Amoebozoan testate amoebae illuminate the diversity of heterotrophs and the complexity of ecosystems throughout geological time

Alfredo L. Porfirio-Sousa[1,2]*, Alexander K. Tice[2,3]*, Luana Morais[4,5], Giulia M. Ribeiro[1], Quentin Blandenier[2], Kenneth Dumack[6], Yana Eglit[7,8,9], Nicholas W. Fry[2], Maria Beatriz Gomes E Souza[1], Tristan Henderson[2], Felicity Kleitz-Singleton[2], David Singer[10], Matthew W. Brown[2]**, Daniel J. G. Lahr[1]**

[1]Department of Zoology, Institute of Biosciences, University of São Paulo, São Paulo, Brazil
[2]Department of Biological Sciences, Mississippi State University, Starkville, Mississippi, USA
[3]Department of Biological Sciences, Texas Tech University, Lubbock 79409, Texas
[4]Department of Geophysics, Institute of Astronomy, Geophysics and Atmospheric Sciences, University of São Paulo, São Paulo, Brazil
[5]Institute of Geosciences, São Paulo State University, Rio Claro, Brazil
[6]Terrestrial Ecology, Institute of Zoology, University of Cologne, Germany
[7]Department of Biology, Dalhousie University, Halifax, NS, Canada
[8]Institute for Comparative Genomics, Dalhousie University, Halifax, NS, Canada
[9]Department of Biology, University of Victoria, Victoria, BC, Canada
[10]Changins, Soil Science and Environment Group, Nyon, Switzerland

* Contributed equally

** Corresponding authors

Email: dlahr@ib.usp.br and matthew.brown@msstate.edu

Author Contributions: Conceptualization, Supervision, Project Administration, Visualization, Methodology, Writing – Original Draft, A.L.P.-S, A.K.T., M.W.B., and D.J.G.L.; Funding Acquisition and Resources, D.J.G.L. and M.W.B.; Sampling and Culture Maintenance, A.L.P.-S, A.K.T., Q.B., K.D., Y.E., M.B.G.S, T.H., F.K.-S, M.W.B., and D.J.G.L; Data Generation, A.L.P.-S, A.K.T.; L.M.; G.M.R., Q.B., K.D., N.F., T.H., D.S., M.W.B., and D.J.G.L.; Phylogenetic Matrices Construction, A.L.P.-S, A.K.T., Q.B., N.F., T.H.; F.K.-S, M.W.B., and D.J.G.L.; Formal Analysis and Investigation, A.L.P.-S, A.K.T.; L.M.; M.W.B., and D.J.G.L.; Review Editing, all authors.

Competing Interest Statement: The authors declare no competing interest.

Deposited as a preprint: bioRxiv, CC BY

Classification: Biological Sciences, Evolution

Keywords: arcellinida; vase-shaped microfossils; phylogenomics; ancestral state reconstruction; Arcellinida classification; eukaryotic evolution;

This PDF file includes:

- Main Text
- Data availability
- Acknowledgements
- References
- Figures 1 to 5

Abstract

Heterotrophic microbial eukaryotes play a pivotal role in aquatic and terrestrial ecosystems, contributing to carbon and nutrient cycles. These microorganisms, capable of phagocytosis, act as predators of bacterial communities and other microeukaryotes, occupying a significant position in complex food webs. The timing of the origin and diversification of heterotrophic microeukaryotes remain unclear. Fossil evidence and molecular data suggest that the emergence of predatory microeukaryotes and the transition to a eukaryote-dominant marine environment occurred around 800 million years ago (mya). Vase-shaped microfossils (VSMs), the fossil record linked to Arcellinida testate amoebae, represent the oldest known evidence of heterotrophic microeukaryotes in marine environments and terrestrial habitats. In this study, we investigate the early divergence and diversification of Arcellinida and related amoebozoan taxa using a relaxed molecular clock approach. Phylogenomic reconstructions reveal a well-resolved tree of amoebozoan testate amoebae, including a monophyletic Arcellinida with three suborders and five infraorders. Through calibration using fossils and rigorous clock models, we estimate the timing of diversification of Arcellinida during the early Neoproterozoic (886 - 764 mya), shedding light on the expansion of life during this period. Our results suggest an established biological complexity in shallow marine ecosystems, involving both phototrophic and heterotrophic microeukaryotes during the Neoproterozoic, followed by an invasion of freshwater systems potentially during Cryogenian and subsequent diversification of Arcellinida in the Phanerozoic. Overall, these findings provide valuable insights into heterotrophic microeukaryotes' evolutionary history and ecological significance in Earth's ecosystems.

Significance Statement

Arcellinida shelled amoebae are heterotrophic microbial eukaryotes with an extensive Neoproterozoic fossil record, the vase-shaped microfossils (VSMs). Here we generated time-calibrated trees by combining additional phylogenomic sampling of the extant group and the VSM fossil record. Our results bring to light key events in the history of microeukaryotes, including: i) the diversification of heterotrophic eukaryotes in marine ecosystems alongside red-algae in the Precambrian; ii) a

Cryogenian establishment of complex terrestrial habitats; iii) a marine mass extinction during late Devonian, and; iv) a post-Silurian explosion of life on terrestrial (including freshwater) habitats. These findings corroborate recent discoveries regarding Precambrian history of eukaryotic protosteranes. Our results provide valuable insights into the evolution of life in marine and terrestrial ecosystems throughout geological time.

Introduction

Heterotrophic microbial eukaryotes play a crucial role in today's aquatic and terrestrial ecosystems by contributing to the carbon and nutrient cycles of the biosphere (Geisen et al., 2018; Singer et al., 2021). These microorganisms, capable of obtaining their food through phagocytosis (engulfment of large particles), act as predators on bacterial communities (bacterivory) as well as phototrophic and other heterotrophic eukaryotes (eukaryovory) (Geisen et al., 2018). As a result, they occupy a significant position in complex food webs, supported by a community of primary producers that fix carbon. Additionally, predation and eukaryovory are evolutionary innovations that have likely contributed to the diversification of eukaryotes (Cohen and Riedman, 2018). While the Last Eukaryotic Common Ancestor (LECA), which lived around two billion years ago, may have been heterotrophic and capable of phagocytosis, the timing and specific conditions under which heterotrophic microeukaryotes have proliferated in Earth's ecosystems remain unclear (Zachar and Szathmáry, 2017; Eme and Ettema 2018; Porter, 2020).

Evidence from fossils, biomarkers, geochemical proxies, genomic data, and molecular clocks indicate that the diversification of eukaryotes occurred during the late Mesoproterozoic Stenian period (1200-1000 Ma) and early Neoproterozoic Tonian period (1000-720 Ma) (Xiao et al., 2018; Li et al., 2020; Tang et al., 2020; Del Cortona et al., 2020; Brocks et al., 2023). This led to a transition from a prokaryotic-dominant to a eukaryotic-dominant marine environment (Porter, 2020; Cohen and Kodner, 2022; Mills et al., 2022; Kang et al., 2023), which occurred around 800 million years ago, likely triggered by increased phosphorus, nitrate, and silica availability and other favorable geochemical and biological conditions (Siever, 1992; Kang et al., 2023; Reinhard et al., 2017; Reinhard et al., 2020). From around this time, Neoproterozoic vase-shaped microfossils (VSMs) represent the remnants of early eukaryotic diversification in the marine fossil record (Porter and Knoll, 2000;

Porter et al., 2003; Morais et al., 2019). Morphological interpretations, supported by ancestral morphological reconstructions, assign many of these VSMs as early members of the Arcellinida testate amoebae group, a diverse lineage of extant heterotrophic microeukaryotes within the Amoebozoa (Porter et al., 2003; Morais et al., 2019; Lahr et al., 2019; Riedman and Porter, 2019). These fossils are abundant and widespread in sedimentary deposits with well-resolved stratigraphies. Since VSMs have been recognized as arcellinids, they are well-accepted to represent the oldest and most diverse fossil evidence of heterotrophic microeukaryotes in marine environments (Porter et al., 2003; Morais et al., 2019; Mus and Moczydowska, 2020; Lahr, 2021). Interestingly, fossils described as arcellinids also represent the earliest direct fossil evidence of phagotrophic microeukaryotes in terrestrial habitats, the VSM *Paleoleptochlamys hassii* found in the 407 million-year-old Rhynie chert, a well-known representative deposit of an early terrestrial ecosystem environment (Strullu-Derrien et al., 2019). Consequently, elucidating the diversity and evolutionary history of Arcellinida is pivotal to illuminate the early evolution and diversification of microbial Eukaryotes both in marine and terrestrial ecosystems (Porter et al., 2003; Morais et al., 2019; Lahr, 2021; Strullu-Derrien et al., 2019).

Recent efforts of sampling diverse amoebozoan testate amoebae in a phylogenomic framework have resolved their deep phylogenetic relationships (Kang et al., 2017; Lahr et al., 2019). Amoebozoa is home to at least two testate amoebae groups: Arcellinida and Corycidia. Arcellinida is a diverse order represented by lineages that build hard shells, which corroborates their capability to leave the exceptionally preserved VSMs in the Neoproterozoic (Porter et al., 2003; Lahr et al., 2019). Corycidia is a recently established subclade of Amoebozoa represented by the lineages of testate amoebae that produce flexible shells and do not branch within Arcellinida (Kang et al., 2017). Despite this recent advancement, most genera of Arcellinida and Corycidia have not been sampled in a phylogenomic framework and their phylogenetic placement and classification remain uncertain (Lahr et al., 2019). Among the outstanding questions open for investigation, we highlight the genus *Microcorycia*. This genus of flexible shell amoebae is currently classified within Corycidia, however *Microcorycia* shares remarkable morphological similarities with some arcellinids by having a hemispherical shell, with an aperture as wide as the widest breadth of the shell, like *Microchlamys, Spumochlamys,* and *Pyxidicula* (Arcellinida) (Badewitz, 2004, Kosakyan et al., 2016). A more complete sampling of

amoebozoa testate amoebae, placing enigmatic taxa such as *Microcorycia*, is required to further appreciate their diversity and evolutionary history.

In addition to expanding the diversity sampling, the potential of combining the VSM record and phylogenomic reconstruction to infer the timing of Arcellinida diversification has not been explored (Riedman and Porter, 2019; Lahr, 2021). Importantly, not only are VSMs interpreted as the oldest direct evidence of heterotrophic organisms, they are also remarkably important for many additional reasons. VSMs are among the few and oldest microfossils that are assigned to an extant group of Eukaryotes, enabling the inference of their general biological characteristics (Porter et al., 2003; Lahr, 2021). Moreover, the highly preserved morphology of these fossils has allowed comparison of VSMs and the results of studies focused on morphological ancestral reconstructions of Arcellinida. Utilizing congruences between the morphologies of VSMs and ancestral reconstruction of arcellinids, specific groups of VSMs have been identified as potential fossils for calibrating the phylogenetic tree of Arcellinida (Lahr et al., 2019; Riedman and Porter, 2019; Lahr, 2021). Chronologically, VSMs predate the geological period characterized by extreme glaciation events that eroded sedimentary rocks worldwide and likely erased diverse post-Tonian fossil records (Keller et al., 2019), with the exception of the Jacadigo group VSMs reported for a glacial setting (Morais et al., 2021). Thus, vase-shaped microfossils serve as a unique testimony during a key geological time period for the evolution of life and the early diversification of Eukaryotes, predating a period where the fossil record is relatively more scarce (Keller et al., 2019; Cohen and Kodner, 2022). In this context, combining the VSMs record and Arcellinida phylogenetic tree calibration opens up avenues to illuminate several aspects of eukaryotic evolution throughout geological time.

Here we investigate the early divergence and diversification of Arcellinida and closely related amoebozoan taxa using a relaxed molecular clock approach. We analyze the diverse record of VSMs and Metazoa fossils to calibrate a well-resolved deep phylogenomic tree. We greatly expanded taxonomic sampling for amoebozoan testate amoebae in phylogenomics, including taxa that lacked precise placement thus far. We constructed a new phylogenomic dataset (utilizing 226 genes) focusing on amoebozoan testate amoebae (Arcellinida and Corycida). While our phylogenomic reconstruction provides novel insights into the organization of amoebozoan testate amoebae diversity, the time calibration sheds light on the

diversification of this diverse group of heterotrophic microeukaryotes. Different calibration strategies and molecular clock models support the divergence of Arcellinida during the early Neoproterozoic, between 886 and 764 mya. The divergence timing of Arcellinida and the diverse VSMs record they left corroborates that the Neoproterozoic witnessed the diversification of Eukaryotes in marine environments, including a recorded establishment of heterotrophic microbial eukaryotes. In summary, our results corroborate a Neoproterozoic diversification of heterotrophic microbial Eukaryotes, with a well-established complexity in shallow marine ecosystems, followed by a potential early land conquest of diverse taxa. Our time trees and Arcellinida fossil record suggest that the Devonian mass extinction on marine ecosystems may have also affected microbial eukaryotes, while terrestrial diversity exploded synchronously to the massive diversification of plants on land and the establishment of complex terrestrial ecosystems.

Results

A resolved tree of amoebozoan testate amoebae

For our phylogenomic reconstruction, we constructed a supermatrix composed of 57 taxa and 226 genes (70,428 amino acid sites) using the PhyloFisher v. 1.2.11 package (Tice et al., 2021). For this supermatrix, we sequenced an additional 13 testate amoebae taxa through single-cell or whole-culture transcriptomics (**Fig. 1; Dataset S01, Tables S1 - S4**). The remaining testate amoebae and sister-group taxa were sampled based on previously available genomes and transcriptomes (**Dataset S01, Table S1**). The resulting phylogenomic tree recovers a monophyletic Arcellinida, with three well-defined suborders (Phryganellina, Organoconcha, and Glutinoconcha) and five infraorders within Glutinoconcha (**Fig. 2; Appendix S01, Figs. S1 and S2**). The Corycidia is also recovered with full support, with two families, Trichosidae and Amphizonelliidae fam. nov. (**Fig. 2**). Nearly all nodes of the tree are fully (=100%) or highly (>92%) supported by Maximum Likelihood nonparametric Real Bootstraps (MLRB), with the exception of a single node within Sphaerothecina clade that has lower support (MLRB = 76%). Within the infraorder Sphaerothecina, the recently raised genus *Galeripora* (previously *Arcella* genus) branches sister to *Arcella*. The newly sequenced *Cyclopyxis* sp. branches within Sphaerothecina but does not form a monophyletic group with previously sequenced

211 *Cyclopyxis lobostoma*. The newly sequenced *Difflugia* cf. *capreolata* (Difflugiidae)
212 also branches as a member of Sphaerothecina, distantly related to the home of other
213 Difflugiidae (Longithecina infraorder). *Heleopera lucida* comb. nov. (previously
214 *Difflugia lucida*; Difflugiidae:Longithecina) branches as a sister group to *Heleopera*
215 *sphagni* and *Heleopera sylvatica* within Volnustoma. The flexible shell testate
216 amoebae *Microcorycia aculeata* and *Microcorycia flava* branch within Arcellinida
217 (Organoconcha suborder) together with *Spumochlamys*, *Microchlamys*, and
218 *Pyxidicula.* Within Phryganellina, *Phryganella paradoxa* and *Cryptodifflugia*
219 *operculata* form a robust clade, sister to *Phryganella acropodia*. Within Corycidia, the
220 newly sequenced *Trichosphaerium* sp. KSS branches as a sister group to a
221 previously sequenced *Trichosphaerium* sp., forming a fully supported clade with the
222 other flexible shell corycids (*Diplochlamys* and *Amphizonella*).
223 For our single-gene reconstructions, we constructed SSU rDNA and
224 Cytochrome C oxidase subunit I (COI) datasets (**Dataset S01, Tables S5 - S6**). The
225 resulting phylogeny based on the SSU rDNA gene recovered Arcellinida monophyly
226 and Organoconcha with low ultrafast bootstrap (UFBoot) support values (<95%), and
227 it failed to recover the Glutinoconcha suborder. Specifically, the infraorder
228 Volnustoma does not branch sister to the remaining Glutinoconcha (**Fig. S3**). On the
229 other hand, SSU recovers the Glutinoconha infraorders with full (= 100%) or high
230 (>98%) support (**Fig. S3**), except the Excentrostoma and Longithecina infraorder
231 (UFBoot = 67%). Some Arcellinida genera currently sampled only based on SSU,
232 such as *Physochila* and *Argynnia,* remain of unclear affinity (**Fig. S3**). The resulting
233 phylogeny of COI recovered a monophyletic Arcellinida clade (**Fig. S4**). Among the
234 suborders, Organoconcha and Phryganellina are recovered, while Phryganellina
235 branches within Glutinoconcha (**Fig. S4**). All Glutinoconcha infraorders are
236 recovered, with the exception of Sphaerothecina (**Fig. S4**). Both single-gene markers
237 do not recover the relationships between and within the suborders and infraorders
238 that are reconstructed through phylogenomics (**Fig. 2 and Figs. S3-S4**).
239
240 ***Time-calibrated tree of amoebozoan testate amoebae***
241 For estimating divergence times in testate amoebae evolution and closely related
242 taxa we expanded our phylogenomic supermatrix to consider a representative
243 sampling of Amorphea, the eukaryotic group composed of Amoebozoa and Obazoa
244 (i.e., composed of Opisthokonta (Fungi, animals, and their protistan relatives) and

two other protistan groups); **Dataset S01, Table S7**). For a rigorous calibration of the tree, we considered the well-documented and accepted fossils *Kimberella*, *Haikouichthys ercaicunensis*, and *Rusophycus* trace to calibrate five nodes within Animals (**Dataset S01, Table S8;** Benton et al., 2015). Similarly, we followed a rigorous and conservative approach in choosing the fossils to calibrate nodes within Arcellinda. We calibrated two of the deeper Arcellinida nodes, those being Organoconcha+Glutinoconcha based on *Melanocyrillium* VSM and Glutinoconcha based on *Limeta* and *Cycliocyrillum* VSMs (**Dataset S01, Tables S8 - S9**). These calibrations follow the most conservative and accepted morphological interpretations of VSMs based on ancestral reconstructions (Lahr et al., 2019; Riedman and Porter, 2019; Lahr, 2021).

Given different priors, clock models, and fossil calibration strategies influence the posterior dates estimated, we performed a total of 16 experiments with four different fossil calibration strategies (i.e., applying Uniform, Skew-Normal, Truncated-Cauchy short-tail, or Truncated-Cauchy long-tail distributions; **Dataset S01, Tables S10**) under an uncorrelated or autocorrelated relaxed clock model with either a drift parameter of α = 2 and β = 2 or α = 1 and β = 10. We ran each experiment in two independent MCMC chains to check for convergence, which was achieved for all analyses (**Appendix S01, Fig. S5**). Comparing the experiments, we observed slight differences between the clock models (*i.e.,* uncorrelated and autocorrelated) and overall younger inferred times with uniform calibration (uncorrelated median = 543 Ma and autocorrelated median = 530 Ma; **Dataset S01, Table S11 and Appendix S01, Fig. S6 - S14**) than with Skew-Normal (uncorrelated median = 578 - 582 Ma and autocorrelated median = 563 Ma; **Dataset S01, Table S11**). The drift parameter (i.e., α = 2 and β = 2 vs. α = 1 and β = 10) had virtually no impact independent of the calibration strategy and clock model (**Dataset S01, Table S11 and Appendix S01, Fig. S15**). Thus, the analyses applying Uniform and Skew-Normal strategies were highly congruent. While Uniform distributions represent a conservative interpretation of the fossil record, Skew-Normal represents a more literal interpretation. Conversely, t-Cauchy distributions represent a more loose interpretation and with our dataset led to the inference of extremely old and unrealistic ages (e.g., Amorphea root as old as 3000 mya; **Dataset S01, Table S11**). To check the influence of Neoproterozoic VSMs calibration on the estimated dates, we performed a total of four experiments calibrating only the nodes within the Animal

clade, applying Uniform and Skew-Normal calibration strategies, under an uncorrelated or autocorrelated relaxed clock model with a drift parameter of α = 2 and β = 2 (**Dataset S01, Table S11**). Comparing these experiments with those that considered VSMs to calibrate nodes within Arcellinida, we observe a congruency between their mean estimated times for all nodes of the tree, while non-calibrated nodes show longer and older confidence intervals (**Dataset S01, Table S11**). Aiming for a conservative approach, the time estimation results shown and discussed herein focus on the time posteriors estimated based on both Uniform and Skew-normal distribution calibration, considering the VSMs, and a drift parameter of α = 2 and β = 2 under the uncorrelated or autocorrelated relaxed clock model.

The molecular clock analyses inferred a mean time for the root of Amorphea to be between 1609 and 1510 mya (95% HPD confidence interval = 1755 - 1342 mya; **Fig. 3**). For Metazoa, the mean time estimated ranged between 892 and 754 mya (95% HPD CI = 968 - 710 mya; **Fig. 3**). The mean estimated for Amoebozoa origin ranged between 1553 - 1406 mya (95% HPD CI = 1707 - 1268 mya; **Fig. 3**). The Arcellinida node is constrained within the mean 835 and 805 mya (95% HPD CI = 886 - 764 mya; **Fig. 3**), estimating an origin for arcellinids during the early Neoproterozoic era (Tonian period; **Fig. 4 A**). For the early divergence time of Arcellinda, the estimated times suggest that the split between Organoconcha and Glutinoconcha branches occurred during the Tonian period (mean = 761 - 739 mya; 95% HPD CI = 804 - 725 mya; **Figs. 3 - 4 B**). Regarding the suborders of Arcellinida, we inferred a mean origin for Phryganellida between 461 - 280 (95% HPD CI = 569 - 187 mya; **Figs. 3 - 4 C**), for Organoconcha 647 - 592 (95% HPD CI = 698 - 509 mya; **Figs. 3 - 4 D**), and for Glutinoconcha between 699 - 684 mya (95% HPD CI = 743 - 656 mya; **Figs. 3 - 4 E**). For the deeper nodes of Glutinoconcha, the most sampled suborder of Arcellinida, the estimated divergence times ranged between Ediacaran and Devonian (mean = 550 - 426 mya; 95% HPD CI = 612 - 368; **Fig. 4 F - H**). The inferred ages for all infraorders of Glutinoconcha suggest a relatively recent diversification ranging from Devonian to early Cretaceous (mean = 352 - 181 mya; 95% HPD CI = 412 - 129 mya; **Figs. 3 - 4 I - M**). Most inferred ages for the nodes representing the origin of the modern extant genera and species of Arcellinida fall within carboniferous and cretaceous (95% HPD CI = 351 - 83 mya; **Dataset S01, Table S11**). The results and time-calibrated trees for all experiments are present in

312 the supplemental material (**Dataset S01, Table S11 and Appendix S01, Fig. S6 -**
313 **S14**).
314
315 **Timing of Arcellinida conquest of land**
316 We performed a statistical inference on the ancestral habitat of key hypothetical
317 ancestors within Arcellinida (**Dataset S01, Table S12 and Appendix S01, Fig. S16**).
318 The ancestral reconstruction inferred with high probability (>88%) a land habitat for
319 Organoconcha and Glutinoconcha hypothetical ancestors (**Fig. 5**), suggesting three
320 independent transition events (3TE) from marine to land environment by arcellinids.
321 Further comparisons of the reconstructed 3TE scenario with alternative scenarios
322 suggest that a marine or land habitat for Organoconcha and Glutinoconcha
323 hypothetical ancestors are statistically equally plausible, while only a land habitat for
324 Phryganellina ancestor is statistically supported (**Fig. 5, Dataset S01, Table S12,**
325 **and Appendix S01, Fig. S16**). Together, the results suggest at least three
326 independent transition events (3TE) from marine to land environment by arcellinids.
327 The 3TE scenario indicates that Arcellinida and Glutinoconcha+Organoconcha
328 hypothetical ancestors were marine, and Glutinoconcha, Organoconcha, and
329 Phryganellina ancestors were already established on land (**Fig. 5**). Together, 3TE
330 ancestral inference and the inferred times for Arcellinida diversification suggest that
331 Glutinoconcha and Organoconcha lineages may have been already established in
332 land habitats during Neoproterozoic/early-Paleozoic, between late Tonian and
333 early-Cambrian, and Phryganellina lineage between late-Neoproterozoic and
334 early-Mesozoic (**Fig. 5**). Alternatively, a marine habitat for Organoconcha or
335 Glutinoconcha hypothetical ancestors implies between four transition events (4TE) to
336 five transition events (5TE). Together, these ancestral inferences and the inferred
337 times for Arcellinida diversification suggest that Sphaerothecina and
338 Hyaliospheniformes common ancestor was already established in land habitats
339 between the early-Ediacaran (Neoproterozoic) and Ordovician (Paleozoic) periods
340 and that Organoconcha lineages may have conquested land between Ediacaran and
341 early-Carboniferous (**Fig. 5**).
342
343 **Discussion**
344
345 ***A resolved tree of amoebozoan testate amoebae***

The phylogenomic dataset constructed in this study improves several aspects of the previously available amoebozoan testate amoebae dataset (Kang et al., 2017; Lahr et al., 2019). Through the PhyloFisher pipeline, we were able to construct a curated phylogenomic matrix that is free of paralogues and contamination (**Appendix S01, SI1**). Moreover, the matrix constructed is an accessible and easy-to-update dataset, since newly sequenced transcriptomes can be easily added through PhyloFisher to further expand the taxonomic sampling of amoebozoan testate amoebae in a phylogenomic approach. In general terms, the phylogenomic tree obtained here is consistent with the previously published phylogenomic tree for amoebozoan testate amoebae (**Fig. 2**; Kang et al., 2017; Lahr et al., 2019). By recovering with full support all major groups of the Arcellinida and Corycidia clades, in accordance with previous reconstructions, we corroborate the robustness of the backbone of the amoebozoan testate amoebae tree. Importantly, the phylogenomic analyses enable us to place previously unsampled taxa within Arcellinida suborders, supporting the taxonomic actions regarding *Heleopera lucida* comb. nov. and the genus *Microcorycia* (**Appendix S01, SI2 - SI3**). Moreover, the phylogenomic tree identifies taxa that will need future taxonomic revision based on further morphological and molecular studies, those being *Difflugia* cf. *capreolata*, the genera *Cyclopyxis*, and *Phryganella*. Detailed discussion on the placement of newly sequenced amoebozoan testate amoebae is presented in **Appendix S01, SI2 - SI3.**

Tonian Vase-shaped microfossils (VSMs) provide robust calibrations for dating the emergence of Arcellinida

Our time-calibrated trees are congruent with several molecular clocks that have sampled the diversity of Eukaryotes (**Fig. 3, Dataset S01, Table S11, and Appendix S01, Fig. S6 - S14**). The estimated ages of the root of our tree fall within the range inferred by a recent investigation into the evolution of eukaryotes (95% confidence interval = 2177 - 1356 million years ago for Amorphea; Strassert et al., 2021). Similarly, the estimated ages for the origin of other major clades align with previous studies, including Obazoa (95% confidence interval = 2305 - 1526 million years ago), opisthokonts (95% confidence interval = 2019 - 1051 million years ago), and animals (95% confidence interval = 833 - 680 million years ago) (Parfrey et al., 2011; Eme et al. 2014; dos Reis et al., 2015; Strassert et al., 2021). Interestingly, these findings also corroborate the inference of the appearance of crown group eukaryotes by

independent analyses of steranes and protosteroids, which suggest a “Tonian revolution” (Brocks et al. 2023). To date, only two studies have calibrated the Arcellinida tree using single markers and the fossil interpretation available at the time (Fiz-Palacios et al., 2014; Lahr et al., 2015). Our inferred ages for Arcellinida are congruent with the study that considered mostly Opisthokonta nodes to calibrate the tree, which had at that time the most robust fossil record within Amorphea and inferred the origin of Arcellinida during the Tonian (mean estimated time = 944 mya) (Lahr et al., 2015). A younger alternative date has been suggested for the origin of Arcellinida, but this result was based on a somewhat controversial methodology for calibration that utilized fossils from late-Paleozoic and Mesozoic of specific Arcellinida genera to calibrate deep nodes of the tree. Naturally, this approach led to inferring a much younger origin of Arcellinida (mean estimated time = 602 mya; Farooqui et al., 2010; Kumar, 2011; Schmidt et al., 2004; Schmidt et al., 2006; Girard et al., 2011; Fiz-Palacios et al., 2014) than our results suggest. Additionally, the assigned ages of some of these fossils used in previous studies are imprecise or had poorly preserved morphology, leading to an ambiguous identification and therefore should not be used as reliable markers for calibration (Lahr et al., 2015). Moreover, the time-calibrated trees we generated without using the VSMs as calibration data inferred times congruent to our main analyses and the ages estimated by Lahr et al. (2015) (**Dataset S01, Table S11 and Appendix S01, SI4**). When these previous studies were conducted, the suborders and infraorders of Arcellinida had not yet been resolved, limiting further comparisons. Notably, the ages we inferred for the origin of modern groups (i.e., genera) are highly congruent with the recent Arcellinida fossil record, which postdate the Carboniferous period and preserve well-accepted fossils assigned to genera like *Arcella*, *Difflugia*, and *Centropyxis* (Schmidt et al., 2004; Farooqui et al., 2010; Girard et al., 2011). Collectively, these earlier investigations and the congruence to the fossil record support the robustness of combining well-established Metazoa and Arcellinida fossil records to calibrate a comprehensive phylogenomic sampling of Amorphea.

VSMs' Affinity to Arcellinida and the Diversity of Heterotrophic Microbial Eukaryotes in Neoproterozoic Oceans

Our time-calibrated tree analysis reveals that Neoproterozoic vase-shaped microfossils most likely represent the origin and early diversification of Arcellinida.

The ages of first appearance assigned to most of the 14 described genera of Neoproterozoic VSMs (811 - 729 million years ago), with the exception of VSM *Taruma rata* which dates from 659 - 638 mya, largely overlap with or postdate the inferred time range for the origin of Arcellinida (886 - 764 million years ago) and the split between Organoconcha and Glutinoconcha suborders (804 - 725 million years ago; **Figure 4 A - B**) (Porter et al., 2003; Riedman et al., 2018; Morais et al., 2019). The close agreement between these dates suggests a strong relationship between these VSMs and the two deepest nodes of Arcellinida. Consequently, we suggest the organisms represented by these microfossils are early branching arcellinids, diverging after the origin of Arcellinida but before the divergence of the Phryganellina, Organoconcha, and Glutinoconcha suborders. It is worth mentioning that some inferred times for Glutinoconcha origin fall within late-Tonian; however, most inferred ages for this suborder fall within the Cryogenian. Thus, when taken together, the time-calibrated trees and the fossil record suggest that even though only two hypothetical ancestors of arcellinids diverged during the Tonian period, the early divergence resulted in remarkable radiation. This is evidenced by the morphologically diverse VSMs, likely left by these ancestors, in the Neoproterozoic rocks worldwide.

This inferred radiation of early arcellinids provides compelling evidence for the capability of Neoproterozoic shallow marine ecosystems to support heterotrophic microbial eukaryotes. Biomarkers, fossil records, and molecular data suggest that at approximately 800 mya the Tonian period witnessed a significant transition from prokaryotic-dominated to eukaryotic-dominated ecosystems (Riedman and Sadler, 2018; Porter, 2020; Cohen and Kodner, 2022; Mills et al., 2022; Kang et al., 2023; Brocks et al. 2023). This transition was likely facilitated by factors such as increased availability of nitrate, phosphorous, silica, and reduced toxicity, which provided a favorable foundation for the diversification of phototrophic eukaryotes (Siever, 1992; Reinhard et al., 2017; Reinhard et al., 2020; Kang et al., 2023). Similarly, molecular clock analyses have indicated the divergence of multiple heterotrophic eukaryotes around the same time (Parfrey et al., 2011; Eme et al., 2014). However, due to the lack of a fossil record for these organisms, their diversity has been challenging to examine. Nevertheless, the timing of Arcellinida early divergence and the diversity of vase-shaped microfossils in shallow marine environments corroborate a remarkable diversification of heterotrophic eukaryotes during the Tonian period.

The probable establishment of communities of phototrophic and heterotrophic eukaryotes in the marine environment during the early Neoproterozoic brings valuable insights into the complex interactions between different microbial eukaryotic groups and the prevailing environmental conditions of that time. The combined need for oxygen and prey are key characteristics of several heterotrophs. During the Tonian period, the ocean exhibited high stratification in terms of oxygen availability, with predominantly anoxic conditions prevailing in the mid and deep oceans (Guilbaud et al., 2015; Wang et al., 2021). Nonetheless, geochemical evidence suggests higher concentrations of oxygen in certain shallow marine waters referred to as oxygen oases (Wang et al., 2021). Probably triggered by photosynthetic biological mats, the concentration of oxygen likely led to improved aerobic conditions and the establishment of benthic oxygen oases, which were conducive to the survival of aerobic eukaryotes (Xiao and Tang, 2018; Del Cortona et al., 2020; Wang et al., 2021). Supporting this scenario, the presence of vase-shaped microfossils (VSMs) in rocks characterized by organic-rich sediments suggests their close association with microbial mats (Porter et al., 2003). Adding to this complexity, the predatory requirements of heterotrophic eukaryotes played a crucial role in the establishment of marine ecosystems (Porter, 2016; Cohen and Riedman, 2018). The organisms represented by these VSMs likely preyed on both the bacterial (bacteriovory) and eukaryotic communities (eukaryovory), similar to extant Arcellinida (Porter et al., 2003; Porter, 2011; Knoll, 2014). Culture observations have demonstrated diverse strategies of extant arcellinids to prey on both phototrophic and heterotrophic organisms, including diatoms and nematodes (Geisen et al., 2015; Dumack et al., 2018). Moreover, several VSMs exhibit holes on their shells interpreted as predation marks, providing further evidence they served as prey to other heterotrophs (Porter, 2016; Cohen and Riedman, 2018). Importantly, predation has been interpreted as one of the triggers for eukaryotic diversification, not only for unicellular lineages but also for macroscopic ones, including animals (Cousyn et al., 2001; Rundle et al., 2003; Meyer and Kassen, 2007; Cohen and Riedman, 2018). Consequently, Arcellinida diversity represented by VSMs stands out as a robust direct record of the establishment of complex shallow marine ecosystems and evolutionary contexts during the Neoproterozoic.

Pre-Cryogenian Eukaryotic Explosion and survival during Earth's Most Severe Ice Ages

The radiation of Arcellinida and the diversity of eukaryotic life just before the Cryogenian period likely played a crucial role in their survival during Earth's most severe ice ages. The Stuartian and Marinoan glaciation periods witnessed extensive ice coverage across the planet, with glaciers even extending into tropical regions (Hoffman and Li, 2009; Rooney et al., 2015; Pu et al., 2016). It is suggested that the survival of life during these Cryogenian glaciations can be explained by the presence of refugia and the adoption of dormancy strategies (Runnegar, 2000; Vincent et al., 2000; Hoffman and Schrag, 2002; Campbell et al., 2011; Campbell et al., 2014; Lawal et al., 2020; Morais et al., 2021). Refugia refers to specific regions or environments that may have offered relatively warmer and less severe conditions serving as shelter for life to persist during glaciation (Campbell et al., 2014). Although habitable environments during the glaciations may have been limited, life could have also adapted by entering a dormant stage, forming cysts that allowed them to withstand adverse conditions (Morgan-Kiss et al., 2006). The presence of cyst-like structures identified inside vase-shaped microfossils serves as direct evidence that these organisms were already capable of entering dormancy stages, similar to extant arcellinids (Mus and Moczydowska, 2000). Additionally, recent discoveries have revealed that several diverse Tonian VSM taxa persisted into the Cryogenian period, further confirming the presence of Arcellinida in marine environments and providing fossil records that showcase the diversity of heterotrophic eukaryotes during glaciation periods (Morais et al., 2021). Importantly, the inferred times for Glutinoconcha and Organoconcha suggest they possibly originated during Cryogenian, which suggests not only survival but also diversification of life during the earth's most severe glaciations (**Figure 4 D - E**). The age of the VSM *Taruma rata*, which is thus far exclusively found in Cryogenian rocks (659 - 638 mya; Morais et al., 2021), overlaps with or postdates our estimates of Glutinoconcha and Organoconcha origins. Thus, *T. rata* may represent a lineage closely related to Glutinoconcha or Organoconcha that diverged during the Cryogenian. Consequently, the VSMs record and the timing of early Arcellinida evolution corroborates the diversification of heterotrophic eukaryotic life shortly before and during the Cryogenian period. This coupled with the exploitation of

refugia habitats and dormancy strategy, may explain the endurance of life during Earth's most severe ice age.

Timing of Arcellinida conquest of land and extinction from marine habitats

The current interpretation of the Neoproterozoic VSMs and the timing of arcellinids' divergence support the origin and early diversification of Arcellinida in shallow marine environments and subsequent multiple transitions to land (Porter et al. 2003; Morais et al., 2021; Lahr, 2021). The most conservative and robust interpretation of these fossils, corroborated by ancestral state reconstruction, poses that the marine VSM *Melanocyrillum* represents a fossil group closely related to Organoconcha+Glutinoconcha hypothetical common ancestor and the VSMs *Limeta* and *Cycliocyrillium* closely related to Glutinoconcha ancestor (Lahr et al., 2019; Riedman and Porter, 2019). This interpretation implies that the split between arcellinids' suborders (Phryganellina, Organoconcha, and Glutinoconcha) happened prior to land conquest. In this scenario, the diversity of Phryganellina, Organoconcha, and Glutinoconcha on land observed today is linked to at least three independent transition events (**Fig. 5**). Multiple transitions into land by arcellinids have already been hypothesized based on the fact that all extant arcellinids inhabit continental environments while their ancient fossils are found exclusively in marine habitats (Porter and Knoll, 2000; Porter et al., 2003; Lahr et al., 2019; Morais et al., 2019; Lahr, 2021, Wang et al. 2023). Importantly, the inferred times of the split between the suborders predate any evidence of major conquest of land by other eukaryotic groups (Wellman and Strother, 2015; Žárský et al., 2022). Although indirect chemical and geological data demonstrate the diversification of prokaryotic life on land as early as 3 bya and body fossils support the presence of eukaryotic life around 1 bya, well-accepted evidence of modern crown eukaryotes diversification on land appears only during Cryogenian (Wellman and Strother, 2015; Žárský et al., 2022). Thus, unless arcellinids dominated land habitats prior to other eukaryotic crown groups, which is unlikely, the fossil record and the inferred times further corroborate the origin and early diversification of Arcellinida lineages in marine environments.

Currently, both an early (Neoproterozoic/early-Paleozoic) or late (late-Paleozoic) land conquest by Arcellinida are equally plausible based on the available fossil record. The VSMs record starts during the Tonian and extends into

the Cryogenian, the latter being the youngest period of the Neoproterozoic in which we have fossil evidence for Arcellinida (Morais et al., 2021; Sharma et al., 2021). Remarkably, recent discoveries show VSMs attributed to arcellinids appearing again in the fossil record during the Devonian, both in marine and terrestrial environments (Strullu-Derrien et al., 2019; Wang et al., 2023). These novel fossils are direct evidence that members of Arcellinida inhabited shallow oceans at least until the Devonian and that during this same period, they were already present on land. It is worth noting that these VSMs do not represent evidence that Arcellinida did not conquer land prior to Devonian, the lack of VSMs record postdating Cryogenian and predating Devonian impairs inferring the time of transition from marine to land habitats by arcellinids.

Our reconstructions of ancestral states indicate that an early conquest of land (Neoproterozoic/early-Paleozoic) represents a plausible scenario (**Fig. 5**), and is congruent with new findings regarding land colonization by eukaryotic life. Together, the inferred times for Arcellinida diversification and the statistically supported scenarios (i.e., 3TE, 4TE, and 5TE) of their land conquest imply that Glutinoconcha lineages dominated land between the late Tonian and Ordovician and Organoconcha lineages colonized land between the Cryogenian and early-Devonian (**Figs. 4 - 5**). Glutinoconcha internal nodes diverged over the Ediacaran and Devonian, with a radiation of its infraorders starting during the Devonian in land habitats (**Figs. 4 - 5**). Novel insights on earlier continental ecosystem formation, challenging the currently prevailing notion, suggest the Cryogenian witnessed the establishment of complex environments on land (Lutzoni et al., 2018; Žárský et al., 2022). Fossil and molecular clock evidence have demonstrated that both eukaryotic phototrophs and absorptive heterotrophs (Fungi) had already dominated terrestrial habitats, modifying the oxygen and nutrient availability of their surroundings (Lutzoni et al., 2018; Žárský et al., 2022). Here we propose that the establishment of terrestrial habitats during Cryogenian and their biotic and abiotic condition may have been permissive for an early Arcellinida terrestrialization. If so, the Organoconcha and Glutinoconcha divergence times may mark the establishment of these two already on Earth's emerging terrestrial ecosystems (**Figs. 4 - 5**). The timing of the origin of Phryganellina is less consistent among experiments impairing more precise interpretations, but all the inferred times fall within the period eukaryotic life was already inhabiting land (**Fig. 5**). In this context, an early land conquest by Arcellinida

implies that during the early establishment of the first land habitats, a complex ecosystem composed of eukaryotic phototrophs and heterotrophs rapidly formed. Importantly, all inferred dates and ancestral reconstructions suggest that from the mid-Paleozoic onward Arcellinida diversified while inhabiting land, and the mid-Paleozoic to early-Mesozoic was marked by the divergence of the five Glutinoconcha infraorders, the same time period estimated for the radiation of land plants (**Figs. 4 - 5**; Donoghue et al., 2021). It is worth mentioning that an early land conquest does not imply arcellinids also disappeared early from marine environments, as evidenced by the presence of VSM in marine environments at least until the Devonian (Wang et al., 2023). Although marine VSMs postdating the Devonian may still be found, the well-known late Devonian mass extinction, linked to the disappearance of diverse macroscopic organisms (Racki, 2005; White et al., 2018), allows us to consider that this extinction event also affected microbial communities, and may be the cause of Arcellinida extinction and current absence from marine environments.

In conclusion, we expanded the phylogenomic sampling of amoebozoan testate amoebae and inferred robust time-calibrated Amorphea trees based on VSMs and metazoan fossil records. The inferred times of Arcellinida's early diversification, the Tonian VSMs, coupled with comprehensive geological and paleontological evidence, corroborate marine arcellinids originated and radiated during Tonian. This suggests that heterotrophic eukaryotes were able to thrive and diversify in marine ecosystems already during their early establishment. The diversification of Arcellinida, and most probably of other eukaryotes, prior to the Cryogenian likely contributed to their survival during Earth’s two most extreme glaciations. While still tentative, an early terrestrialization and Devonian marine extinction of Arcellinida is congruent with the timing of earth’s terrestrial habitat formation, a Devonian marine mass extinction event, and our time-calibrated trees, serving thus as a robust null hypothesis. The potential of novel VSMs discoveries and the expansion of phylogenomic sampling will allow further illumination of Arcellinida evolution. Ultimately, we demonstrate the robustness of Arcellinida evolution, vase-shaped microfossil record, and molecular investigation as key to shedding light on the diversification of heterotrophic microbial Eukaryotes and the establishment of both marine and terrestrial ecosystems from the Neoproterozoic.

Material and Methods

Sampling, RNA extraction, and sequencing

We generated transcriptomes for 13 previously genomically unsampled amoebozoan testate amoeba species through mRNA extraction from either monoclonal cultures or single-cells isolated from natural samples (**Dataset S01, Table S1**). We synthesized cDNA from RNA extractions using Smart-seq2 (Picelli et al., 2014). We prepared our cDNA libraries for sequencing on the Illumina platform using a NexteraXT DNA Library Preparation Kit (Illumina) following the manufacturer's recommended protocol. Libraries were then pooled and sequenced (**Appendix S01, SI1; Dataset S01, Table S1**).

Trimming, Transcriptome Assembly, and Quality Assessment

We trimmed primers, adaptors, and low-quality bases from raw Illumina reads using the Trimmomatic v. 0.36 (Bolger et al., 2014). We then assembled the surviving reads with Trinity v. 2.12.0 (Haas et al., 2013). We predicted amino acid sequences (proteomes) from the assembled transcriptomes using Transdecoder v. 5.5.0. Finally, we assessed the completeness of all newly sequenced transcriptomes using BUSCO v. 5.3.2 (Manni et al., 2021) (**Dataset S01, Table S1**). Further details on trimming, transcriptome assembly, and quality assessment are presented in **Appendix S01, SI1.**

Phylogenomic dataset construction

We constructed our amoebozoan phylogenomic dataset using the database and tools found in PhyloFisher v. 1.2.11 (Tice et al., 2021) following the steps outlined at https://thebrownlab.github.io/phylofisher-pages/detailed-example-workflow. Our final concatenated matrix used in subsequent phylogenetic analyses consisted of 226 genes (70,428 amino acid sites) and 57 amoebozoan taxa (**Dataset S01, Table S4**). From each individual ortholog that was concatenated to create the aforementioned matrix, we constructed single ortholog trees to be used as input for coalescent-based phylogenomic analyses. Further details on our approach for phylogenomic dataset construction are presented in **Appendix S01, SI1.**

Phylogenomic analyses

650 We performed maximum likelihood phylogenetic reconstruction using our final matrix
651 with IQ-TREE2 v. 2.0-rc (Minh et al. 2020). We initially inferred a tree from our matrix
652 under the LG+C20+G4 model. We used the resulting tree as a guide tree to infer a
653 Posterior Means Site Frequency model (Wang et al., 2018) using the ML model
654 LG+C60+G4+PMSF in IQ-TREE2. We assessed the topological support for the
655 resulting tree by 100 Real nonparametric Bootstrap replicates in IQ-TREE (IQ-TREE
656 v. 2.1.2 COVID-edition) using the PMSF model. The topological support values
657 inferred from MLRB were mapped onto the ML tree using RAxML v. 8.2.12
658 (Stamatakis, 2014) using the "-f b" option. We carried out coalescent-based
659 phylogenomic analyses with ASTRAL-III v. 5.7.3 (Zhang et al., 2018). Statistical
660 support for our ASTRAL-III analysis was assessed via local posterior probability
661 values.
662
663 ***SSU and COI phylogenetic analysis***
664 We retrieved Small Subunit ribosomal RNA (SSU) and Cytochrome C oxidase
665 subunit I (COI) gene sequences from transcriptomes through similarity searches
666 implemented in Blast+ v. 2.14.1 (Camacho et al., 2009). To perform similarity
667 searches we built SSU and COI datasets of Arcellinida based on sequences
668 available on NCBI and used them on independent blastn searches (Camacho et al.,
669 2009) to retrieve SSU and COI sequences from the available transcriptomes. For
670 SSU and COI phylogenetic reconstructions we performed multiple sequence
671 alignments with MAFFT v. 7.453 (Katoh et al., 2002). We performed automated
672 alignment trimming to remove from the final alignments highly variable and poorly
673 aligned regions using trimAl v. 1.2 (Capella-Gutiérrez et al., 2009). We obtained
674 trimmed SSU and COI alignments composed of 764 and 1358 nucleotides,
675 respectively, that we used for the phylogenetic analyses. We inferred all the
676 maximum-likelihood trees using ModelFinder (Kalyaanamoorthy et al., 2017) and
677 obtained node supports with the ultrafast bootstrap (Hoang et al., 2018), both
678 implemented in the IQ-TREE v. 2.2.2.7 software (Minh et al., 2020).
679
680 ***Bayesian molecular dating***
681 *Dataset and Topology*
682 Utilizing previously identified orthologs already present in the publicly available
683 PhyloFisher database, we expanded our amoebozoan phylogenomic dataset to

include a representative sampling of Amorphea. Amorphea is the eukaryotic clade composed of Amoebozoa, Fungi, Animals, and some other unicellular lineages. These ortholog sequences were aligned, trimmed, and concatenated as described above (*Phylogenomic dataset construction section*). Our final expanded dataset used in the subsequent phylogenetic reconstruction and for the molecular dating analysis consisted of 230 genes (73,467 amino acid sites) and 96 taxa (**Dataset S01, Table S7**). Maximum likelihood phylogenetic analysis was performed as described above (*Phylogenomic analyses section*).

Fossil constraints

As external calibration information, we considered widely accepted fossils to calibrate five internal nodes of the Metazoa clade and two internal nodes of Arcellinida (Benton et al., 2015; dos Reis et al., 2015; Lahr et al. 2019; Porter and Riedman 2019; **Dataset S01, Tables S8 - S9**). We derived the fossil calibration for Arcellinida nodes based on previous analyses and interpretations of the morphological relationship between vase-shaped microfossils and extant arcellinids (Lahr et al., 2019; Porter and Riedman 2019). Interpretations of Bayesian and maximum likelihood ancestral reconstructions of arcellinids shell morphology suggest a morphological congruence between the VSM *Melanocyrillium* to the Glutinoconcha+Organoconcha hypothetical ancestor and between the VSMs *Limeta* and *Cycliocyrillum* to Glutinoconcha hypothetical ancestor (Lahr et al. 2019). Currently, these congruences represent the most conservative interpretation of VSMs affinity to Arcellinida and have been evaluated by independent members of the paleontological community specialized in these microfossils (Porter and Riedman 2019). Importantly, the diversity and morphology of these Neoproterozoic VSMs found in sedimentary deposits around the world have been consistently investigated, and their stratigraphy continuously refined over time, consequently enabling us to constrain these fossils' ages (**Dataset S01, Tables S9;** Knoll and Vidal 1980; Binda and Bockhari, 1980; Knoll and Calder 1983; Knoll et al., 1989; Green et al., 1988; Knoll et al., 1991; Porter et al., 2003; Dehler et al., 2010; Sergeev and Schopf, 2010; Strauss et al., 2014; Riedman et al., 2018; Cohen et al., 2017; Moczydłowska et al., 2017; Freitas et al., 2021; Rooney et al., 2018; Morais et al., 2019; Mus et al., 2020). Thus, based on the ancestral morphological reconstructions and the discussion of vase-shaped microfossils morphology and stratigraphy present in the literature, we

consider *Melanocyrillium* to calibrate Glutinoconcha+Organoconcha node with a minimum time of 729 mya and maximum of 750 mya and *Limeta* and *Cycliocyrillum* to calibrate Glutinoconcha node with a minimum time of 659 mya and a maximum of 750 mya (Lahr et al., 2019; Porter and Riedman 2019; **Dataset S01, Table S8**). To determine these age limits in a conservative approach, we have listed all available reported occurrences of these VSMs across the planet and have restricted the dates to the maximum minimal date reported and minimum maximal date reported, thus encompassing the total distribution of times reported for these fossils (**Dataset S01, Table S9**).

Since molecular dating considers the fossil information as statistical distributions, and different distributions may impact the time estimation differently, we followed dos Reis et al. (2015) strategy and used a total of four different distributions to represent the calibrations derived from the fossil record: i. uniform; ii. skew-normal; iii. truncated-cauchy short-tail; and iv. truncated-cauchy long-tail. For the uniform distribution, we mapped onto the nodes to be calibrated the fossils' minimum (t_L) and maximum (t_u) ages and the left (p_L= *0.001*) and right (p_u= *0.025*) tail probabilities following the format 'B(t_L, t_u, p_L, p_u)' (**Dataset S01, Table S10**). For the Skew-Normal distribution, we mapped onto the nodes the fossils' minimum age (location) and the scale (α) and shape (β) of the distribution following the format 'SN(location, α, β)', the Skew-Normal distribution distributions were calculated and plotted with MCMCtreeR (**Dataset S01, Table S10**). For truncated-Cauchy we mapped onto the nodes the fossils' minimum age (t_L), with mode (p = 0), tail parameter (c = 0.1 for short-tail and c = 10 for long-tail), and left tail probability (p_L=0.01) following the format 'L(t_L, p, c, p_L)' (**Dataset S01, Table S10**).

Divergence time estimation

We performed Bayesian molecular dating with the MCMCTree program, implemented within the PAML package (**Dataset S01, Table S10**; Yang, 2007). We estimated a mean substitution rate of 0.03135 replacement site^{-1}10^{-8}Myr^{-1} for our dataset with IQ-TREE2 v. 2.0.6 (Minh et al., 2020) phylogenetic dating under the LG+G model. Within the PAML package, we set the overall substitution rate ('rgene_gamma = α, β' parameter) as a gamma-Dirichlet prior following dos Reis et al., 2015, with α = 2 and β = 63.78. This substitution rate was implemented for all dating experiments. For the rate drift parameter ('sigma^2_gamma = α, β'), we

752 independently implemented two alternatives, α = 2 and β = 2 or α = 1 and β = 10.
753 Similarly, we considered both uncorrelated and autocorrelated relaxed clock models.
754 For all experiments, we analyzed the data under the LG+G model as a single
755 partition, constrained the root age between 1.6 - 3.2 Ga, and considered a uniform
756 birth-death tree prior and 100 million years as one time unit. We performed a total of
757 16 experiments, four experiments for each calibration strategy (i.e., uniform,
758 skew-normal, truncated-cauchy short-tail, or truncated-cauchy long-tail), varying the
759 rate drift parameter (i.e., α = 2 and β = 2 or α = 1 and β = 10), clock model (i.e.,
760 uncorrelated and autocorrelated). For each experiment, we ran two independent
761 MCMC chains to verify convergence, discarding the first 2,000 iterations as burn-in
762 and considering the following 20 million generations. To check the influence of
763 calibration using Neoproterozoic VSMs on the estimated dates, we performed a total
764 of four experiments calibrating only the nodes within the Animal clade, applying
765 Uniform and Skew-Normal calibration strategies, under an uncorrelated or
766 autocorrelated relaxed clock model with a drift parameter of α = 2 and β = 2,
767 following the same approach described above, as detailed in **Appendix S01, SI4.**
768
769 ***Ancestral reconstruction of Arcellinida habitat***
770 We applied maximum likelihood (ML) method, implemented in BayesTraits v. 4.0.1
771 (Meade and Pagel, 2022), to statistically reconstruct the ancestral habitat states of
772 arcellinids and compare evolutionary scenarios. Currently, the diverse vase-shaped
773 microfossils record and ancestral morphological reconstructions suggest that
774 Arcellinida origin and the split between Organoconcha and Glutinoconcha took place
775 in marine environments during the Neoproterozoic (Lahr et al., 2019; Riedman and
776 Porter, 2019). However, a post-Tonian gap in the fossil record, probably influenced
777 by extreme glaciation events that eroded and modified sedimentary rocks (Keller et
778 al., 2019; Morais et al., 2021; Lahr, 2021), impairs inferring the timing and context of
779 land conquest events that happened over Arcellinida evolutionary history. Thus, we
780 combine the fossil evidence and the phylogenomic reconstruction with branch
781 lengths to reconstruct the ancestral habitat states (i.e., marine habitat *vs.* land
782 habitat) of key Arcellinida clades through the BayesTraits MultiState method.
783 Specifically, we used the 100 topologies with branch lengths obtained for the
784 phylogenomic Real Bootstrap topological support assessment (see *Phylogenomic*
785 *analyses section*) and implemented ancestral reconstruction analyses fossilizing

Arcellinida and Organoconcha+Glutinoconcha nodes as marine (**DatasetS01, Table S12 and Appendix S01, Fig. S16**). Aiming for a conservative approach, we further compared the reconstructed scenario with alternative hypothetical scenarios through BayesTraits fossilizing strategy (**DatasetS01, Table S12 and Appendix S01, Fig. S16**). Specifically, we modeled and compared hypothetical scenarios by fixing nodes to specific values (i.e., marine or land), calculating their maximum likelihood values and applying a likelihood ratio test (LRT), which we considered as significant a difference of LRT ≥ 2 (Pagel, 1999).

Data Availability

Raw sequencing files are deposited at NCBI SRA repository under the Bioproject PRJNA1032600. Phylogenomic supermatrix, single gene marker datasets, and input information for the molecular clock and ancestral reconstruction analyses are presented in DatasetS01.

Acknowledgments

We thank Enrique Lara for the discussions and critiques. We thank Cori Tice for her help with R and advice on plot aesthetics. We thank Alastair G.B. Simpson for providing the culture of *Trichosphaerium* sp. KSS. This work was supported in part by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grants 2019/22692-8 and 2021/09529-0 awarded to A.L.P.S. and grant 2019/22815-2 awarded to D.J.G.L., and by the United States National Science Foundation (NSF) Division of Environmental Biology (DEB) grant 2100888 (http://www.nsf.gov) awarded to M.W.B. KD was funded by the German Research Foundation grant 399699069 - DU 1863/1.

References

Badewitz, H.J., 2004. The genus Microcorycia Cockerell, 1911 (Testacealobosia, Rhizopoda, Protozoa). A critical monograph of the genus including a first description of a new species: Microcorycia scutella n. sp. Lauterbornia, 50, pp.111-146.

Brocks, J.J., Nettersheim, B.J., Adam, P., Schaeffer, P., Jarrett, A.J., Güneli, N., Liyanage, T., van Maldegem, L.M., Hallmann, C. and Hope, J.M., 2023. Lost world of complex life and the late rise of the eukaryotic crown. Nature, pp.1-7.

Benton, M.J., Donoghue, P.C., Asher, R.J., Friedman, M., Near, T.J. and Vinther, J., 2015. Constraints on the timescale of animal evolutionary history. Palaeontologia Electronica, 18(1), pp.1-106.

Binda, P.L. and Bokhari, M.M., 1980. Chitinozoanlike microfossils in a late Precambrian dolostone from Saudi Arabia. Geology, 8(2), pp.70-71.

Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), pp.2114-2120.

Brocks, J.J., Nettersheim, B.J., Adam, P., Schaeffer, P., Jarrett, A.J., Güneli, N., Liyanage, T., van Maldegem, L.M., Hallmann, C. and Hope, J.M., 2023. Lost world of complex life and the late rise of the eukaryotic crown. Nature, pp.1-7.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L., 2009. BLAST+: architecture and applications. BMC bioinformatics, 10, pp.1-9.

Campbell, A.J., Waddington, E.D. and Warren, S.G., 2011. Refugium for surface life on Snowball Earth in a nearly-enclosed sea? A first simple model for sea-glacier invasion. Geophysical research letters, 38(19).

Campbell, A.J., Waddington, E.D. and Warren, S.G., 2014. Refugium for surface life on Snowball Earth in a nearly enclosed sea? A numerical solution for sea-glacier invasion through a narrow strait. Journal of Geophysical Research: Oceans, 119(4), pp.2679-2690.

Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics, 25(15), pp.1972-1973.

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(5), pp.683-701.

Cohen, P.A. and Kodner, R.B., 2022. The earliest history of eukaryotic life: uncovering an evolutionary story through the integration of biological and geological data. Trends in Ecology & Evolution, 37(3), pp.246-256.

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(2), pp.173-180.

Del Cortona, A., Jackson, C.J., Bucchini, F., Van Bel, M., D'hondt, S., Škaloud, P., Delwiche, C.F., Knoll, A.H., Raven, J.A., Verbruggen, H. and Vandepoele, K., 2020. Neoproterozoic origin and multiple transitions to macroscopic growth in green seaweeds. Proceedings of the National Academy of Sciences, 117(5), pp.2551-2559.

Cousyn, C., De Meester, L., Colbourne, J.K., Brendonck, L., Verschuren, D. and Volckaert, F., 2001. Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. Proceedings of the National Academy of Sciences, 98(11), pp.6256-6260.

Dehler, C.M., Fanning, C.M., Link, P.K., Kingsbury, E.M. and Rybczynski, D., 2010. Maximum depositional age and provenance of the Uinta Mountain Group and Big Cottonwood Formation, northern Utah: Paleogeography of rifting western Laurentia. Bulletin, 122(9-10), pp.1686-1699.

Donoghue, P.C., Harrison, C.J., Paps, J. and Schneider, H., 2021. The evolutionary emergence of land plants. Current Biology, 31(19), pp.R1281-R1298.

Dumack, K., Kahlich, C., Lahr, D.J. and Bonkowski, M., 2018. Reinvestigation of Phryganella paradoxa (Arcellinida, Amoebozoa) Penard 1902. Journal of Eukaryotic Microbiology, 66(2), pp.232-243.

Eme, L. and Ettema, T.J., 2018. The eukaryotic ancestor shapes up.

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 Harbor Perspectives in Biology, 6(8), p.a016139.

Farooqui, A., Kumar, A., Jha, N., Pande, A.C. and Bhattacharya, D.D., 2010. A thecamoebian assemblage from the Manjir formation (early Permian) of northwest Himalaya, India. Earth Science India, 3.

Fiz-Palacios, O., Leander, B.S. and Heger, T.J., 2014. Old lineages in a new ecosystem: diversification of arcellinid amoebae (Amoebozoa) and peatland mosses. PLoS One, 9(4), p.e95238.

Freitas, B.T., Rudnitzki, I.D., Morais, L., Campos, M.D.R., Almeida, R.P.D., Warren, L.V., Boggiani, P.C., Caetano-Filho, S., Bedoya-Rueda, C., Babinski, M.B.T.T.R. and Fairchild, T.R., 2021. Cryogenian glaciostatic and eustatic fluctuations and massive Marinoan-related deposition of Fe and Mn in the Urucum District, Brazil. Geology, 49(12), pp.1478-1483.

Geisen, S., Mitchell, E.A., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L.D., Jousset, A., Krashevska, V., Singer, D. and Spiegel, F.W., 2018. Soil protists: a fertile frontier in soil biology research. FEMS Microbiology Reviews, 42(3), pp.293-323.

Geisen, S., Rosengarten, J., Koller, R., Mulder, C., Urich, T. and Bonkowski, M., 2015. Pack hunting by a common soil amoeba on nematodes. Environmental microbiology, 17(11), pp.4538-4546.

Girard, V., Néraudeau, D., Adl, S.M. and Breton, G., 2011. Protist-like inclusions in amber, as evidenced by Charentes amber. European Journal of protistology, 47(2), pp.59-66.

Green, J.W., Knoll, A.H. and Swett, K., 1988. Microfossils from oolites and pisolites of the upper Proterozoic Eleonore Bay Group, central East Greenland. Journal of Paleontology, 62(6), pp.835-852.

Guilbaud, R., Poulton, S.W., Butterfield, N.J., Zhu, M. and Shields-Zhou, G.A., 2015. A global transition to ferruginous conditions in the early Neoproterozoic oceans. Nature Geoscience, 8(6), pp.466-470.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B.O., Lieber, M. and MacManes, M.D., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols, 8(8), pp.1494-1512.

Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q. and Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Molecular biology and evolution, 35(2), pp.518-522.

Hoffman, P.F. and Li, Z.X., 2009. A palaeogeographic context for Neoproterozoic glaciation. Palaeogeography, Palaeoclimatology, Palaeoecology, 277(3-4), pp.158-172.

Hoffman, P.F. and Schrag, D.P., 2002. The snowball Earth hypothesis: testing the limits of global change. Terra nova, 14(3), pp.129-155.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A. and Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14(6), pp.587-589.

Kang, J., Gill, B., Reid, R., Zhang, F. and Xiao, S., 2023. Nitrate limitation in early Neoproterozoic oceans delayed the ecological rise of eukaryotes. Science Advances, 9(12), p.eade9647.

Kang, S., Tice, A.K., Spiegel, F.W., Silberman, J.D., Pánek, T., Čepička, I., Kostka, M., Kosakyan, A., Alcântara, D.M., Roger, A.J. and Shadwick, L.L., 2017. Between a pod and a hard test: the deep evolution of amoebae. Molecular biology and evolution, 34(9), pp.2258-2270.

Kumar, A., 2011. Acid-resistant Cretaceous thecamoebian tests from the Arabian Peninsula: a suggestion for study of agglutinated rhizopods in palynological slides. Journal of Micropalaeontology, 30(1), pp.1-5.

Katoh, K., Misawa, K., Kuma, K.I. and Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research, 30(14), pp.3059-3066.

Keller, C.B., Husson, J.M., Mitchell, R.N., Bottke, W.F., Gernon, T.M., Boehnke, P., Bell, E.A., Swanson-Hysell, N.L. and Peters, S.E., 2019. Neoproterozoic glacial origin of the Great Unconformity. Proceedings of the National Academy of Sciences, 116(4), pp.1136-1145.

Knoll, A.H., 2014. Paleobiological perspectives on early eukaryotic evolution. Cold Spring Harbor Perspectives in Biology, 6(1), p.a016121.

Knoll, A.H. and Calder, S., 1983. Microbiotas of the late Precambrian Ryssö Formation, Nordauslandet, Svalbard. Palaeontology 26:467–496.

Knoll, A.H., Swett, K. and Burkhardt, E., 1989. Paleoenvironmental distribution of microfossils and stromatolites in the Upper Proterozoic Backlundtoppen Formation, Spitsbergen. Journal of Paleontology, 63(2), pp.129-145.

Knoll, A.H., Swett, K. and Mark, J., 1991. Paleobiology of a Neoproterozoic tidal flat/lagoonal complex: the Draken Conglomerate Formation, Spitsbergen. Journal of paleontology, pp.531-570.

Knoll, A.H. and Vidal, G., 1980. Late Proterozoic vase-shaped microfossils from the Visingsö Beds, Sweden. Geologiska Föreningen i Stockholm Förhandlingar, 102(3), pp.207-211.

Kosakyan, A., Gomaa, F., Lara, E. and Lahr, D.J., 2016. Current and future perspectives on the systematics, taxonomy and nomenclature of testate amoebae. European journal of Protistology, 55, pp.105-117.

Lahr, D.J., 2021. An emerging paradigm for the origin and evolution of shelled amoebae, integrating advances from molecular phylogenetics, morphology and paleontology. Memórias do Instituto Oswaldo Cruz, 116.

Lahr, D.J., Kosakyan, A., Lara, E., Mitchell, E.A., Morais, L., Porfirio-Sousa, A.L., Ribeiro, G.M., Tice, A.K., Panek, T., Kang, S. and Brown, M.W., 2019. Phylogenomics and morphological reconstruction of Arcellinida testate amoebae highlight diversity of microbial eukaryotes in the Neoproterozoic. Current Biology, 29(6), pp.991-1001.

Lahr, D.J., Bosak, T., Lara, E. and Mitchell, E.A., 2015. The Phanerozoic diversification of silica-cycling testate amoebae and its possible links to changes in terrestrial ecosystems. PeerJ, 3, p.e1234.

Lawal, H.M., Schilde, C., Kin, K., Brown, M.W., James, J., Prescott, A.R. and Schaap, P., 2020. Cold climate adaptation is a plausible cause for evolution of multicellular sporulation in Dictyostelia. Scientific reports, 10(1), p.8797.

Li, G., Chen, L., Pang, K., Zhou, G., Han, C., Yang, L., Lv, W., Wu, C., Wang, W. and Yang, F., 2020. An assemblage of macroscopic and diversified carbonaceous compression fossils from the Tonian Shiwangzhuang Formation in western Shandong, North China. Precambrian Research, 346, p.105801.

Lutzoni, F., Nowak, M.D., Alfaro, M.E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A.E., Lewis, L.A., Swofford, D.L., Hibbett, D. and Hilu, K., 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. Nature Communications, 9(1), p.5451.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A. and Zdobnov, E.M., 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular biology and evolution, 38(10), pp.4647-4654.

Meade, A. and Pagel, M., 2022. Ancestral State Reconstruction Using BayesTraits. In Environmental Microbial Evolution: Methods and Protocols (pp. 255-266). New York, NY: Springer US.

Meyer, J.R. and Kassen, R., 2007. The effects of competition and predation on diversification in a model adaptive radiation. Nature, 446(7134), pp.432-435.

Mills, D.B., Boyle, R.A., Daines, S.J., Sperling, E.A., Pisani, D., Donoghue, P.C. and Lenton, T.M., 2022. Eukaryogenesis and oxygen in Earth history. Nature Ecology & Evolution, 6(5), pp.520-532.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A. and Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution, 37(5), pp.1530-1534.

Moczydłowska, M., Pease, V., Willman, S., Wickström, L. and Agić, H., 2017. A Tonian age for the Visingsö Group in Sweden constrained by detrital zircon dating and biochronology: implications for evolutionary events. Geological Magazine, 155(5), pp.1175-1189.

Morais, L., Lahr, D.J.G., Rudnitzki, I.D., Freitas, B.T., Romero, G.R., Porter, S.M., Knoll, A.H. and Fairchild, T.R., 2019. Insights into vase-shaped microfossil diversity and Neoproterozoic biostratigraphy in light of recent Brazilian discoveries. Journal of Paleontology, 93(4), pp.612-627.

Morais, L., Freitas, B.T., Fairchild, T.R., Toniolo, T.F., Campos, M.D.R., Prado, G.M.E.M., Silva, P.A.S., Rudnitzki, I.D., Lahr, D.J.G., Leme, J.D.M. and Philippot, P., 2021. Diverse vase-shaped microfossils within a Cryogenian glacial setting in the Urucum Formation (Brazil). Precambrian Research, 367, p.106470.

Mus, M.M. and Moczydowska, M., 2000. Internal morphology and taphonomic history of the Neoproterozoic vase-shaped microfossils from the Visings Group, Sweden. Norsk Geologisk Tidsskrift, 80(3), pp.213-228.

Mus, M.M., Moczydłowska, M. and Knoll, A.H., 2020. Morphologically diverse vase-shaped microfossils from the Russøya Member, Elbobreen Formation, in Spitsbergen. Precambrian Research, 350, p.105899.

Pagel, M., 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Systematic biology, 48(3), pp.612-622.

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. Proceedings of the National Academy of Sciences, 108(33), pp.13624-13629.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S. and Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. Nature protocols, 9(1), pp.171-181.

Porter, S., 2011. The rise of predators. Geology, 39(6), pp.607-608.

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. Proceedings of the Royal Society B: Biological Sciences, 283(1831), p.20160221.

Porter, S.M., 2020. Insights into eukaryogenesis from the fossil record. Interface Focus, 10(4), p.20190105.

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(3), pp.360-385.

Porter, S.M., Meisterfeld, R. and Knoll, A.H., 2003. Vase-shaped microfossils from the Neoproterozoic Chuar Group, Grand Canyon: a classification guided by modern testate amoebae. Journal of Paleontology, 77(3), pp.409-429.

Porter, S.M. and Riedman, L.A., 2019. Evolution: ancient fossilized amoebae find their home in the tree. Current Biology, 29(6), pp.R212-R215.

Pu, J.P., Bowring, S.A., Ramezani, J., Myrow, P., Raub, T.D., Landing, E., Mills, A., Hodgin, E. and Macdonald, F.A., 2016. Dodging snowballs: Geochronology of the Gaskiers glaciation and the first appearance of the Ediacaran biota. Geology, 44(11), pp.955-958.

Racki, G., 2005. Toward understanding Late Devonian global events: few answers, many questions. In Developments in Palaeontology and Stratigraphy (Vol. 20, pp. 5-36). Elsevier.

Reinhard, C.T., Planavsky, N.J., Gill, B.C., Ozaki, K., Robbins, L.J., Lyons, T.W., Fischer, W.W., Wang, C., Cole, D.B. and Konhauser, K.O., 2017. Evolution of the global phosphorus cycle. Nature, 541(7637), pp.386-389.

Reinhard, C.T., Planavsky, N.J., Ward, B.A., Love, G.D., Le Hir, G. and Ridgwell, A., 2020. The impact of marine nutrient abundance on early eukaryotic ecosystems. Geobiology, 18(2), pp.139-151.

Dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C. and Yang, Z., 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Current biology, 25(22), pp.2939-2950.

Riedman, L.A., Porter, S.M. and Calver, C.R., 2018. Vase-shaped microfossil biostratigraphy with new data from Tasmania, Svalbard, Greenland, Sweden and the Yukon. Precambrian Research, 319, pp.19-36.

Riedman, L.A. and Sadler, P.M., 2018. Global species richness record and biostratigraphic potential of early to middle Neoproterozoic eukaryote fossils. Precambrian Research, 319, pp.6-18.

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., 2018. Coupled Re-Os and U-Pb geochronology of the Tonian Chuar Group, Grand Canyon. Bulletin, 130(7-8), pp.1085-1098.

Rooney, A.D., Strauss, J.V., Brandon, A.D. and Macdonald, F.A., 2015. A Cryogenian chronology: Two long-lasting synchronous Neoproterozoic glaciations. Geology, 43(5), pp.459-462.

Rundle, H.D., Vamosi, S.M. and Schluter, D., 2003. Experimental test of predation's effect on divergent selection during character displacement in sticklebacks. Proceedings of the National Academy of Sciences, 100(25), pp.14943-14948.

Runnegar, B., 2000. Loophole for snowball Earth. Nature, 405(6785), pp.403-404.

Schmidt, A.R. and Dilcher, D.L., 2007. Aquatic organisms as amber inclusions and examples from a modern swamp forest. Proceedings of the National Academy of Sciences, 104(42), pp.16581-16585.

Schmidt, A.R., Ragazzi, E., Coppellotti, O. and Roghi, G., 2006. A microworld in Triassic amber. Nature, 444(7121), pp.835-835.

Schmidt, A.R., Schönborn, W. and Schäfer, U., 2004. Diverse fossil amoebae in German Mesozoic amber. Palaeontology, 47(2), pp.185-197.

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. Journal of Paleontology, 84(3), pp.363-401.

Sharma, M., Shukla, Y. and Sergeev, V.N., 2021. Microfossils from the Krol ‘A’of the Lesser Himalaya, India: Additional supporting data for its early Ediacaran age. Palaeoworld, 30(4), pp.610-626.

Siever, R., 1992. The silica cycle in the Precambrian. Geochimica et Cosmochimica Acta, 56(8), pp.3265-3272.

Singer, D., Seppey, C.V., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L., Blandenier, Q., Debroas, D., de Groot, G.A., De Vargas, C. and Domaizon, I., 2021.

Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. Environment International, 146, p.106262.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), pp.1312-1313.

Strassert, J.F., Irisarri, I., Williams, T.A. and Burki, F., 2021. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nature Communications, 12(1), p.1879.

Strauss, J.V., Rooney, A.D., Macdonald, F.A., Brandon, A.D. and Knoll, A.H., 2014. 740 Ma vase-shaped microfossils from Yukon, Canada: Implications for Neoproterozoic chronology and biostratigraphy. Geology, 42(8), pp.659-662.

Strullu-Derrien, C., Kenrick, P., Goral, T. and Knoll, A.H., 2019. Testate amoebae in the 407-million-year-old Rhynie Chert. Current Biology, 29(3), pp.461-467.

Tang, Q., Pang, K., Yuan, X. and Xiao, S., 2020. A one-billion-year-old multicellular chlorophyte. Nature Ecology & Evolution, 4(4), pp.543-549.

Tice, A.K., Žihala, D., Pánek, T., Jones, R.E., Salomaki, E.D., Nenarokov, S., Burki, F., Eliáš, M., Eme, L., Roger, A.J. and Rokas, A., 2021. PhyloFisher: a phylogenomic package for resolving eukaryotic relationships. PLoS Biology, 19(8), p.e3001365.

Vincent, W.F., Gibson, J.A.E., Pienitz, R., Villeneuve, V., Broady, P.A., Hamilton, P.B. and Howard-Williams, C., 2000. Ice shelf microbial ecosystems in the high arctic and implications for life on snowball earth. Naturwissenschaften, 87, pp.137-141.

Wang, H.C., Minh, B.Q., Susko, E., Roger, A.J. 2018. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. Syst Biol. 67(2), pp.216-235.

Wang, H., Liu, A., Li, C., Feng, Q., Tang, S., Cheng, M. and Algeo, T.J., 2021. A benthic oxygen oasis in the early Neoproterozoic ocean. Precambrian Research, 355, p.106085.

Wang, K., Xu, H.H., Liu, B.C., Bai, J., Wang, Y., Tang, P., Lu, J.F. and Wang, Y., 2023. Shallow-marine testate amoebae with internal structures from the Lower Devonian of China. Iscience, 26(5).

Wellman, C.H. and Strother, P.K., 2015. The terrestrial biota prior to the origin of land plants (embryophytes): a review of the evidence. Palaeontology, 58(4), pp.601-627.

White, D.A., Elrick, M., Romaniello, S. and Zhang, F., 2018. Global seawater redox trends during the Late Devonian mass extinction detected using U isotopes of marine limestones. Earth and Planetary Science Letters, 503, pp.68-77.

Xiao, S. and Tang, Q., 2018. After the boring billion and before the freezing millions: evolutionary patterns and innovations in the Tonian Period. Emerging Topics in Life Sciences, 2(2), pp.161-171.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution, 24(8), pp.1586-1591.

Zachar, I. and Szathmáry, E., 2017. Breath-giving cooperation: Critical review of origin of mitochondria hypotheses: Major unanswered questions point to the importance of early ecology. Biology Direct, 12, pp.1-26.

Žárský, J., Žárský, V., Hanáček, M. and Žárský, V., 2022. Cryogenian glacial habitats as a plant terrestrialisation cradle–the origin of the anydrophytes and Zygnematophyceae split. Frontiers in Plant Science, 12, p.735020.

1225 Zhang, C., Rabiee, M., Sayyari, E. and Mirarab, S., 2018. ASTRAL-III: polynomial
1226 time species tree reconstruction from partially resolved gene trees. BMC
1227 bioinformatics, 19(6), pp.15-30.

Figures

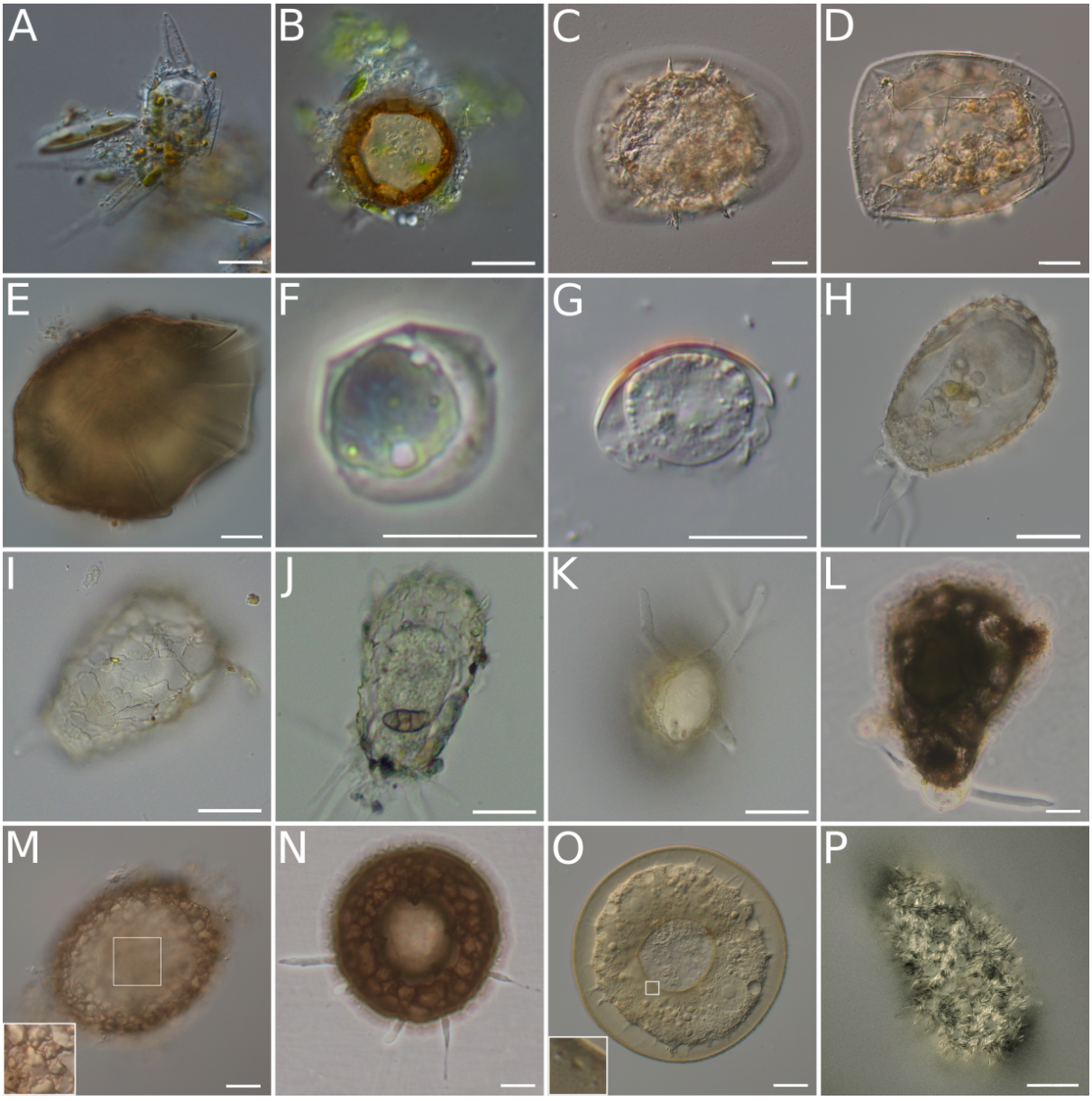

Figure 1. Sampled Testate Amoebae — Arcellinida and Corycidia. The pictured organisms were photodocumented prior to molecular processing and represent individuals or cultures from which the transcriptomic data was obtained. The scale bars represent 20 µm, except when specified. **A.** *Phryganella paradoxa* T, lateral view. **B.** *Phryganella acropodia* A, apertural view. **C - D.** *Microcorycia aculeata*, dorsal view (C) and apertural view (D); **E.** *Microcorycia flava*, dorsal view focusing on the flexible part of the shell; **F.** *Spumochlamys sp.*, dorsal view; **G.** *Spumochlamys bryora*, lateral view; **H - K.** *Heleopera lucida* comb. nov. (previously *Difflugia lucida*), lateral view focusing on the cell within the shell (H), lateral view focusing on the shell (I - J), and apertural view focusing on the compressed aspect of the shell (K); **L.** *Difflugia* cf. *capreolata*, lateral view, scale bar 40 µm; **M.** *Netzelia lobostoma*, lateral view, white square focusing on details of the shell; **N.** *Cyclopyxis sp.*, apertural view, scale bar 40 µm; **O.** *Galeripora sp.*, apertural view, white square focusing on the pores which surround the shell aperture; **P.** *Trichosphaerium sp.* KSS. Measured

1244 morphometric characteristics of the newly sequenced testate amoebae taxa are
1245 present on **Dataset S01, Table S2**.

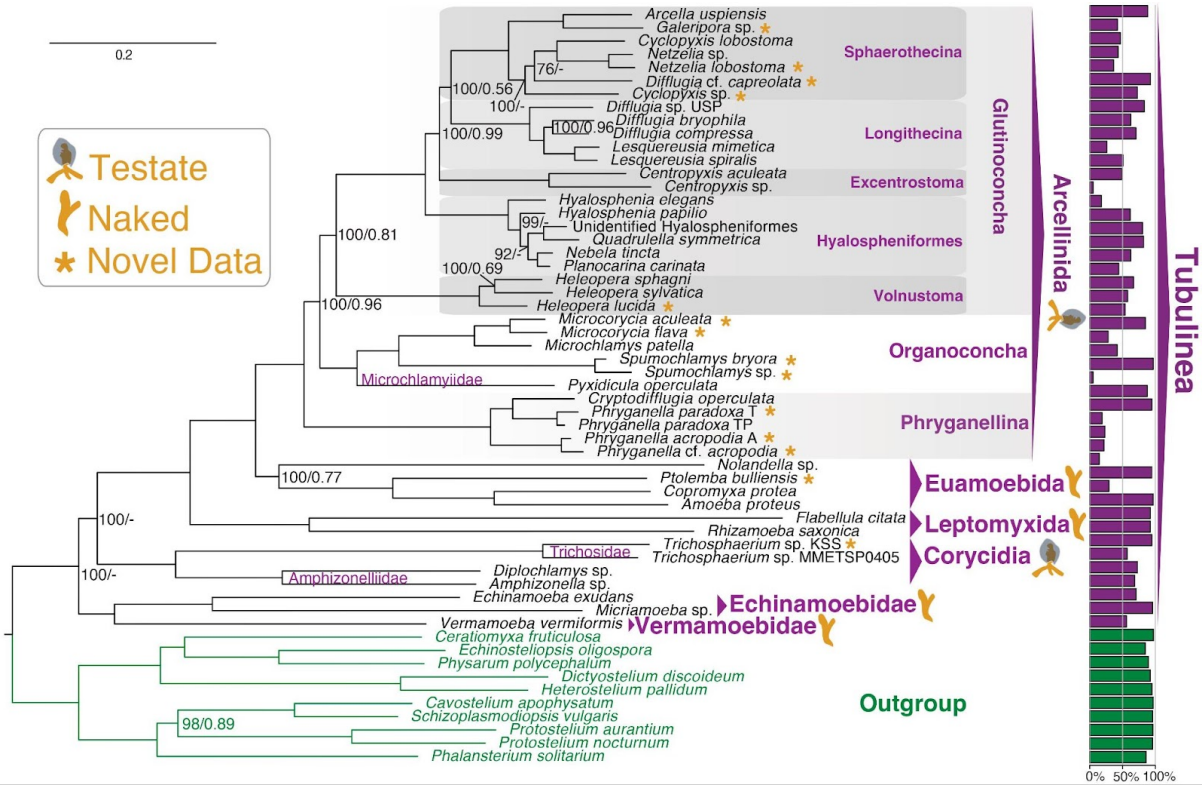

Figure 2. The Tree of amoebozoan testate amoebae. 226 genes (70,428 amino acid sites) phylogeny of amoebozoan testate amoebae rooted with Evosea (Amoebozoa). The tree was initially built using IQ-TREE2 v. 2.0-rc1 under the LG+C20+G4 model of protein evolution and further used to infer a Posterior Means Site Frequency model using the ML model LG+C60+G4+PMSF. Topological support was assessed by 100 Maximum Likelihood Real Bootstrap Replicates (MLRB) and local posterior probability values (LPP) calculated using ASTRAL-III v. 5.7.3, and are shown in the format (MLRB/LPP).

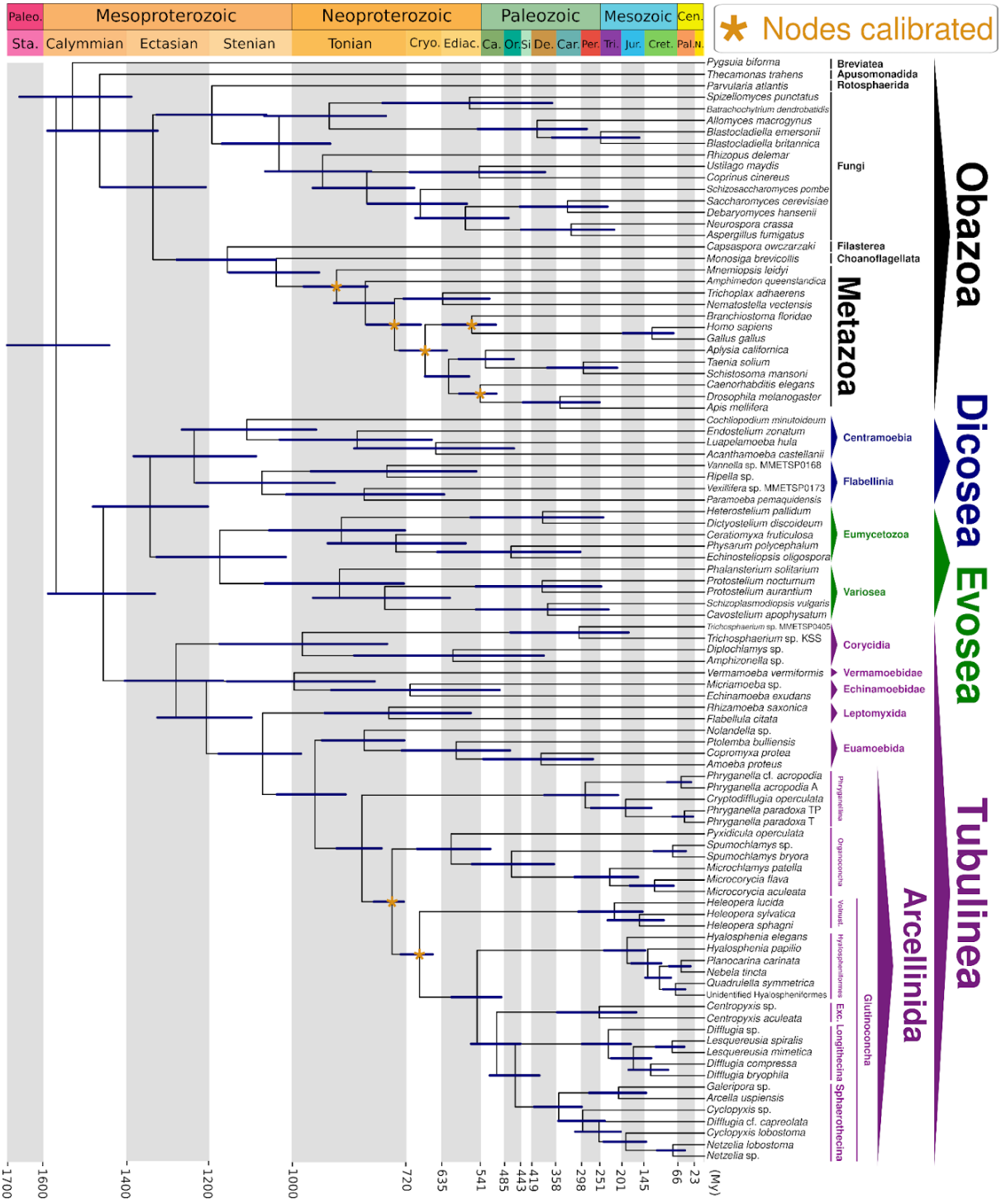

Figure 3. Amorphea time-calibrated tree. Divergence times displayed here were inferred under autocorrelated relaxed clock model, applying a Skew-Normal distribution for node calibration and drift parameter of α = 2 and β = 2. Bars at nodes are 95% highest probability density confidence intervals. Asterisks indicate the nodes calibrated based on external fossil information. The results and time-calibrated trees for all experiments are present in the supplemental material (**Dataset S01, Table S11 and Appendix S01, Fig. S6 - S14**). Abbreviations: Exc.- Excentrotoma; Vonust. - Volnustoma; Paleo. - Paleoproterozoic; Sta. - Statherian; Cryo. - Cryogenian; Ediac. - Ediacaran; Ca. - Cambrian; Or. Ordovician; Si - Silurian; De. - Devonian; Car. Carboniferous; Per. - Permian; Tri. - Triassic; Jur. - Jurassic;

Cret. - Cretaceous; Cen. - Cenozoic; Pal. - Paleogene; N. - Neogene; My - Million Years.

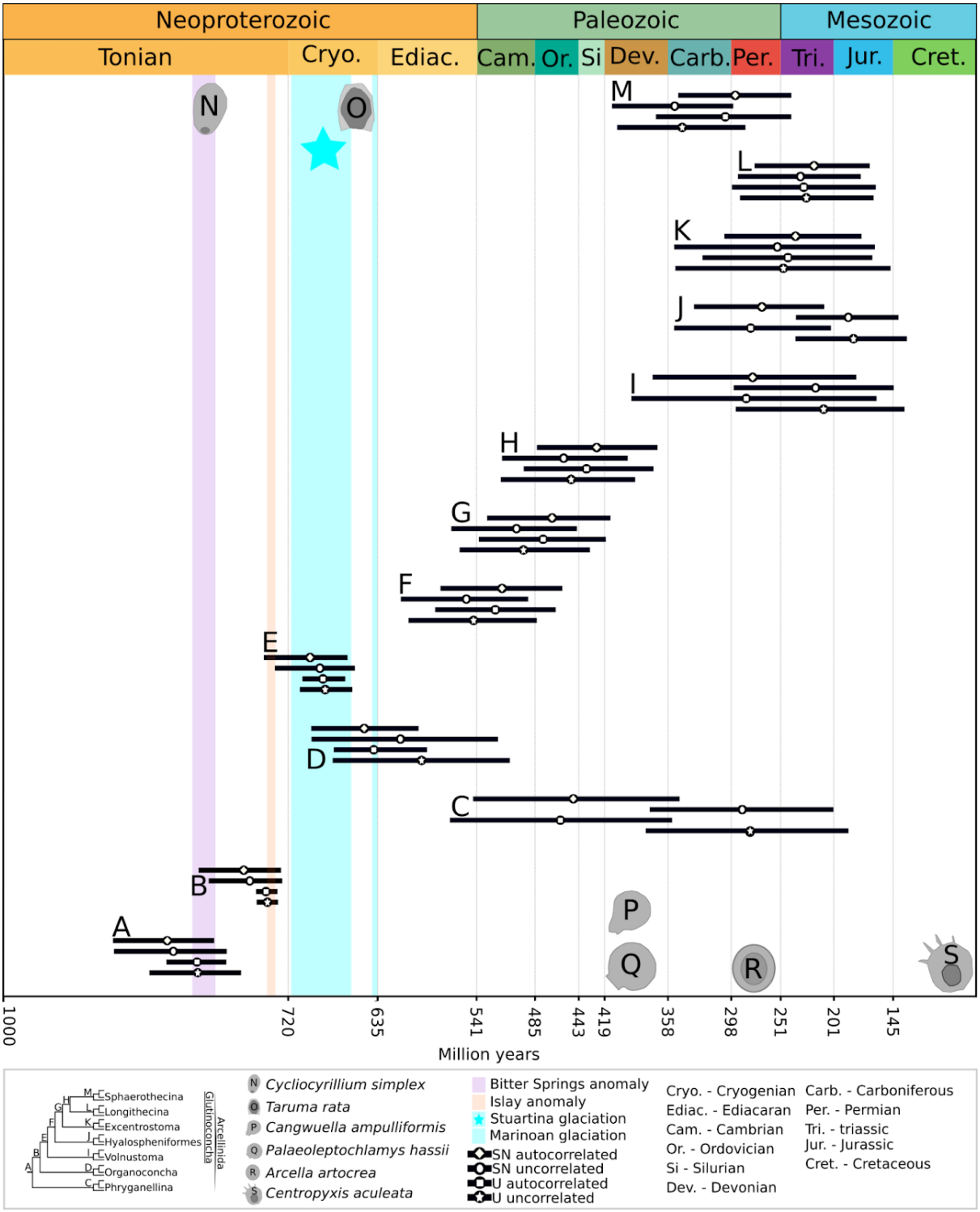

Figure 4. Estimated times using different calibration strategies (Uniform - U - or Skew-Normal - SN) and clock models (uncorrelated or autocorrelated) for nodes of interest throughout geological time, highlighting key carbon anomalies, ice age events, and fossil record. **A.** Arcellinida node; **B.** Organoconcha+Glutinoconcha node; **C.** Phryganellina node. **D.** Orgnoconcha node; **E.** Glutinoconcha node; **F.** Hyalospheniformes - Sphaerothecina node; **G.** Excentrostoma - Sphaerothecina node; **H.** Longithecina - Sphaerothecina node; **I.** Volnustoma node; **J.** Hyalospheniformes node; **K.** Excentrostoma node; **L.** Longithecina node; **M.** Sphaerothecina node. **N - O.** Time range of Neoproterozoic vase-shaped

microfossils represented by the maximum date of ~800 million years (my) reported for *Cyclocyrillium simplex* (N) and the minimum date of 638 my reported for *Taruma rata* (O). **P - Q.** Devonian vase-shaped microfossils represented by the marine fossil *Cangwuella ampulliformis* (P) and the 407-million years old terrestrial *Palaeoleptochlamys hassii* (Q). **R - S.** Time range of late Paleozoic and Mesozoic fossils of extant arcellinids represented by the maximum date of ~298 my reported for the putative fossil of *Arcella artocrea* (R) (Farooqui et al. 2014) and the minimum date of ~93 my reported for the putative fossil of *Centropyxis aculeata* (S) (Schmidt and Dilcher 2007). The bars are 95% highest probability density confidence intervals. The results and time-calibrated trees for all experiments are present in the supplemental material (**Dataset S01, Table S11 and Appendix S01, Fig. S6 - S14**).

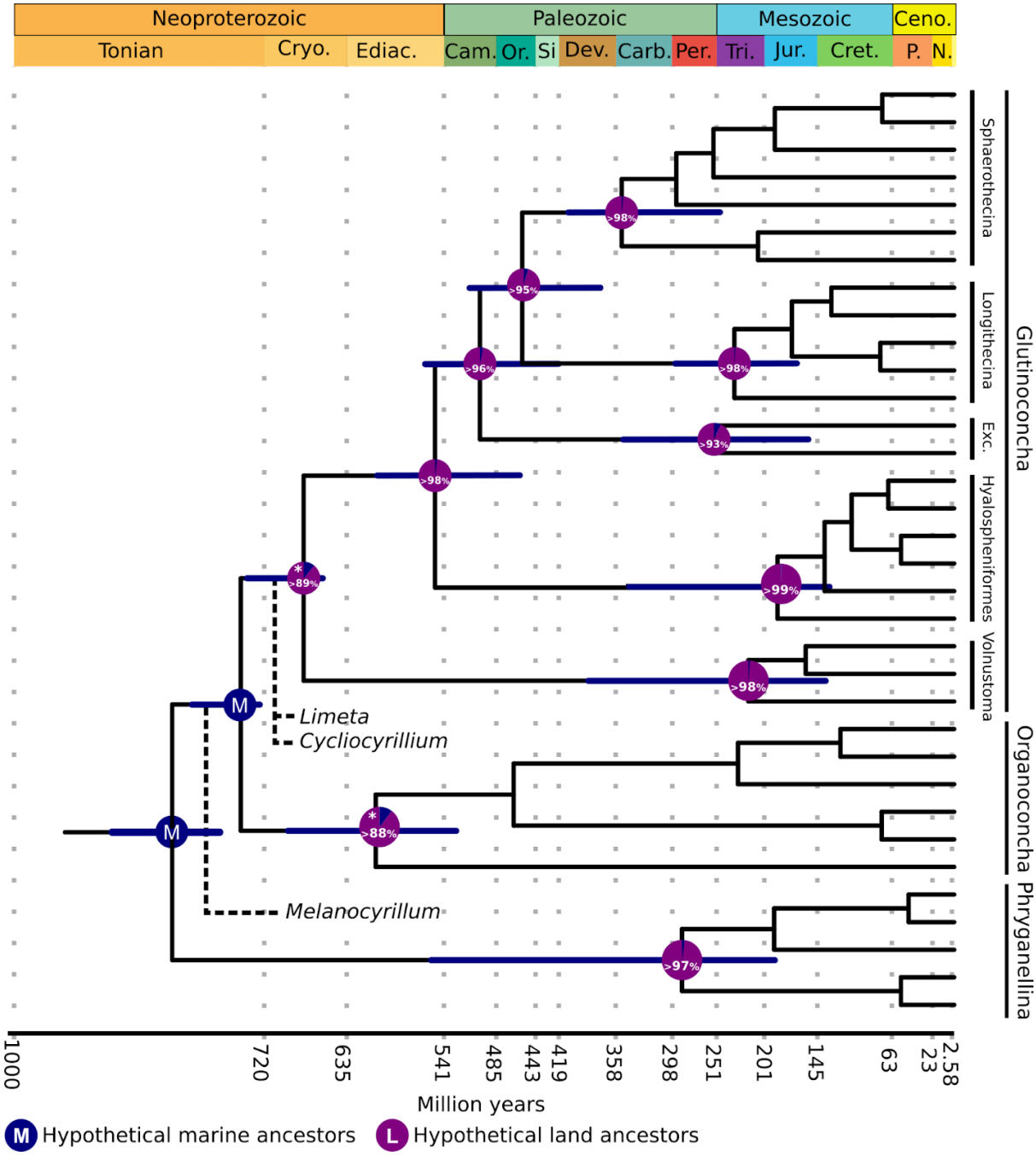

Figure 5. Ancestral reconstructions of Arcellinida habitats using BayesTraits. Pie charts at each node indicate the mean probabilities of a hypothetical marine ancestor (blue) or a hypothetical land ancestor (purple). Asterisks indicate nodes for which a marine or land ancestor is statistically equally plausible based on the fossilization approach and Likelihood Ratio test (LRT) implemented to test alternative hypothetical scenarios. Bars at nodes are 95% highest probability density confidence intervals based on the 16 experiments considering the different calibration strategies, drift parameters, and clock models. The complete results are shown on **Dataset S01, Table S12 and Appendix S01, Fig. S16**.