PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Galindo LJ, Torruella G, Moreira D, Eglit Y, Simpson AGB, Völcker E, Clauß S, López-García P. 2019 Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi. *Phil. Trans. R. Soc. B* **374**: 20190094.
http://dx.doi.org/10.1098/rstb.2019.0094

Accepted: 26 May 2019

One contribution of 18 to a discussion meeting issue 'Single cell ecology'.

Subject Areas:
evolution, ecology, genomics, microbiology, taxonomy and systematics

Keywords:
nucleariid amoebae, opisthokont, single-cell, *Pompholyxophrys*, *Lithocolla*, Rotosphaerida

Authors for correspondence:
Guifré Torruella
e-mail: guifftc@gmail.com
Purificación López-García
e-mail: puri.lopez@u-psud.fr

†These authors contributed equally to the study.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4646486.

THE ROYAL SOCIETY PUBLISHING

Combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of fungi

Luis Javier Galindo[1,†], Guifré Torruella[1,†], David Moreira[1], Yana Eglit[2], Alastair G. B. Simpson[2], Eckhard Völcker[3], Steffen Clauß[3] and Purificación López-García[1]

[1]Unité d'Ecologie, Systématique et Evolution, CNRS, Université Paris-Sud, Université Paris-Saclay, AgroParisTech, 91400 Orsay, France
[2]Department of Biology, and Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Nova Scotia, Canada
[3]Penard Laboratory, Cape Town, 7708 South Africa

LJG, 0000-0003-3642-3200; GT, 0000-0002-6534-4758; DM, 0000-0002-2064-5354; YE, 0000-0002-2178-3062; AGBS, 0000-0002-4133-1709; EV, 0000-0002-6841-1389; PL-G, 0000-0002-0927-0651

Nucleariid amoebae (Opisthokonta) have been known since the nineteenth century but their diversity and evolutionary history remain poorly understood. To overcome this limitation, we have obtained genomic and transcriptomic data from three *Nuclearia*, two *Pompholyxophrys* and one *Lithocolla* species using traditional culturing and single-cell genome (SCG) and single-cell transcriptome amplification methods. The phylogeny of the complete 18S rRNA sequences of *Pompholyxophrys* and *Lithocolla* confirmed their suggested evolutionary relatedness to nucleariid amoebae, although with moderate support for internal splits. SCG amplification techniques also led to the identification of probable bacterial endosymbionts belonging to Chlamydiales and Rickettsiales in *Pompholyxophrys*. To improve the phylogenetic framework of nucleariids, we carried out phylogenomic analyses based on two datasets of, respectively, 264 conserved proteins and 74 single-copy protein domains. We obtained full support for the monophyly of the nucleariid amoebae, which comprise two major clades: (i) *Parvularia–Fonticula* and (ii) *Nuclearia* with the scaled genera *Pompholyxophrys* and *Lithocolla*. Based on these findings, the evolution of some traits of the earliest-diverging lineage of Holomycota can be inferred. Our results suggest that the last common ancestor of nucleariids was a freshwater, bacterivorous, non-flagellated filose and mucilaginous amoeba. From the ancestor, two groups evolved to reach smaller (*Parvularia–Fonticula*) and larger (*Nuclearia* and related scaled genera) cell sizes, leading to different ecological specialization. The *Lithocolla* + *Pompholyxophrys* clade developed exogenous or endogenous cell coverings from a *Nuclearia*-like ancestor.

This article is part of a discussion meeting issue 'Single cell ecology'.

1. Introduction

Nucleariids are non-flagellated, free-living, phagotrophic filose amoebae [1]. 18S rRNA gene molecular phylogenies placed *Nuclearia* as a deep branch within the opisthokonts [2,3], particularly as sister clade to fungi [4,5], as subsequently corroborated by phylogenomic analyses [6,7]. They are thus part of the Holomycota (Nucletmycea), the opisthokont lineage containing fungi and its relatives [8]. The last opisthokont common ancestor probably was a phagotrophic cell with a single flagellum and polarized cell shape, a feature that is shared with the deepest-branching fungi and their aphelid [9] and rozellid [10] relatives [11].

Therefore, nucleariids underwent substantial evolutionary change from that ancestor which we need to understand to infer the global evolutionary history of Holomycota, including key biological traits such as the fungal multicellularity [12] or the transition to parasitism [13].

So far, only a few studies of nucleariid species are available, including some morphological descriptions [1,14–16] and molecular phylogenetic studies [2–5,17–20]. Nevertheless, many *incertae sedis* species await molecular characterization [21–25]. Historically, owing to the lack of clear external features distinguishable under optical microscopy, nucleariids have been assigned to a variety of amoeboid taxa [26,27]. *Nuclearia* Cienkowski, 1865, is the most commonly observed and characterized genus [1,28,29]. Until the late twentieth century, this genus was associated with other naked filose amoebae in several different and conflicting taxonomies [8,30,31]. Patterson, using transmission electron microscopy data, separated nucleariids from other filose amoebae, united distinct genera (e.g. *Nuclearella* Frenzel, 1897; *Nuclearina* Frenzel, 1897, *Nucleosphaerium* Cann and Page, 1979) into *Nuclearia*, clarified its systematics [1,14], and confirmed its relationship with *Vampyrellidium perforans* [16,32] (not to be confused with the cercozoan *Vampyrella* [33]) and the scale-bearing filose amoeba *Pompholyxophrys* [15]. It was further proposed that other silica-scaled amoebae with a secreted silica-mineral coat composed of silicified particles (i.e. idiosomes), like *Pinaciophora* and *Rabdiophrys* (not to be confused with the centrohelid *Raphidiophrys* [34]) were related to *Nuclearia* [19,20,22,25,33]. In agreement with Patterson, Page grouped *Nuclearia* and *Pompholyxophrys* inside the Cristidiscoidida [35]. Later, Mikrjukov suggested that *Elaeorhanis* [36] and *Lithocolla* [37], two scaled filose amoebae with coats composed of aggregated exogenous material (i.e. xenosomes), were also related to nucleariids [22] and claimed priority of the name Rotosphaerida over Cristidiscoidea to group all nucleariid amoebae [38]. Since then, molecular phylogeny analyses have placed *Fonticula* [5,26,39] and *Parvularia* [20,40] together with *Nuclearia* as a sister clade to the rest of Holomycota, although the 18S rDNA gene marker could not resolve the internal relationships between nucleariid clades.

To solve some of these uncertainties, we sampled putative nucleariid species from freshwater and marine environments, including naked (*Nuclearia* sp.) and scale-bearing (*Pompholyxophrys* sp. and *Lithocolla* sp.) amoebae. We obtained molecular data using traditional culturing and single-cell genomic techniques and inferred a robust phylogenetic framework that leads to an improved understanding of the biodiversity of these organisms and a clarification of the systematics of the whole nucleariid clade.

2. Methods

(a) Biological material

Lithocolla globosa (electronic supplementary material, figure S1) was isolated from a marine sediment sample from Splitnose Point near Ketch Harbour, Nova Scotia, Canada (44.477 N, 63.541 W) and grown in culture with *Navicula pseudotenelloides* NAVIC33 as food source. Single *Lithocolla* cells were micromanipulated with an Eppendorf PatchMan NP2 micromanipulator using a 110 µm VacuTip microcapillary (Eppendorf) in an inverted microscope Leica Dlll3000 B, cells were washed in clean water drops before storing them into individual tubes. *Pompholyxophrys* cells (electronic supplementary material, figure S2) were collected from a freshwater lake near Zwönitz, Germany both by manual micromanipulation and by using the previously described equipment into tubes in sets of 20–30 cells or as single cells (without washing steps when manually collected) [41]. Both *Nuclearia delicatula* and *Nuclearia thermophila* (electronic supplementary material, figure S3) were isolated from the mixed freshwater culture JP100 from Sciento (UK) maintained with *Oscillatoria*-like filamentous cyanobacteria, and with the presence of *Poterioochromonas*-like (stramenopile) and *Echinamoeba*-like (amoebozoan) contaminants (electronic supplementary material, figure S3A–D). *Nuclearia thermophila* was isolated by micromanipulation (using previously cited equipment) from the initial JP100 culture. Individual *Nuclearia pattersoni* XT1 cells were collected after washing steps using the previously described micromanipulator equipment from the intestine of a dissected *Xenopus tropicalis* tadpole grown in the laboratory.

(b) DNA and RNA purification, 18S rRNA gene amplification and sequencing

To assess the identity of our nucleariid amoebae, we first obtained 18S rRNA gene sequences from cultures and single-cell isolates by polymerase chain reaction (PCR) amplification using distinct combinations of primers 82F (5′-GAAACTGCGAATGGCTC-3′), 612F (5′-GCAGTTAAAAAGCTCGTAGT-3′), 1379R (5′-TGTGTACAAAGGGCAGGGAC-3′) and 1498R (5′-CACCTACGGAAACCTTGTTA-3′). Amplicon cloning was performed with the TOPO-TA cloning kit (Invitrogen) following the instructions of the manufacturers. RNA was purified from the cultures of *N. delicatula*, *N. thermophila*, the mixed culture of *L. globosa* and its food *Navicula* sp. using the kit RNeasy Micro (Qiagen, Venlo, The Netherlands) including a DNAse treatment. In addition, whole transcriptome amplification (WTA) and whole genome amplification (WGA) of micromanipulated single cells was carried out using REPLI-g WTA/WGA Kits (Qiagen) for *N. pattersoni*, *L. globosa* and *Pompholyxophrys*. For a batch of 20 *Pompholyxophrys* cells, DNA was first released with the PicoPure DNA extraction kit (Applied Biosystems) and then WGA was performed (table 1). Paired-end sequences were obtained by polyA RNAseq or Nextera library construction and sequencing was performed with an Illumina HiSeq SBS Kit v4 2500 2 × 125 bp by Eurofins Genomics (Ebersberg, Germany) or by the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain) for the Nextera libraries.

(c) Molecular data assembly, decontamination and annotation

Reads were screened with FastQC [42] before and after quality/Illumina adapter trimming with Trimmomatic v0.33 [43] in paired-end mode with the following parameters: ILLUMINACLIP:adapters.fasta:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:28. Resulting reads were assembled with SPAdes 3.9.1 [44]. To predict protein sequences, we co-assembled the *L. globosa* dataset and sequences from the two *Pompholyxophrys* species (*P.* sp. and *P. punicea*), after verifying that they belonged to the same species by 18S rRNA gene phylogenetic analyses. Two co-assembly rounds were performed before and after decontamination by BlobTools v0.9.19 [45]. In the case of *Lithocolla*, the predicted *Navicula* proteome was used to further eliminate sequences from its prey using BLASTp [46]. Decontaminated predicted protein sequences were obtained using Transdecoder v2 (http:transdecoder.github.io) with default parameters and Cd-hit v4.6 [47] with 100% identity. Proteins were annotated with the EggNOG v4.5 [48] database with DIAMOND as mapping mode, and the taxonomic scope to adjust automatically (table 1). We have deposited the new nucleariid 18S rRNA gene sequences in GenBank with accession numbers MK547173–MK547179, and *Pompholyxophrys* bacterial endosymbionts 16S rRNA gene

Table 1. List of protist single-cells/culture samples, sequence statistics and number of phylogenetic markers retrieved from genome/transcriptome datasets. (WTA stands for whole transcriptome amplification and WGA for whole genome amplification.)

						individual species assemblies				
cell/culture identifier	DNA/RNA (culture or few/single-cell)	read-pairs	yield (Gb)	GBE 264 markers (%)	SCPD 74 markers (%)	no. of contigs/ scaffolds	no. of proteins	no. of 'clean' proteins	GBE 264 markers (%)	SCPD 74 markers (%)
L. globosa MK547176										
culture SnPLi with *Navicula*	RNAseq (culture)	41 033 000	23 854	199 (75.37)	60	70 737	72 580	9277	211 (79.92)	65 (87.83)
LG140, LG144, LG145	WGA (few-cells)	77 313 319	15 462	35 (13.25)	27					
LG147	WTA (single-cell)	37 212 410	18 681	81 (30.68)	24					
N. pseudotenelloides										
NAVIC33 culture	RNAseq (culture)	44 463 054	13 428	—	—	36 618	3350	—	—	—
Pompholyxophrys sp. MK547174										
LG126	WGA (single-cell)	73 107 816	14 621	3 (1.13)	1 (1.34)	86 851	39 399	1094	82 (31.06)	19 (25.67)
LG130	WTA (single-cell)	37 135 207	18 642	80 (30.3)	18 (24.32)					
P. punicea MK547175										
LG129	WTA (single-cell)	39 500 923	19 829	125 (47.34)	31 (41.89)	227 098	82 091	3121	144 (54.54)	34 (45.94)
20cellsWGA	WGA (few-cells)	47 517 660	23 854	36 (13.63)	9 (12.16)					
LG127	WGA (single-cell)	68 532 623	13 706	0	0	2356	—	—		
N. pattersoni XT1 MK547179										
XT1	WTA (single-cell)	7 062 454	4237	33 (12.5)	0	453 169	41 060	—	—	—
N. delicatula JP100 MK547177										
culture JP100 contaminated with other eukaryotes	RNAseq (culture)	83 127 257	10 390	234 (88.63)	59 (79.72)	56 177	54 191	—	—	—
N. thermophila JP100 MK547178										
culture JP100 cleaned Sep/Nov	RNAseq (culture)	128 552 236	32 139	251 (95.07)	72 (97.29)	70 205	65 150	—	—	—

sequences with accession numbers MK616425–MK616429. Transcriptome and genome sequence data have been submitted to NCBI SRA under the Bioproject PRJNA517920. Decontaminated predicted proteins, phylogenetic datasets and trees have been deposited in Figshare [49].

(d) 18S and 16S rRNA gene phylogenies

We compiled the 18S rRNA gene sequences included in three previous studies of nucleariids, including environmental sequences [20,50,51], and aligned them with our newly obtained sequences. We generated a dataset of 207 sequences and 1756 bp. For bacterial endosymbionts, we used the 16S rRNA gene sequences of *Nuclearia* sp. endosymbionts identified in the previous study [28] as queries to find homologues by BLASTn [46] in all nucleariid assemblies (*Parvularia*, 2 *Nuclearia* and 2 *Fonticula* species). Selected sequences of potential endosymbionts along with their closest BLAST hits were included in phylogenetic trees to have representatives of closely related bacteria. We worked with three datasets, one complete dataset of 100 sequences and 1503 bp, and two subsets of this first dataset for the Chlamydiae group (18 sequences and 1454 bp) and the Rickettsiales group (26 sequences and 1390 bp). All alignments were made using MAFFT v7 [52]. Trimming of the alignment was performed manually for the 18S rRNA gene sequences and with TrimAl in automated1 mode [53] for the 16S rRNA gene sequences.

Maximum-likelihood (ML) phylogenetic trees were inferred using IQTree v1.6 [54]. For the 18S rRNA gene ML trees, the GTR + R8 + F0 evolutionary model was used to assess branch support with 1000 ultrafast bootstraps (UFBS), single branch tests SH-like approximate likelihood ratio test based on the Shimodaira-Hasegawa (SH) algorithm for tree comparison [55] and approximate Bayes test [56]. In addition, 1000 non-parametric bootstraps [57] were obtained with the TIM3 + F + I + G4 model as the best-fitting one based on the Bayesian information criterion (BIC) from ModelFinder [58]. For the 16S rRNA gene ML trees, the best fit model chosen by BIC [59] was the GTR model (for the complete dataset and for the Rickettsiales dataset) and the TIM3 model (for the Chlamydiae dataset) both with F + I + G4. Bayesian inference (BI) phylogenies were inferred using MrBayes v3.2.6 [60]. For both the 16S and 18S rRNA gene BI trees, the GTR + G + I model was used, with four Markov chain Monte Carlo (MCMC) chains for 1 000 000 generations, sampling every 100 trees and burn-in of the first 2500 saved trees.

(e) Phylogenomic analyses

Two distinct datasets, a dataset modified from Mikhailov *et al*. [9,61] (dataset GBE: 264 protein alignments) and Torruella *et al*. [9] (dataset SCPD: 74 single-copy domains) were updated with data from seven new nucleariid species. For both datasets, orthologues were identified by tBLASTn, aligned with MAFFT v7 and trimmed with TrimAl with the automated1 option. Alignments were visualized and manually edited with Geneious v6.0.6 and single gene trees obtained with FastTree v2.1.7 [62] with default parameters. Single gene trees were then manually checked and corrected for paralogous and/or contaminating sequences. All datasets were assembled into a supermatrix with Alvert.py from the package Barrel-o-Monkeys [63]. Resulting matrices were called SCPD21_23481aa and GBE22_97918aa. No orthologous markers were retrieved for *N. pattersoni* XT1 in the SCPD dataset. For both datasets, BI phylogenetic trees were reconstructed using PhyloBayes-MPI v1.5 [64] under the CAT-Poisson model, two MCMC chains for each dataset were run for greater than 15 000 generations, saving one every 10 trees. Analyses were stopped once convergence thresholds were reached after a burn-in of 25% (i.e. maximum discrepancy less than 0.1 and minimum effective size greater than 100 calculated using bpcomp). ML phylogenetic trees were inferred with IQ-Tree v1.6 under the LG + R5 + C60 model. Statistical support was obtained with 1000 UFBS [65] and 1000 replicates of the SH-like approximate likelihood ratio test [56]. All trees were visualized with FigTree [66].

Fully detailed materials and methods can be found in the electronic supplementary material.

3. Results and discussion

(a) *Pompholyxophrys* and *Lithocolla* are free-living nucleariid amoebae

We obtained 18S rRNA gene sequences from two cultures of *Nuclearia* (*N. delicatula* JP100 and *N. thermophila* JP100), one single cell from another *Nuclearia* species (*N. pattersoni* XT1), two single cells and one few cells (20 cells) from *Pompholyxophrys* species and one culture of *L. globosa* (table 1 and the electronic supplementary material). This represents the first molecular data for both *Pompholyxophrys* and *Lithocolla*. We included our new sequences in a large 18S rRNA gene dataset containing all available nucleariid sequences. Phylogenetic analyses of this dataset confirmed the monophyly of *Nuclearia* species and their relationship with the environmental sister clade NUC-1, whereas the environmental clade NUC-2 was sister to the *Parvularia* clade (figure 1 and electronic supplementary material, figure S4A–C) [20]. *Fonticula alba* exhibited a long branch sister to the group containing the *Pompholyxophrys* and *Lithocolla* sequences. This group also contained several environmental sequences originally called marine fonticulids [19] but recent metabarcoding studies [45,46] have found freshwater representatives intermixed with the marine ones. The morphology and behaviour of *Lithocolla* cells in culture strongly resemble *Nuclearia* (electronic supplementary material, figure S1). Also its exogenous aggregative cell covering suggests a higher similarity to naked *Nuclearia* than to *Pompholyxophrys* [22]. However, our results support a closer phylogenetic relationship of *Pompholyxophrys* and *Lithocolla* as compared to *Nuclearia* (figure 1). Nevertheless, the internal topology of this large Pompholyxophrys–Lithocolla group, which additionally encompasses two large clades of environmental sequences (with currently not known representative species), remains unclear. This is probably owing to the limited signal of the 18S rRNA marker at this level of resolution.

Although some *Nuclearia* have been found in brackish water [1], all published environmental sequences clustering with *Nuclearia* come from soil or freshwater systems (as deduced from sequence metadata deposited in GenBank) and *Parvularia*, as *Nuclearia*, seems to be exclusively freshwater. *Pompholyxophrys* has also been found only in freshwater systems [15,22] but it is sister to a clade of marine environmental sequences (figure 1 and electronic supplementary material, figure S4A–C). Although our *Lithocolla* sequence clustered within an exclusively marine clade, this genus has been observed also in freshwater environments [37].

Nuclearia species are capable of growing in eutrophic and/or contaminated environments. For example, they can ingest toxic filamentous cyanobacteria that can thrive in perturbed environments as their sole food source [29,41]. This capability appears to be related to their association with symbiotic bacteria that degrade toxic metabolites, as microcystin, contained in the cyanobacteria ingested by *Nuclearia* [28,29,67]. Our *N. pattersoni* single cell was recovered by micromanipulation from the gut content of a dissected *X. tropicalis* tadpole grown in the laboratory. When collected, this cell was alive and moving, suggesting that it was a commensal in the amphibian gut. In agreement

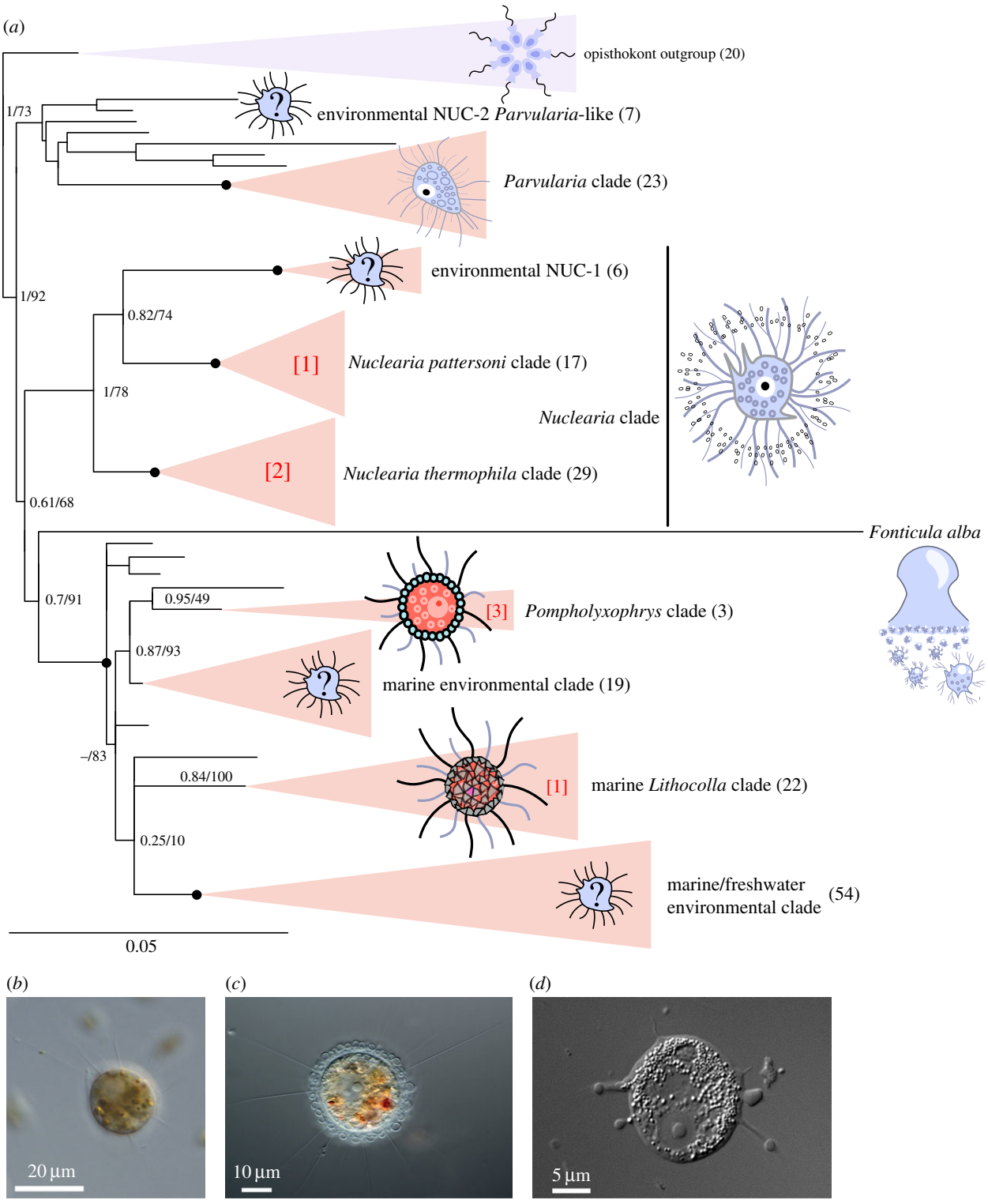

Figure 1. (*a*) ML phylogenetic tree of nucleariid 18S rRNA gene sequences. The tree was reconstructed from an alignment of 1756 nucleotide positions of 207 sequences, including the three *Nucleariа*, three *Pompholyxophrys* and one *L. globosa* sequences obtained in this study as well as all nucleariid sequences available in GenBank with the GTR + R8 model. Major groups were collapsed (the complete tree is shown in the electronic supplementary material, figure S4A). Statistical supports are Bayesian posterior probabilities (PP) obtained under the GTR + G + I model on the left and ML ultrafast bootstrap (UFBS) on the right. Branches with support values higher or equal to 0.99 PP and 95% UFBS are indicated by black dots. Clades without known representatives are indicated with a question mark. The number of sequences is shown in parenthesis and the number of sequences obtained in this study is shown in red brackets. (*b–d*) From left to right optical microscopy images of *L. globosa*, *P. punicea* and *N. thermophila* JP100. Scale bars are 20, 10 and 5 µm, respectively. (Online version in colour.)

with this idea, *N. pattersoni* was originally described from fish gills [17]. Whether *Nucleariа* maintains preferential ecological interactions with metazoans or not remains to be determined. By contrast, multiple observations suggest that *Pompholyxophrys* species, as many other silica-based scale-bearing amoebae, are free-living and develop in clear freshwater bodies, wet *Sphagnum* moss, and peat bogs [68,69].

(b) Endosymbiotic bacteria in nucleariids

Single-cell approaches allowed us to examine an important ecological aspect of these amoebae, namely their relationships with intracellular bacteria. Bacterial endosymbionts have been previously observed in nucleariids [31], with the first molecular data coming from a *Rickettsia* endosymbiont in *N. pattersoni* [17] and the gammaproteobacterium *Candidatus*

Endonucleobacter rarus in *N. thermophila* [67]. Dirren and Posch [28] characterized several bacterial endosymbionts in different species and strains of *N. thermophila* and *N. delicatula*. They observed that the specificity of the symbiosis might vary depending on the host *Nuclearia* species. In some cases, the same endosymbiont species was found in the same host (*N. thermophila*) from different places, but in other cases, the same host (*N. delicatula*) may harbour different endosymbionts.

We generated four single/few-cell transcriptomes (SCT) and four single/few-cells genomes (SCG) for *Lithocolla*, *Pompholyxophrys* and *Nuclearia* (table 1), and using as a reference the bacterial endosymbiont 16S rRNA gene dataset from Dirren and Posch [28], we searched for endosymbiotic candidate species. However, we not only searched in our SCTs/SCGs but also in our RNAseq data and in all other nucleariid data available in public databases (*Parvularia*, two *Fonticula* species and two *Nuclearia* species).

We retrieved 13 bacterial 16S rRNA gene sequences, five of which branched together with well-known bacterial intracellular lineages (figure 2; electronic supplementary material, figure S5). These sequences were only found in the *Pompholyxophrys* assemblies, including two SCTs (*Pompholyxophrys* LG126 and LG127) and one SCG from *P. punicea* (20-cells WGA).

One of these bacterial sequences (*Pompholyxophrys* sp. LG126 (2)) branched within the Chlamydiae (figure 2*a*), along with sequences of known bacterial endosymbionts of the amoebae *Acanthamoeba* sp. and *Hartmannella vermiformis*. The other four sequences branched within the Rickettsiales (figure 2*b*). *Pompholyxophrys punicea* LG127 seemed to harbour two different *Rickettsia*-like endosymbionts. One of them, LG127 (1), branched within a clade of *Rickettsia* species endosymbionts of different hosts, including metazoans and, interestingly, *N. pattersoni* [70]. The second sequence LG127 (2) and a second sequence from *Pompholyxophrys* sp. LG126 (1) were identical. The last endosymbiont candidate sequence came from the *P. punicea* 20-cells WGA assembly and, although clearly branching within the Rickettsiales, had no close relatives. Thus, the same endosymbiont can be found in different cells from the same natural sample, as in the case of *Pompholyxophrys* sp. LG126 (1) and LG127 (2). Conversely, different endosymbionts can coexist in the same cell as well, as seen in *P. punicea* LG127 (1 and 2), in this case belonging to the same bacterial clade of Rickettsiales. A single cell can also harbour endosymbionts from phylogenetically distant groups as seen in *Pompholyxophrys* sp. LG126 (1 and 2), containing representatives of Chlamydiae and Rickettsiales (figure 2).

Our results are consistent with the findings of Dirren and Posch [28], showing that symbiont acquisition in nucleariids seems to be rather promiscuous. It is also worth noting that we only found endosymbiont sequences in the *Pompholyxophrys* assemblies. We could not recover any bacterial sequence from our *Nuclearia* assemblies, even though we have worked with the same *Nuclearia* species studied by Dirren and Posch [28]. However, because we only analyzed with *Nuclearia* transcriptome sequences, we cannot completely discard the presence of endosymbionts.

(c) Phylogenomic analyses

To establish a solid phylogenetic framework for nucleariids, and because the 18S rRNA gene has limited resolution power, we generated genome and transcriptome data for several nucleariids (table 1). Although the percentage of orthologue gene markers recovered for the two datasets was low (especially for *Pompholyxophrys* assemblies) (table 1), we could retrieve a sufficient number of gene marker sequences from our new assemblies for three *Nuclearia* species, two *Pompholyxophrys* species and *Lithocolla* (table 1). We also used publicly available data from two *Nuclearia* species [7], two *Fonticula* species [5,71] and *Parvularia atlantis* [20], adding representative members of other opisthokont lineages as outgroup. With these sequence datasets, we updated two datasets of conserved phylogenetic markers previously used to study the phylogeny of holomycotan clades [9,61]: the GBE dataset (264 proteins) and the SCPD dataset (74 single-copy protein domains—without *N. pattersoni* XT1 as no gene markers were retrieved for this species) (electronic supplementary material, figures S6A–D). As in the 18S rRNA gene phylogeny, all previously recognized nucleariids (*Nuclearia*, *Fonticula* and *Parvularia*) clustered together with *Lithocolla* and the two *Pompholyxophrys* species with maximum support in ML and BI analyses for both datasets, forming a sister clade to other Holomycota (figure 3). However, the relationships between the different genera were not the same as in the 18S rRNA gene tree, in particular regarding the placement of *Fonticula*. *Fonticula* appeared as a long branch sister clade to *Lithocolla* and *Pompholyxophrys* (with low statistical support) in the 18S rRNA gene tree (figure 1). However, in the phylogenomic analyses, the two *Fonticula* species clustered with *Parvularia* with high statistical support (figure 3). All the five *Nuclearia* species (with the same internal topology as in the 18S rRNA gene tree) clustered with *Lithocolla* and the two *Pompholyxophrys*. Thus, two separated clades formed, one containing all *Nuclearia* species and one containing the scale-bearing *Pompholyxophrys* and *Lithocolla*, both with maximum support values.

(d) Evolutionary implications

Our robust phylogenomic tree of nucleariids allows us to discuss the evolutionary history of several nucleariid characters, although molecular data are still missing for genera putatively related to nucleariids, such as *Vampyrellidium*, *Pinaciophora*, *Elaeorhanis* or *Rabdiophrys* (see the electronic supplementary material for detailed taxonomical discussion).

The last common ancestor of opisthokonts was probably phagotrophic with amoeboid polarized cell shape and a single flagellum, features that can be found in extant examples such as choanoflagellates [72], pigoraptors [73] or aphelids [9]. All known nucleariids lack flagella, suggesting that the last common ancestor of all nucleariids had already lost the ancestral flagellum. It is also worth mentioning that the nucleariid ancestor probably originated in freshwater environments, as suggested by the 18S rRNA gene tree analysis in which all the basal branches (including environmental clades) are occupied by freshwater lineages. The non-polarized and plastic cell shape surrounded by hyaline pseudopodia (branching filopodia) of nucleariids seems concomitant with the loss of flagella. Although there are few studies on nucleariid biology, cell movement by 'walking' on the benthos [29] and planktonic stages with equally radiating filopodia (electronic supplementary material, figures S1–S3) arise as major common features of nucleariids, together with a mucilaginous coat involved in different functions (from encystation to encapsulation of ectosymbionts or scales [1,14,29,67]. Although the current knowledge about this group is limited, we can already speculate about evolutionary patterns regarding cell size, food

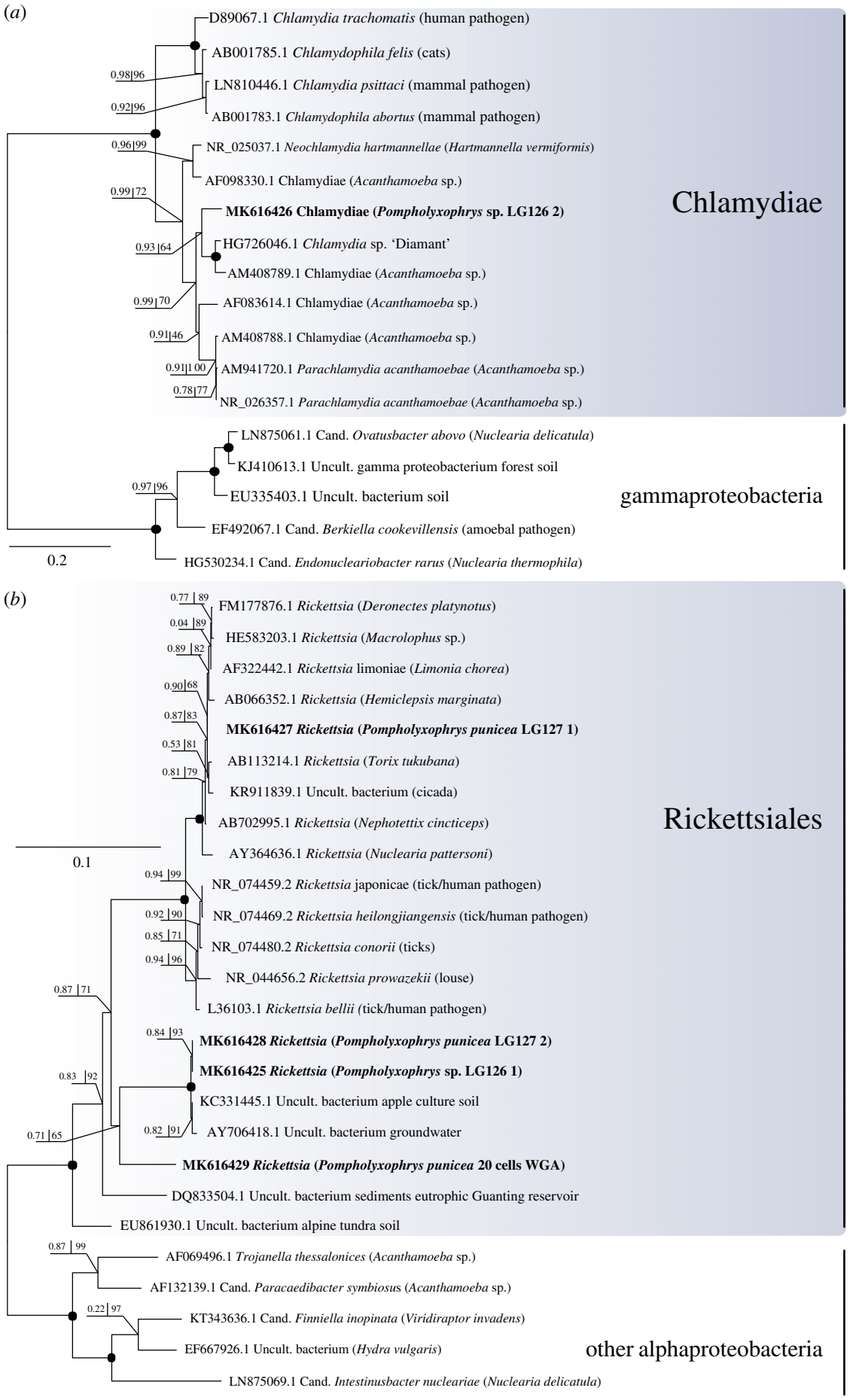

Figure 2. ML phylogenetic tree of 16S rRNA genes showing likely nucleariid bacterial endosymbionts (in bold). (*a*) Chlamydiae tree including one sequence from *Pompholyxophrys* sp. LG126 (2) and inferred under the TIM3 + F + I + G4 model using 1454 conserved nucleotide positions. (*b*) Rickettsiales tree including four sequences obtained in this study and inferred under the GTR + F + I + G4 model using 1390 conserved nucleotide positions. Statistical supports shown are Bayesian PP obtained under GTR + G + I on the left and ML UFBS on the right. Endosymbiont hosts are indicated in parenthesis.

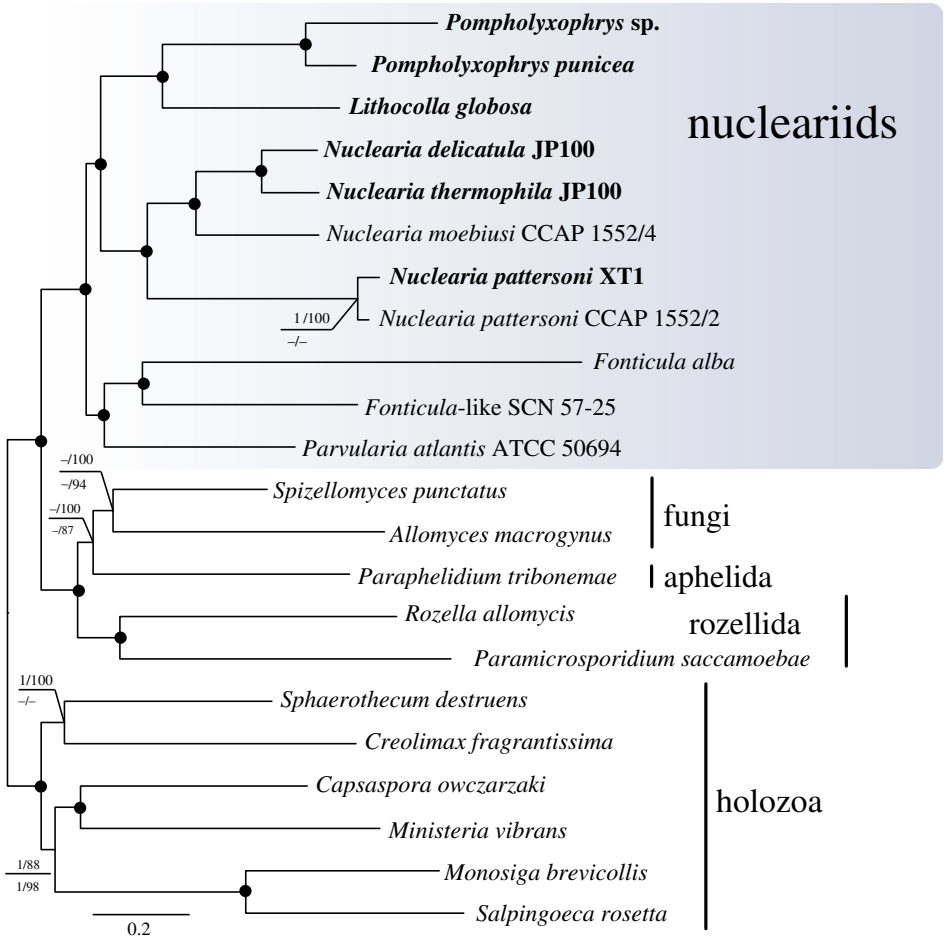

Figure 3. ML phylogenomic tree based on the GBE protein dataset. The tree was reconstructed using 264 conserved proteins, 22 species and 96 276 conserved amino acid positions with the LG + R5 + C60 model. Upper values correspond to supports obtained from the GBE dataset and lower values to those obtained from the single-copy protein domain (SCPD21; without *N. pattersoni* XT1) dataset. Bayesian PP under the CAT-Poisson model are shown on the left and ML UFBS supports are shown on the right. Branches with support values higher or equal to 0.99 PP and 95% UFBS are indicated by black dots. Species names in bold correspond to those for which we have obtained transcriptome and/or genome sequences in this study. (Online version in colour.)

source, ecological niche and cell-coverings (figure 4). From the last common nucleariid ancestor, two clades evolved, one characterized by smaller cells (*Parvularia–Fonticula*) and other with larger cells (*Nucleria* and scaled nucleariids). These differential cell sizes correlate with different ecological specializations in terms of prey and lifestyle. *Parvularia* and *Fonticula* are both exclusively bacterivorous and part of nanoplankton, the first never reaches more than 6 µm [20] and the latter no more than 12 µm [5,39] in size. *Fonticula alba*, which seems to evolve faster than other nucleariids (see branch lengths in figures 1 and 2), grows better in agar plates than in liquid medium (D. López-Escardó 2017, personal communication), and uses its mucilaginous coat to aggregate cells and form fruiting bodies [74]. Hence, *F. alba* looks more adapted to soil environments than to the water column preferred by other nucleariids. Although *Parvularia* and *Nucleria* share many common features (justifying the original identification of *Parvularia* as a nucleariid [20]) *Nucleria* cells are much bigger (from approximately 10 up to 60 µm, depending on the life stage and culture conditions [28]; electronic supplementary material, figure S3). *Lithocolla* (electronic supplementary material, figure S1) and *Pompholyxophrys* (electronic supplementary material, figure S2) range from 20 to 45 µm [15,23,37]. This microplanktonic (greater than 20 µm) size allows them to feed on filamentous cyanobacteria, algae or even other eukaryotes. Finally, *Fonticula*, *Parvularia* and *Nucleria* seem very plastic in terms of cell shape, being round, amorphous or extremely elongated. However, cells became less polymorphic in the genera that acquired the capacity to cover themselves either with xenosomes (probably as a by-product of phagocytosis), as in *Lithocolla* and maybe *Elaeorhanis* [27] (electronic supplementary material, figure S1, [41]), or with idiosomes, as in *Pompholyxophrys* (electronic supplementary material, figure S2) and maybe *Pinaciophora* [25].

Despite these evolutionary implications, deciphering the evolutionary history of nucleariids will require additional data. Indeed, although nucleariids are a pivotal group at the onset of the Holomycota divergence, they remain an under-sampled group, as suggested by environmental data and the many described and likely related species that still lack molecular data. As most nucleariids lack cultured representatives in the laboratory, single-cell techniques will be an invaluable tool to expand the known diversity of uncultured nucleariids, helping to reconcile genomic information with morphology and ecological features, including the presence and role of ecto- and endo-symbiotic bacteria.

(e) Culturing versus single-cell genomes/transcriptomes

In this study, we have used a combination of single-cell techniques (including steps of whole genome/transcriptome amplification) and whole RNA extraction from cultured

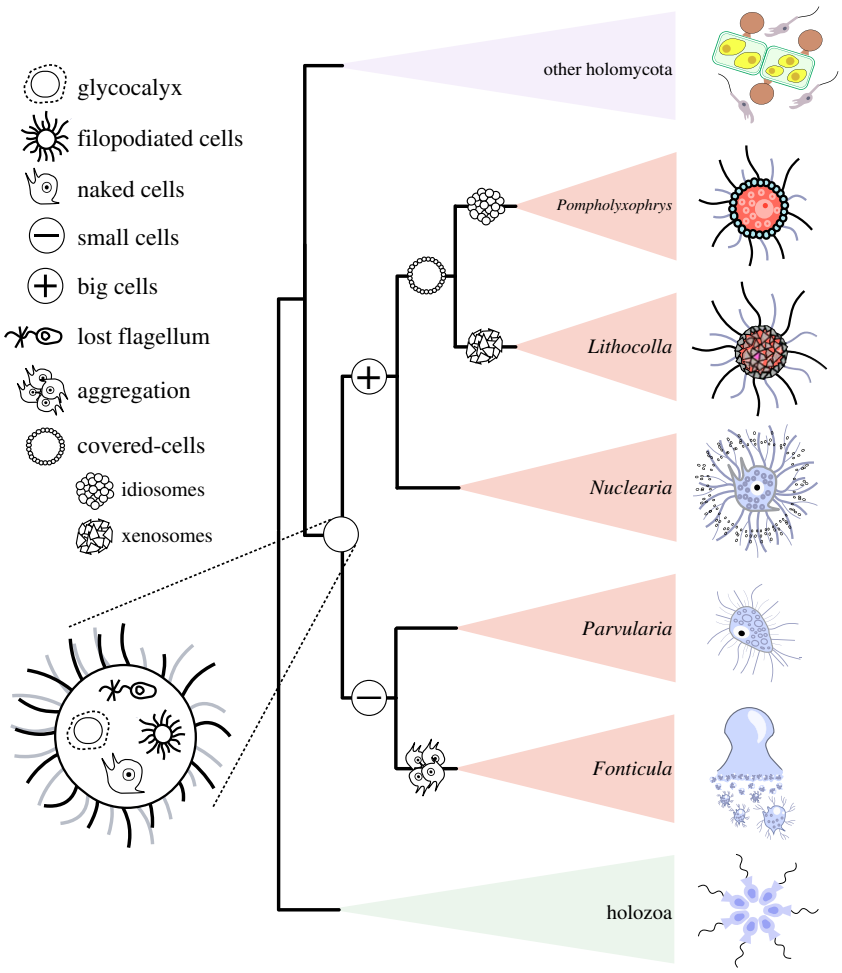

Figure 4. Schematic opisthokont phylogeny displaying cellular characteristics of nucleariids (cell size, presence/absence of cell-cover and its nature, lack of flagellated stages, filopodia, and the presence of a glycocalyx and the capacity to aggregate) and their probable ancestral states in some nodes. (Online version in colour.)

material (without amplification steps) to sequence genomic material from several nucleariid species. Our SCGs/SCTs obtained after WGA/WTA steps produced different results when blasted against our biggest and most complete multigene dataset (GBE). In the case of *Lithocolla*, we obtained two single-cell assemblies, one from a few-cells genome amplification (SCG; LG140,144,145) and one from an SCT (SCT; LG147), recovering 30.68% and 13.25% of the GBE dataset proteins, respectively. The SCTs outperformed the SCGs in *Lithocolla*. In comparison, we recovered 75.37% of the proteins when performing traditional whole RNA extraction and sequencing from a culture.

In the case of the two *Pompholyxophrys* species, we only could obtain single/few-cell genomes/transcriptomes, because no cultures were available. Our *Pompholyxophrys* assemblies displayed different protein recovery percentages ranging from 30 to 47% for the SCTs (LG130 and LG129) and 0 to 13.63% for the SCGs and few-cells genome (LG127, LG126 and 20cellsWGA). Again, the SCTs seemed to outperform the SCGs in terms of protein recovery in this particular case.

Both SCGs/SCTs proved to be useful to obtain enough data to place *Lithocolla* and *Pompholyxophrys* in our multigene phylogeny with strong support. It also allowed us to unveil the hidden diversity in the group, because what initially we thought to be a single *Pompholyxophrys* species were actually two different species (*Pompholyxophrys* sp. and *P. punicea*) as revealed by both 18S rRNA gene and multigene trees.

Nevertheless, not surprisingly, the best results were obtained after RNA extraction of cultures, e.g. *Lithocolla* (75.37%), a result that we confirmed for *N. delicatula* JP100 and *N. thermophila* JP100, for which we recover 88.63% and 95.07% of the dataset proteins, respectively. Culturing approaches, if achievable, remain the best strategies to produce a high amount of high-quality data. However, most protist species are not easily amenable to culture. Therefore, single-cell 'omics', although still far from allowing high or even levels of completeness often allow, as in this particular study, retrieving enough conserved markers to run robust phylogenomic analyses. Further progress in single-cell approaches leading to the retrieval of higher and more homogeneous coverages will hopefully allow more in-depth comparative genomics and population genomics of protists directly sampled from natural communities.

Data accessibility. 18S and rRNA gene sequences have been deposited in GenBank with accession nos. MK547173–MK547179 and MK616425–MK616429, respectively. Transcriptome and genome sequence data have been submitted to NCBI SRA under the Bioproject PRJNA517920.

Authors' contributions. L.J.G., G.T., D.M. and P.L.-G. conceived, coordinated the study and wrote the manuscript. G.T. micromanipulated and obtained *Nuclearia*, *Lithocolla* and *Navicula* RNA. L.J.G. micromanipulated and obtained *Lithocolla* and *Pompholyxophrys* DNA and RNA. D.M. micromanipulated and amplified *N. pattersoni* RNA. S.C. collected freshwater samples and micromanipulated *Pompholyxophrys* cells. Y.E. isolated, cultured and characterized *Lithocolla*. E.V. identified and obtained images from *Pompholyxophrys*. G.T., L.J.G.

and D.M. reconstructed 18S and 16S rRNA gene phylogenies. G.T. and L.J.G. assembled genome and transcriptome sequences, cleaned the assemblies, performed phylogenomic analyses and contributed equally to this work. All authors gave final approval for publication.

Competing interests. We have no competing interests to declare.

Funding. This work was funded by the European Research Council Advanced Grant 'ProtistWorld' (grant no. 322669) and the Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN project SINGEK (http://www.singek.eu/; grant agreement no. H2020-MSCA-ITN-2015-675752). G.T. was financed by the European Marie Sklodowska-Curie Action (704566 AlgDates).

Acknowledgements. We thank David López-Escardó for help with *Nuclearia* molecular identification and advice on duplex PCR, John O'Brien for sample collection for *Lithocolla*, Hélène Timpano and the UNICELL platform for help with single-cell methods, Giselle Walker for help with an unpublished taxonomy of eukaryotes, and the website repositories Microworld, Protist information server, The World of Protozoa and Biodiversity Heritage Library for access to valuable information.

References

1. Patterson DJ. 1984 The genus *Nuclearia* (Sarcodina, Filosea): species composition and characteristics of the taxa. *Arch. Mikrobiol.* **128**, 127–139. (doi:10.1016/s0003-9365(84)80034-2)
2. Zettler L, Nerad TA, O'Kelly CJ, Sogin ML. 2001 The nucleariid amoebae: more protists at the animal-fungal boundary. *J. Eukaryot. Microbiol.* **48**, 293–297. (doi:10.1111/j.1550-7408.2001.tb00317.x)
3. Medina M, Collins AG, Taylor JW, Valentine JW, Lipps JH, Amaral-Zettler L, Sogin ML. 2003 Phylogeny of Opisthokonta and the evolution of multicellularity and complexity in Fungi and Metazoa. *Int. J. Astrobiol.* **2**, 203–211. (doi:10.1017/S1473550403001551)
4. Steenkamp ET, Wright J, Baldauf SL. 2006 The protistan origins of animals and fungi. *Mol. Biol. Evol.* **23**, 93–106. (doi:10.1093/molbev/msj011)
5. Brown MW, Spiegel FW, Silberman JD. 2009 Phylogeny of the 'forgotten' cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. *Mol. Biol. Evol.* **26**, 2699–2709. (doi:10.1093/molbev/msp185)
6. Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T. 2008 Multigene phylogeny of choanozoa and the origin of animals. *PLoS ONE* **3**, e2098. (doi:10.1371/journal.pone.0002098)
7. Liu Y, Steenkamp ET, Brinkmann H, Forget L, Philippe H, Lang BF. 2009 Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. *BMC Evol. Biol.* **9**, 1–11. (doi:10.1186/1471-2148-9-272)
8. Adl SM *et al.* 2018 Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119. (doi:10.1111/jeu.12691)
9. Torruella G, Grau-Bové X, Moreira D, Karpov SA, Burns JA, Sebé-Pedrós A, Völcker E, López-García P. 2018 Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. *Commun. Biol.* **1**, 231. (doi:10.1038/s42003-018-0235-z)
10. James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, Stajich JE. 2013 Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Curr. Biol.* **23**, 1548–1553. (doi:10.1016/j.cub.2013.06.057)
11. Karpov SA, Mamkaeva MA, Aleoshin VV, Nassonova E, Lilje O, Gleason FH. 2014 Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front. Microbiol.* **5**, 1–11. (doi:10.3389/fmicb.2014.00112)
12. Nagy LG, Kovács GM, Krizsán K. 2018 Complex multicellularity in fungi: evolutionary convergence, single origin, or both? *Biol. Rev.* **93**, 1778–1794. (doi:10.1111/brv.12418)
13. Magallón S *et al.* 2018 Contemporaneous radiations of fungi and plants linked to symbiosis. *Nat. Commun.* **9**, 1–11. (doi:10.1038/s41467-018-07849-9)
14. Patterson DJ. 1983 On the organization of the naked filose amoeba, *Nuclearia moebiusi* Frenzel, 1897 (Sarcodina, Filosea) and its implications 1. *J. Protozool.* **30**, 301–307. (doi:10.1111/j.1550-7408.1983.tb02920.x)
15. Patterson DJ. 1985 On the organization and affinities of the amoeba, *Pompholyxophrys punicea* Archer, based on ultrastructural examination of individual cells from wild material. *J. Eukaryot. Microbiol.* **32**, 241–246. (doi:10.1111/j.1550-7408.1985.tb03044.x)
16. Patterson DJ, Surek B, Melkonian M. 1987 The ultrastructure of *Vampyrellidium perforans* Surek & Melkonian and its taxonomic position among the naked filose Amoebae1. *J. Protozool.* **34**, 63–67. (doi:10.1111/j.1550-7408.1987.tb03133.x)
17. Dyková I, Veverková M, Fiala I, Macháčková B, Pecková H. 2003 *Nuclearia pattersoni* sp. n. (Filosea), a new species of amphizoic amoeba isolated from gills of roach (*Rutilus rutilus*), and its rickettsial endosymbiont. *Folia Parasitol.* (*Praha*). **50**, 161–170. (doi:10.14411/fp.2003.030)
18. Yoshida M, Nakayama T, Inouye I. 2009 *Nuclearia thermophila* sp. nov. (Nucleariidae), a new nucleariid species isolated from Yunoko Lake in Nikko (Japan). *Eur. J. Protistol.* **45**, 147–155. (doi:10.1016/j.ejop.2008.09.004)
19. del Campo J, Mallo D, Massana R, de Vargas C, Richards TA, Ruiz-Trillo I. 2015 Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. *Environ. Microbiol.* **17**, 3195–3207. (doi:10.1111/1462-2920.12759)
20. López-Escardó D, López-García P, Moreira D, Ruiz-Trillo I, Torruella G. 2017 *Parvularia atlantis* Gen. et Sp. Nov., a nucleariid filose amoeba (Holomycota, Opisthokonta). *J. Eukaryot. Microbiol.* **38**, 42–49. (doi:10.1111/jeu.12450)
21. Thomsen HA. 1978 On the identity of the Heliozoan *Pinaciophora fluviatilis* and *Potamodiscus kalbei*; with the description of eight new *Pinaciophora* species. *Protistologica* **14**, 359–373. (doi:10.1016/s0003-9365(78)80027-x)
22. Mikrjukov KA. 1999 Taxonomic revision of scale-bearing Heliozoon-like amoebae (Pompholyxophryidae, Rotosphaerida). *Acta Potozool.* **38**, 119–131.
23. Nicholls KH, Durrschmidt M. 1985 Scale structure and taxonomy of some species of *Raphidocystis*, *Raphidiophrys*, and *Pompholyxophrys* (Heliozoea) including descriptions of six new taxa. *Can. J. Zool.* **63**, 1944–1961. (doi:10.1139/z85-288)
24. Roijackers RMM, Siemensma FJ. 1988 A study of cristidiscoidid amoebae (Rhizopoda, Filosea), with descriptions of new species and keys to genera and species. *Arch. fur Protistenkd.* **135**, 237–253. (doi:10.1016/S0003-9365(88)80072-1)
25. Nicholls KH. 2012 New and little-known marine species of *Pinaciophora*, *Rabdiaster* and *Thomseniophora* gen. nov. (Rotosphaerida: Pompholyxophryidae). *J. Mar. Biol. Assoc. United Kingdom* **93**, 1211–1229. (doi:10.1017/S002531541200135X)
26. Cavalier-Smith T. 1993 Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994.
27. Patterson DJ, Simpson AGB, Rogerson A. 2000 Amoebae of uncertain affinities. In *An illustrated guide to the protozoa, 2nd edn*, vol. II (eds JJ Lee, GF Leedale, P Bradbury), pp. 804–827. Lawrence, KS: Society of Protozoologists.
28. Dirren S, Posch T. 2016 Promiscuous and specific bacterial symbiont acquisition in the amoeboid genus *Nuclearia* (Opisthokonta). *FEMS Microbiol. Ecol.* **92**, 1–16. (doi:10.1093/femsec/fiw105)
29. Dirren S, Pitsch G, Silva MOD, Posch T. 2017 Grazing of *Nuclearia thermophila* and *Nuclearia delicatula* (Nucleariidae, Opisthokonta) on the toxic cyanobacterium *Planktothrix rubescens*. *Eur. J. Protistol.* **60**, 87–101. (doi:10.1016/j.ejop.2017.05.009)
30. Cienkowski L. 1865 Beiträge zur Kenntnis der Monaden. *Arch. für mikroskopische Anat.* **1**, 203–232. (doi:10.1007/BF02961414)
31. Cann JP, Page FC. 1979 *Nucleosphaerium tuckery* nov. ge. nov. sp—a new freshwater filose amoeba without motile form in a new family Nucleariidae (Filosea: Aconchulinida) feeding by ingestion only. *Arch. für Protistenkd.* **122**, 226–240. (doi:10.1016/S0003-9365(79)80034-2)

32. Surek B, Melkonian M. 1980 The filose amoeba *Vampyrellidium perforans* nov. sp. (Vampyrellidae, Aconchulinida): axenic culture, feeding behaviour and host range specifity. *Arch. für Protistenkd.* **123**, 166–191. (doi:10.1016/S0003-9365(80)80003-0)
33. Hess S, Sausen N, Melkonian M. 2012 Shedding light on vampires: the phylogeny of vampyrellid amoebae revisited. *PLoS ONE* **7**, e31165. (doi:10.1371/journal.pone.0031165)
34. Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ. 2016 Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B Biol. Sci.* **283**, 20152802. (doi:10.1098/rspb.2015.2802)
35. Page FC. 1987 The classification of 'Naked' Amoebae (Phylum Rhizopoda). *Arch. für Protistenkd.* **133**, 199–217. (doi:10.1016/S0003-9365(87)80053-2)
36. Schulze FE. 1874 Rhizopodenstudien II. *Arch. für mikroskopische Anat.* **10**, 377–400. (doi:10.1007/BF02960329)
37. Penard E. 1904 *Les héliozoaires d'eau douce*. Genève: H. Kündig. See https://www.biodiversitylibrary.org/item/16282.
38. Rainer H. 1968 Urtiere, Protozoa—Wurzelfuessler, Rhizopoda—Sonnentierchen, Heliozoa–Systematik und Taxonomie, Biologie, Verbreitung und Ökologie der Arten der Erde. In *Die Tierwelt Deutschlands und der angrenzenden Meeresteile nach ihren Merkmalen und nach ihrer Lebensweise* (eds M Dahl, F Peus), p. 56. Jena, Germany: Gustav Fischer Verlag.
39. Worley AC, Raper KB, Hohl M. 1979 *Fonticula alba*: a new cellular slime mold (Acrasiomycetes). *Mycologia* **71**, 746–760. (doi:10.1080/00275514.1979.12021068)
40. Torruella G *et al*. 2015 Phylogenomics reveals convergent evolution of lifestyles in close relatives of animals and fungi. *Curr. Biol.* **25**, 2404–2410. (doi:10.1016/j.cub.2015.07.053)
41. Cann JP. 1986 The feeding behavior and structure of *Nuclearia delicatula*. *J. Protozool.* **33**, 392–396. (doi:10.1111/j.1550-7408.1986.tb05629.x)
42. Andrews S. 2010 FastQC: a quality control tool for high throughput sequence data. See https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
43. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
44. Nurk S *et al*. 2013 Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **20**, 714–737. (doi:10.1089/cmb.2013.0084)
45. Laetsch DR, Blaxter ML, Leggett RM. 2017 BlobTools: interrogation of genome assemblies. *F1000Research* **6**, 1–16. (doi:10.12688/f1000research.12232.1)
46. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009 BLAST plus: architecture and applications. *BMC Bioinf.* **10**, 1–9. (doi:10.1186/1471-2105-10-421)
47. Li W, Godzik A. 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659. (doi:10.1093/bioinformatics/btl158)
48. Huerta-Cepas J *et al*. 2016 EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293. (doi:10.1093/nar/gkv1248)
49. Galindo LG, Torruella G, Eglit Y, Simpson AGB, Völcker E, Clauß S, Moreira D, López-García P. 2019 Data from combined cultivation and single-cell approaches to the phylogenomics of nucleariid amoebae, close relatives of Fungi. *Figshare* (doi:10.6084/m9.figshare.7814411.v1)
50. Heger TJ, Giesbrecht IJW, Gustavsen J, del Campo J, Kellogg CTE, Hoffman KM, Lertzman K, Mohn WW, Keeling PJ. 2018 High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. *Environ. Microbiol.* **20**, 1185–1203. (doi:10.1111/1462-2920.14061)
51. Arroyo AS, López-Escardó D, Kim E, Ruiz-Trillo I, Najle SR. 2018 Novel diversity of deeply branching holomycota and unicellular holozoans revealed by metabarcoding in Middle Paraná River, Argentina. *Front. Ecol. Evol.* **6**, 1–17. (doi:10.3389/fevo.2018.00099)
52. Katoh K, Misawa K, Kuma K, Miyata T. 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066. (doi:10.1093/nar/gkf436)
53. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. (doi:10.1093/bioinformatics/btp348)
54. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
55. Guindon S, Gascuel O. 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704. (doi:10.1080/10635150390235520)
56. Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O, Nisimova MAA, II MAG, Ranc JEAN. 2011 Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* **60**, 685–699. (doi:10.1093/sysbio/syr041)
57. Felsenstein J. 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N. Y).* **39**, 783–791. (doi:10.2307/2408678)
58. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589. (doi:10.1038/nmeth.4285)
59. Posada D. 2008 jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256. (doi:10.1093/molbev/msn083)
60. Ronquist F, Huelsenbeck JP. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)
61. Mikhailov KV, Simdyanov TG, Aleoshin VV. 2016 Genomic survey of a hyperparasitic microsporidian *Amphiamblys* sp. (Metchnikovellidae). *Genome Biol. Evol.* **9**, 454–467. (doi:10.1093/gbe/evw235)
62. Price MN, Dehal PS, Arkin AP. 2009 Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650. (doi:10.1093/molbev/msp077)
63. Leigh JWJW, Roger AJ. 2008 Barrel-o-Monkeys. See http://rogerlab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm.
64. Lartillot N, Lepage T, Blanquart S. 2009 PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288. (doi:10.1093/bioinformatics/btp368)
65. Minh BQ, Nguyen MAT, Von Haeseler A. 2013 Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195. (doi:10.1093/molbev/mst024)
66. Rambaut A. 2009 FigTree. See http://tree.bio.ed.ac.uk/software/figtree/.
67. Dirren S, Salcher MM, Blom JF, Schweikert M, Posch T. 2014 Ménage-à-trois: the Amoeba *Nuclearia* sp. from Lake Zurich with its ecto- and endosymbiotic bacteria. *Protist* **165**, 745–758. (doi:10.1016/j.protis.2014.08.004)
68. Leonov MM. 2012 Heliozoans (Heliozoa, Sarcodina, Protista) of fresh and marine waters of the European part of Russia: species composition, morphology, and distribution. *Inl. Water Biol.* **3**, 344–355. (doi:10.1134/s1995082910040073)
69. Wujek DE. 2015 Freshwater silica-scaled heterotrophic protista: heliozoa, thaumatomonad flagellates, amoebae, and bicosoecids, from the Lake Itasca Region, Minnesota. *J. Minnesota Acad. Sci.* **78**, 1–14. (doi:10.1007/s10773-016-2998-7)
70. Feehan CJ, Johnson-Mackinnon J, Scheibling RE, Lauzon-Guay JS, Simpson AGB. 2013 Validating the identity of *Paramoeba invadens*, the causative agent of recurrent mass mortality of sea urchins in Nova Scotia, Canada. *Dis. Aquat. Organ.* **103**, 209–227. (doi:10.3354/dao02577)
71. Kantor RS, van Zyl AW, van Hille RP, Thomas BC, Harrison STL, Banfield JF. 2015 Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics. *Environ. Microbiol.* **17**, 4929–4941. (doi:10.1111/1462-2920.12936)
72. Hoffmeyer TT, Burkhardt P. 2016 Choanoflagellate models—*Monosiga brevicollis* and *Salpingoeca rosetta*. *Curr. Opin. Genet. Dev.* **39**, 42–47. (doi:10.1016/j.gde.2016.05.016)
73. Hehenberger E, Tikhonenkov DV, Kolisko M, del Campo J, Esaulov AS, Mylnikov AP, Keeling PJ. 2017 Novel predators reshape holozoan phylogeny and reveal the presence of a two-component signaling system in the ancestor of animals. *Curr. Biol.* **27**, 2043–2050. (doi:10.1016/j.cub.2017.06.006)
74. Deasey MC, Olive LS. 1981 Role of golgi apparatus in sorogenesis by the cellular slime mold *Fonticula alba*. *Science* **213**, 561–563. (doi:10.1126/science.213.4507.561)