. *8*, https://doi.org/10.1371/currents.tol.e68a099364bb1a1e129a17b4e06b0c6b.
16. Pawlowski, J., Holzmann, M., and Tyszka, J. (2013). New supraordinal classification of Foraminifera: Molecules meet morphology. Mar. Micropaleontol. *100*, 1–10.
17. Theriot, E.C., Ashworth, M.P., Nakov, T., Ruck, E., and Jansen, R.K. (2015). Dissecting signal and noise in diatom chloroplast protein encoding genes with phylogenetic information profiling. Mol. Phylogenet. Evol. *89*, 28–36.
18. Parks, M.B., Wickett, N.J., and Alverson, A.J. (2018). Signal, uncertainty, and conflict in phylogenomic data for a diverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). Mol. Biol. Evol. *35*, 80–93.

19. Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A.A., Lang, B.F., Simpson, A.G.B., and Roger, A.J. (2013). Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. Proc. Biol. Sci. *280*, 20131755.

20. Singer, D., Kosakyan, A., Pillonel, A., Mitchell, E.A., and Lara, E. (2015). Eight species in the *Nebela collaris* complex: *Nebela gimlii* (Arcellinida, Hyalospheniidae), a new species described from a Swiss raised bog. Eur. J. Protistol. *51*, 79–85.

21. Lara, E., Roussel-Delif, L., Fournier, B., Wilkinson, D.M., and Mitchell, E.A. (2016). Soil microorganisms behave like macroscopic organisms: patterns in the global distribution of soil euglyphid testate amoebae. J. Biogeogr. *43*, 520–532.

22. Smirnov, A.V., Chao, E., Nassonova, E.S., and Cavalier-Smith, T. (2011). A revised classification of naked lobose amoebae (Amoebozoa: lobosa). Protist *162*, 545–570.

23. Cavalier-Smith, T., Chao, E.E., and Lewis, R. (2016). 187-gene phylogeny of protozoan phylum Amoebozoa reveals a new class (Cutosea) of deep-branching, ultrastructurally unique, enveloped marine Lobosa and clarifies amoeba evolution. Mol. Phylogenet. Evol. *99*, 275–296.

24. Anderson, O.R. (2018). A half-century of research on free-living amoebae (1965-2017): Review of biogeographic, ecological and physiological studies. Acta Protozool. *57*, 1.

25. Gomaa, F., Todorov, M., Heger, T.J., Mitchell, E.A., and Lara, E. (2012). SSU rRNA phylogeny of Arcellinida (Amoebozoa) reveals that the largest Arcellinid genus, *Difflugia* Leclerc 1815, is not monophyletic. Protist *163*, 389–399.

26. Lahr, D.J., Grant, J.R., and Katz, L.A. (2013). Multigene phylogenetic reconstruction of the Tubulinea (Amoebozoa) corroborates four of the six major lineages, while additionally revealing that shell composition does not predict phylogeny in the Arcellinida. Protist *164*, 323–339.

27. Rooney, A.D., Austermann, J., Smith, E.F., Li, Y., Selby, D., Dehler, C.M., Schmitz, M.D., Karlstrom, K.E., and Macdonald, F.A. (2017). Coupled Re-Os and U-Pb geochronology of the Tonian Chuar Group, Grand Canyon. Geol. Soc. Am. Bull. *130*, 1085–1098.

28. Sergeev, V.N., and Schopf, J.W. (2010). Taxonomy, paleoecology and biostratigraphy of the late Neoproterozoic Chichkan microbiota of South Kazakhstan: The marine biosphere on the eve of metazoan radiation. J. Paleontol. *84*, 363–401.

29. Meisterfeld, R. (2002). Order Arcellinida Kent, 1880. In The Illustrated Guide to the Protozoa, Volume 2, J.J. Lee, ed. (ISOP), pp. 2827–2860.

30. Gomaa, F., Lahr, D.J.G., Todorov, M., Li, J., and Lara, E. (2017). A contribution to the phylogeny of agglutinating Arcellinida (Amoebozoa) based on SSU rRNA gene sequences. Eur. J. Protistol. *59*, 99–107.

31. Pagel, M., Meade, A., and Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. *53*, 673–684.

32. Porter, S.M., Meisterfeld, R., and Knoll, A.H. (2003). Vase-shaped micro-fossils from the Neoproterozoic Chuar Group, Grand Canyon: a classification guided by modern testate amoebae. J. Paleontol. *77*, 409–429.

33. Bloeser, B. (1985). *Melanocyrillium*, a new genus of structurally complex late Proterozoic microfossils from the Kwagunt Formation (Chuar Group), Grand Canyon, Arizona. J. Paleontol. *59*, 741–765.

34. Dumack, K., Kahlich, C., Lahr, D.J.G., and Bonkowski, M. (2018). Reinvestigation of *Phryganella paradoxa* (Arcellinida, Amoebozoa) Penard 1902. J. Eukaryot. Microbiol. *0*, 1–12.

35. Bovee, E. (1985). Class Lobosea Carpenter, 1861. In An Illustrated Guide to the Protozoa, J.J. Lee, S.H. Hutner, and E.C. Bovee, eds. (Kansas: ISOP), pp. 158–211.

36. Lara, E., Heger, T.J., Mitchell, E.A., Meisterfeld, R., and Ekelund, F. (2007). SSU rRNA reveals a sequential increase in shell complexity among the euglyphid testate amoebae (Rhizaria: Euglyphida). Protist *158*, 229–237.

37. Kosakyan, A., Lahr, D.J., Mulot, M., Meisterfeld, R., Mitchell, E.A., and Lara, E. (2016). Phylogenetic reconstruction based on COI reshuffles the taxonomy of Hyalosphenid shelled (testate) amoebae and reveals the convoluted evolution of shell plate shapes. Cladistics *32*, 606–623.

38. Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature *401*, 877–884.

39. Mus, M.M., and Moczydowska, M. (2000). Internal morphology and taphonomic history of the Neoproterozoic vase-shaped microfossils from the Visings Group, Sweden. Nor. Geol. Tidsskr. *80*, 213–228.

40. Eme, L., Sharpe, S.C., Brown, M.W., and Roger, A.J. (2014). On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. Cold Spring Harb. Perspect. Biol. *6*, a016139.

41. Parfrey, L.W., Lahr, D.J., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. USA *108*, 13624–13629.

42. Lara, E., and Gomaa, F. (2017). Symbiosis between testate amoebae and photosynthetic organisms. In Algal and cyanobacteria symbioses (World Scientific), pp. 399–419.

43. Knoll, A.H. (2014). Paleobiological perspectives on early eukaryotic evolution. Cold Spring Harb. Perspect. Biol. *6*, a016121.

44. Butterfield, N.J. (2015). Early evolution of the Eukaryota. Palaeontology *58*, 5–17.

45. Cohen, P.A., and Riedman, L.A. (2018). It's a protist-eat-protist world: recalcitrance, predation, and evolution in the Tonian–Cryogenian ocean. Emerging Topics in Life Sciences *2*, 173–180.

46. Loron, C.C., Rainbird, R.H., Turner, E.C., Greenman, J.W., and Javaux, E.J. (2018). Implications of selective predation on the macroevolution of eukaryotes: evidence from Arctic Canada. Emerging Topics in Life Sciences *2*, 247–255.

47. Sperling, E.A., Wolock, C.J., Morgan, A.S., Gill, B.C., Kunzmann, M., Halverson, G.P., Macdonald, F.A., Knoll, A.H., and Johnston, D.T. (2015). Statistical analysis of iron geochemical data suggests limited late Proterozoic oxygenation. Nature *523*, 451–454.

48. Pogge von Strandmann, P.A., Stüeken, E.E., Elliott, T., Poulton, S.W., Dehler, C.M., Canfield, D.E., and Catling, D.C. (2015). Selenium isotope evidence for progressive oxidation of the Neoproterozoic biosphere. Nat. Commun. *6*, 10157.

49. Geisen, S., Rosengarten, J., Koller, R., Mulder, C., Urich, T., and Bonkowski, M. (2015). Pack hunting by a common soil amoeba on nematodes. Environ. Microbiol. *17*, 4538–4546.

50. Porter, S.M., Agić, H., and Riedman, L.A. (2018). Anoxic ecosystems and early eukaryotes (Emerging Topics in Life Sciences).

51. Lenton, T.M., Boyle, R.A., Poulton, S.W., Shields-Zhou, G.A., and Butterfield, N.J. (2014). Co-evolution of eukaryotes and ocean oxygenation in the Neoproterozoic era. Nat. Geosci. *7*, 257.

52. Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

53. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494–1512.

54. Tice, A.K., Shadwick, L.L., Fiore-Donno, A.M., Geisen, S., Kang, S., Schuler, G.A., Spiegel, F.W., Wilkinson, K.A., Bonkowski, M., Dumack, K., et al. (2016). Expansion of the molecular and morphological diversity of Acanthamoebidae (Centramoebida, Amoebozoa) and identification of a novel life cycle type within the group. Biol. Direct *11*, 69.

55. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

56. Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. *10*, 210.

57. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

58. Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A.K., Shiratori, T., Ishida, K.I., Hashimoto, T., Simpson, A.G.B., and Roger, A.J. (2018). Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. Genome Biol. Evol. *10*, 427–433.

59. Mai, U., and Mirarab, S. (2017). Treeshrink: efficient detection of outlier tree leaves. Comparative Genomics, 116–140.

60. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

61. Wang, H.C., Minh, B.Q., Susko, E., and Roger, A.J. (2017). Modeling site hetero-geneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. *67*, 216–235.

62. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. *62*, 611–615.

63. Swofford, D.L., and Sullivan, J. (2003). Phylogeny inference based on parsimony and other methods using PAUP*. In The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny, pp. 160–206.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Living Amoebae Samples	See Table S1	N/A
Limeta lageniformis fossil	Scientific Palaeontological collection, Institute of Geosciences, University of São Paulo, Brazil.	Slide GP/5T:2529F
Palaeoamphora urucumense fossil	Scientific Palaeontological collection, Institute of Geosciences, University of São Paulo, Brazil.	Slide GP/5T-2534F
Taruma rata fossil	Scientific Palaeontological collection, Institute of Geosciences, University of São Paulo, Brazil.	Slide GP/5T 2533 B
Cycliocyrillium torquata fossil	Earth Science Department collection, University of California Santa Barbara, USA	Slides J1204-16.8 and F930-15.5
Mellanocyrillum hexodiadema fossil	Earth Science Department collection, University of California Santa Barbara, USA	Slide F930-15.5
Chemicals, Peptides, and Recombinant Proteins		
TRI-Reagent	Sigma	Catalog #T9424
NEXTERA-XT	Illumina	Catalog #FC-131-1096
Deposited Data		
Raw Sequences	https://www.ncbi.nlm.nih.gov/bioproject/513164	Bioproject : PRJNA380424

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Daniel J. G. Lahr (dlahr@ib.usp.br).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Biological Samples

We isolated live testate amoebae from natural samples in Brazil, Switzerland, and USA, using inverted microscopes and standard protistology pipetting techniques. In sum, samples are freshly collected and isolated under an inverted microscope. Amoebae to be processed by single-cell methods are then cleaned through serial dilution in filtered and autoclaved mineral water. Amoebae to be cultured are isolated and cleaned as above, and inoculated into a sterile vial containing mineral water and 1:100 solution of cereal grass media. These techniques are described in more detail [26], details and coordinates of sampled environments are in Table S1. We strategically sampled representative species for the major morphological forms of Arcellinida. All species were identified in comparison to original descriptive literature. All species were vouchered by light microscopy, several by scanning electron microscopy, and we further generated SSU-rDNA trees to confirm identity.

METHOD DETAILS

RNA extraction and Sequencing

Testate amoebae were either cultured or processed with single-cell RNA extraction. Cultured samples (*Arcella intermedia, Cryptodifflugia operculata,* and *Pyxidicula operculata*) were allowed to grow until a density of at least 5,000 individuals was achieved. Cultures were then cleaned, concentrated and processed using total RNA extraction standard method (TRI-reagent, Sigma), as described in [2]. *Difflugia compressa* was also subjected to total RNA extraction following the same procedures, however, the original sample was made of approximately 400 individuals isolated from the environment, cleaned, and starved overnight. All other samples were subjected to single-cell RNA extraction and amplification using the SMARTSEQ2 method [52], as previously described in [2]. Immediately before single-cell library preparation, each individual cell was photodocumented, generating a voucher image of the exact sequenced individual (Figure S1). We prepared libraries from RNA samples for all species using NEXTERA-XT, and sequenced using either MiSeq or HiSeq platforms (details in Table S1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatics pipeline and Phylogenetic Reconstructions

Raw sequences were cleaned and trimmed for quality using TRIMMOMATIC, then contigs were assembled using Trinity [53]. Nucleotide sequences were translated with Transdecoder and putative orthologous gene sequences from the dataset of Kang and collaborators [2] were acquired from each predicted proteome. All proceeding steps were performed as in [2]. The phylogenomic dataset of 324 genes was constructed as developed in [54] and implemented in [2]. A reference dataset of 324 aligned proteins described in [2] was used as the starting point for the current analysis, from which data from testate amoebozoans as well as our new data generated from our efforts here were selected. Extensive efforts were made to exclude contamination and paralogs. The dataset was constructed from orthologs identified as both highly transcribed (and therefore likely to be present in RNaseq data) and globally distributed across the eukaryotic tree [2]. For each of the 324 orthologs of interest, a representative sequence was used, most often from *Arabidopsis thaliana* or *Homo sapiens*, as queries for tBlastN or BlastP approaches (Data S2, Reference Ortholog Queries). Potential homolog sequences from our novel RNaseq data and other publically available data (see Data S2) were identified using a threshold e-value of 1e-10. From these putative orthologs, we performed BlastP against the OrthoMCL v. 5.0 database and obtained all sequences matched below a threshold e-value of 1e-10. The candidate sequences that matched the correct OrthoMCL ortholog ID, and which did not correspond to prokaryotic sequences, were designated as putative orthologs. These putative orthologs were added to the existing protein alignments that contained sequences from a set of taxa representing all major eukaryotic lineages. The resultant gene clusters were re-aligned using MAFFT-Linsi [55]. Ambiguously aligned sequences were trimmed and culled using BMGE [56]. We obtained individual maximum-likelihood trees from these orthologs using RAxML v. 8.0 [57] under the LG model+gamma distribution of rate heterogeneity, with 4 discrete gamma rate classes (LG+GAMMA), under the LG model + gamma distribution of rate heterogeneity, with 4 discrete gamma rate classes (LG+GAMMA). Each tree was ML bootstrapped (MLBS) by 100 pseudoreplicates. A set of custom python scripts was used to 1) annotate each sequence based on its top hit in RefSeq non-redundant protein database (release 87); 2) compare MLBS values to a consensus tree of well-supported eukaryotic groupings to identify paralogy and contaminations that were supported by bootstrap value above 70%. Outputs of these scripts were plotted onto single gene trees with branches color-coded by the major eukaryotic assemblages they belong to [19, 58]. Trees were then examined individually by eye, all potential contaminants and paralogs were removed. Sequences shorter than 30% of the trimmed alignment were also removed [59]. 250 genes that did not show any level of paralogy and were present in at least 22 of the 36 taxa used for the final analyses were kept. The individual genes from those taxa were re-aligned, re-trimmed and concatenated by the custom pipeline known as ORTHOLAGER [2]. This resulted in a final supermatrix of 85,595 amino acid sites. Maximum likelihood (ML) trees were inferred using IQ-Tree v. 1.5.5 [60]. The best-fitting available model based on the Akaike Information Criterion (AIC) was the LG+C60+F+GAMMA site heterogenous mixture model with class weights optimized from the dataset and four discrete gamma categories. ML trees were estimated under this model for dataset. We then used this model and best ML tree under the LG+C60+F+GAMMA model to estimate the 'posterior mean site frequencies' (PMSF) model [61] from the dataset. This LG+C60+F+GAMMA PMSF model was used to re-estimate ML trees, and for a real bootstrap analysis of the dataset, with 500 pseudoreplicates (Figure 2). Bayesian inferences were performed using Phylobayes-MPI v1.6j [62], under the CAT-GTR+GAMMA model, with four discrete GAMMA categories. A total of six independent Markov chain Monte Carlo chains were run for 8,800 generations, sampling every second generation. Two of the chains converged (at 4,040 generations, which was used as the burnin), with the largest discrepancy in posterior probabilities (PPs) (maxdiff) = 0. The topology of the converged chains has the same backbone topology as the ML tree above with minor differences in terminal branches, and it fully supported with Bayesian posterior probabilities of 1 for all nodes in that tree (Figure S3). We have additionally generated a ML tree of the SSU rDNA gene, for comparative reasons with the previous literature (Figure S2).

Ancestral State Reconstructions

We performed Maximum Likelihood ancestral state reconstructions with the program Multistate [31, 38] as part of the package BayesTraits (available at http://www.evolution.rdg.ac.uk/BayesTraits.html), strictly following the provided guidelines given by the authors in the documentation package. Analyses were run in parallelized virtual machines. We fixed format issues and performed re-rooting either by hand or using PAUP* [63]. We coded morphological characters broadly, using the 7 most relevant characteristics for identification of testate amoebae (Table 1). We first generated 500 boot-strapped trees in IQTree, with branch lengths calculated as explained above [60]. For the ancestral state reconstruction, it is important to determine how many variables are appropriate to include in the model. We made a reconstruction where all variables were allowed to vary freely (each variable represents the probability of character change between each of the four possible states, non-reversible). We then made reconstructions by restricting the reciprocal pairs of variables individually - i.e., we asked the question whether we could represent the chance of a character going from 1 to 2 as the same if it was going from 2 to 1. The likelihoods resulting from these reconstructions were compared against each other using a likelihood ratio test, the resulting values were matched against a chi-square distribution to determine p values, this allowed the use of degrees of freedom in the comparison, hence penalizing models with more parameters. The order of transformation never made a significant difference in any of the cases tested, thus we have chosen to restrict all reciprocal pairs,

hence minimizing the number of free parameters in our model. The resulting data was then pooled together and the results of the 1,000 iterations were averaged for interpretation of results and figure-making (Table 2).

DATA AND SOFTWARE AVAILABILITY

Raw sequencing files are deposited at NCBI SRA repository under the Bioproject PRJNA513164. Alignments and trees of individual genes used to generate the phylogenomic analysis, intermediate files from ancestral reconstruction analysis as well as custom Python scripts used to identify paralogy and contamination are available upon request.