Contents lists available at ScienceDirect



Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



Phylogenomics supports the monophyly of the Cercozoa

Nicholas A.T. Irwin^{a,*}, Denis V. Tikhonenkov^{a,b}, Elisabeth Hehenberger^{a,1}, Alexander P. Mylnikov^b, Fabien Burki^{a,2}, Patrick J. Keeling^a

^a Department of Botany, University of British Columbia, Vancouver V6T 124, British Columbia, Canada
^b Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok 152742, Russia

ARTICLE INFO

ABSTRACT

Keywords: Cercozoa Rhizaria Phylogeny Phylogenomics Single-cell transcriptomics The phylum Cercozoa consists of a diverse assemblage of amoeboid and flagellated protists that forms a major component of the supergroup, Rhizaria. However, despite its size and ubiquity, the phylogeny of the Cercozoa remains unclear as morphological variability between cercozoan species and ambiguity in molecular analyses, including phylogenomic approaches, have produced ambiguous results and raised doubts about the monophyly of the group. Here we sought to resolve these ambiguities using a 161-gene phylogenetic dataset with data from newly available genomes and deeply sequenced transcriptomes, including three new transcriptomes from *Auriganonas solis, Abollifer prolabens*, and a novel species, *Lapto gusevi* n. gen. n. sp. Our phylogenomic analysis strongly supported a monophyletic Cercozoa, and approximately-unbiased tests rejected the paraphyletic topologies observed in previous studies. The transcriptome of *L. gusevi* represents the first transcriptomic data from the large and recently characterized Aquavolonidae-Treumulida-Novel Clade 12' group, and phylogenomics supported its position as sister to the cercozoan subphylum, Endomyxa. These results provide insights into the phylogeny of the Cercozoa and the Rhizaria as a whole.

1. Introduction

The Cercozoa comprise a diverse protozoan phylum composed of ecologically and morphologically diverse organisms. The majority of cercozoans are heterotrophic amoeboflagellates that are abundant in soil and aquatic environments. However, a number of lineages exhibit unique trophic modes, relying on phototrophy (e.g., the chlorarachniophytes and Paulinella chromatophora) or intracellular parasitism (e.g., Plasmodiophora brassicae). This diversity is compounded by morphological variation as cercozoans can also adopt, and sometimes interchange between, zooflagellate and amoeboid morphotypes. As a result of this variability, the phylum Cercozoa and its various subphyla, such as the Reticulofilosa, Monadofilosa, and Endomyxa, only became recognized through molecular phylogenetic analysis. Initial indications of these relationships came from phylogenies based on ribosomal RNA (rRNA) as well as alpha- and beta- tubulins (Bhattacharya et al., 1995; Cavalier-Smith, 1998; Cavalier-Smith and Chao, 2003a; Keeling et al., 1998; Moreira et al., 2007; Wylezich et al., 2002). Further sampling of actin, tubulin, RNA polymerase II, and polyubiquitin, as well as phylogenomic approaches, revealed that the Cercozoa are closely related to foraminiferans and radiolarians, and together these groups came to comprise the supergroup Rhizaria (Archibald et al., 2003; Bass et al., 2005; Burki et al., 2010; Keeling, 2001; Longet et al., 2003; Takishita et al., 2005). But although these analyses provided insights into the higher order taxonomy of the Cercozoa, they also raised doubts about the group's monophyly.

Phylogenomic analyses have consistently recovered a paraphyletic Cercozoa, with the Endomyxa often clustering with the Foraminifera or the Retaria (Foraminifera and Radiolaria) (Brown et al., 2012; Burki et al., 2013, 2010; Cavalier-Smith et al., 2018; Krabberød et al., 2017; Sierra et al., 2015, 2013). However, poor taxon sampling and datasets with substantial amounts of missing data have reduced support for these phylogenies, leaving it unresolved whether the Endomyxa are part of the Cercozoa or not. A number of molecular synapomorphies have been identified and used in an attempt to consolidate the phylogenetics, but these do not address this particular issue. For example, polyubiquitin insertions were originally used to support the clustering of the Cercozoa and Retaria, but this feature appears to be ancestral to Rhizaria and fails to support the partitioning of its phyla (Archibald et al., 2003; Bass et al., 2005; Burki et al., 2010). In contrast, insertions

https://doi.org/10.1016/j.ympev.2018.09.004 Received 24 July 2018; Accepted 6 September 2018 Available online 11 October 2018 1055-7903/ © 2018 Elsevier Inc. All rights reserved.

^{*} Corresponding author.

E-mail address: nickatirwin@gmail.com (N.A.T. Irwin).

¹ Current address: Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA.

² Current address: Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

in the mitochondrial ribosomal protein L1 appear to be common to the Cercozoa to the exclusion of the Endomyxa, whereas a single nucleotide deletion in the small subunit (SSU) rRNA support its monophyly (Burki et al., 2010; Cavalier-Smith et al., 2018; Cavalier-Smith and Chao, 2003b). Despite the ambiguity of these data, a recent reclassification of Rhizaria formally removed the Endomyxa from the Cercozoa, placing it within the Retaria, with the primary justification for the amendment being phylogenomics (Cavalier-Smith, 2018).

To more thoroughly test the monophyly of the Cercozoa, we sought to increase the phylogenomic sampling of cercozoan taxa, in particular the Endomyxa and recently recognized relatives of the Endomyxa (Bass et al., 2018), and assess the effect of better sampling on phylogenomic analyses. Overall, these analyses support the monophyly of the Cercozoa, and reject the alternative hypotheses where they are paraphyletic.

2. Materials and methods

2.1. Collection, identification, and transcriptomics of cercozoan species

Clone Lap-1 was obtained from the sediment of a shallow pool near a forest road in Cát Tiên National Park (107°23'184'E, 11°26'752"N), Dong Nai Province, S.R. Vietnam on May 13, 2012. The sample was collected at 20 cm depth (Temp. 34.9 °C, pH 6.7, Eh 148 mV, Conductivity 19 µS/cm, DO 5.6 mg/L) and contained organic detritus and plant debris. Field studies in Vietnam were conducted under permits issued by the administration of Cát Tiên National Park, Vietnam, and authorized by the Vietnam-Russian Tropical Centre, Coastal Branch (Nha Trang, Vietnam). Clone AF-21 (identified as Aurigamonas solis, Vickerman 2005) was isolated from a sample of water and plant debris collected from a fountain in Victoria Garden (16°31'23.1"W, 28°23'21.1"N), La Otorava, Tenerife, Canary Islands on October 20, 2014. The samples were examined on the third, sixth and ninth day of incubation in accordance with methods described previously (Tikhonenkov et al., 2008). Following isolation, Lap-1 and AF-21 were propagated on the bodonid, Parabodo caudatus strain BAS-1, which was grown in Pratt's medium using the bacterium Pseudomonas fluorescens as food (Tikhonenkov et al., 2014). AF-21 is currently being stored in a collection of live protozoan cultures at the Institute for Biology of Inland Waters, Russian Academy of Sciences, however, Lap-1 perished after six months of cultivation.

Lap-1 and AF-21 were identified using a combination of microscopic and molecular approaches. Light microscopy observations of Lap-1 and AF-21 (Fig. S1A-E) were made using a Zeiss AxioScope A.1 equipped with a DIC water immersion objective (63x) and an AVT HORN MC-1009/S analog video camera. The SSU rRNA gene of Lap-1 (GenBank accession number: MG818165) and AF-21 (GenBank accession number: MH886395) was amplified by polymerase chain reaction (PCR) using the general eukaryotic primers PF1 (5'-GCGCTACCTGGTTGATCCT GCC-3') and FAD4 (5'-TGATCCTTCTGCAGGTTCACCTAC-3') and EukA (5'-AACCTGGTTGATCCTGCCAGT-3') and EukB (5'-GATCCTTCTGCA GGTTCACCTAC-3'), respectively (Keeling, 2002; Medlin et al., 1988; Tikhonenkov et al., 2014). PCR products were subsequently cloned (Lap-1) or sequenced directly (AF-21) using Sanger dideoxy sequencing.

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.09.004.

For cDNA preparation, cells grown in clonal laboratory cultures were harvested when the cells had reached peak abundance and after the majority of the prey had been eaten. Cells were collected by centrifugation (2000g, room temperature) onto the 0.8 µm membrane of a Vivaclear mini column (Sartorium Stedim Biotech Gmng, Cat. No. VK01P042). Total RNA was then extracted using a RNAqueous-Micro Kit (Invitrogen, Cat. No. AM1931) and converted into cDNA using a SMARTer Pico PCR cDNA Synthesis Kit (Clonetech) for Lap-1 and the Smart-Seq2 protocol (Picelli et al., 2014) for AF-21.

A single, gliding, wide-oval, flattened protist of approximately

20 µm with a long posterior trailing flagellum was isolated from a sea water sample taken on June 13, 2017 in the Strait of Georgia, British Columbia, Canada (123° 28′50″W, 49°10′366″N) at 220 m depth using a Niskin bottle. The cell was roughly identified as Abollifer prolabens (Fig. S1F-J) following light microscopy observations using a Leica DMIL LED microscope equipped with a DIC objective (63x) and a Sony α 6000 camera. A single A. prolabens cell was picked from the sample using a glass micropipette and transferred to a 0.2 mL thin-walled PCR tube containing 2 µL of cell lysis buffer (0.2% Triton X-100 and RNase inhibitor (Invitrogen)). cDNA was generated from the single cell using the Smart-Seq2 protocol (Picelli et al., 2014). One microlitre of non PCRpre-amplified cDNA (collected after reverse transcription) was used as a template for SSU rRNA gene amplification using the eukarvotic primers 18SFU (5'-ATGCTTGTCTCAAAGGRYTAAGCCATGC-3') and 18SRU (5'-CWGGTTCACCWACGGAAACCTTGTTACG-3') (Tikhonenkov et al., 2016). The PCR product was then sequenced by Sanger dideoxy sequencing and deposited in GenBank (accession number: MH886394). The newly generated SSU rRNA gene sequences were then used to further assess the identity of the newly collected taxa by using BLASTn to look for similar sequences in the non-redundant NCBI database (Altschul et al., 1990) (Table S1).

Sequencing libraries were prepared using either TruSeq (Lap-1) or NexteraXT (AF-21, *A. prolabens*) protocols and sequencing was performed on an Illumina MiSeq using either 250 bp (Lap-1) or 300 bp (AF-21, *A. prolabens*) paired end reads. Raw reads are available in the NCBI Short Read Archive (SRA) (Bioproject number: PRJNA490214, SRR7816690–SRR7816692).

2.2. Transcriptome assembly

Raw Illumina sequencing reads from *A. solis* (AF-21), *A. prolabens*, Lap-1, and *P. chromatophora* (downloaded from the NCBI SRA, SRR3221671) were merged using PEAR v0.9.6 and the quality of the paired reads was confirmed in FastQC (Andrews, 2010; Zhang et al., 2014). Adapter and primer sequences were subsequently trimmed using Trimmomatic v0.36 and transcriptomes were assembled using Trinity v2.4.0 (Bolger et al., 2014; Grabherr et al., 2011). The resulting contigs were then filtered for bacterial and kinetoplastid contaminants using BlobTools as well as blastn and blastx searches against the NCBI nt database and the Swiss-Prot database, respectively (Altschul et al., 1990; Laetsch and Blaxter, 2017). Coding sequences were predicted using a combination of TransDecoder v3.0.1 as well as a similarity search against the Swiss-Prot database (Haas et al., 2013).

2.3. Ortholog identification and concatenation

Along with the aforementioned taxa, the following transcriptomes and genomes were screened for orthologs and integrated into a previously developed phylogenomic dataset; *Bigelowiella longifila* (MMETSP1359), *Partenskyella glossopodia* (MMETSP1318), *Amorphochlora amoebiformis* (MMETSP0042), *Plasmodiophora brassicae* (PRJEB8376), and *Spongospora subterranea* (PRJEB9159) (Keeling et al., 2014; Schwelm et al., 2015).

Each dataset was then searched for a set of 263 genes that have been used in previous phylogenomic analyses (Burki et al., 2016, 2013, 2012; Hehenberger et al., 2017). Briefly, all of the sequences in the 263 gene-set, representing a wide range of eukaryotes, were used as queries to search the above datasets using blastp (Altschul et al., 1990). The hits were then filtered using an e-value threshold of $1e^{-20}$ and a query coverage of 50%. For each gene, a maximum of four non-redundant hits from each dataset was added to the original gene-set. Each of the gene-sets was then aligned using MAFFT L-INS-i v7.222 and trimmed using trimAl v1.2 with a gap-threshold of 80% (Capella-Gutiérrez et al., 2009; Katoh and Standley, 2013). In order to identify paralogs and contaminants, single gene trees were constructed for each gene using IQ-Tree v1.5.4 and the LG+G4 model with support from 1000 ultrafast

bootstraps (Hoang et al., 2018; Nguyen et al., 2015). The resultant trees were manually examined in FigTree v1.4.2 and contaminants and or paralogous sequences were flagged and removed (Rambaut, 2012). The final cleaned gene-sets were filtered so that they contained a maximum of 40% missing positions and then concatenated in SCaFoS v1.2.5 (Roure et al., 2007). The resulting concatenated alignment consisted of 161 genes spanning 44,786 amino acid positions from 40 taxa.

2.4. Phylogenomic tree building and topology tests

Phylogenomic analyses were performed using the concatenated alignment. The maximum likelihood (ML) trees were constructed using IQ-Tree v1.5.4 with statistical support assessed using both 1000 ultrafast and 100 non-parametric bootstrap pseudoreplicates (Hoang et al., 2018; Nguyen et al., 2015). Both an empirical profile mixture model (LG + C40 + F + G4) as well as a free rate heterogeneity model (LG + F + R6, selected using ModelFinder) were used for phylogenetic inference (Kalyaanamoorthy et al., 2017; Quang et al., 2008). When using the complex LG + C40 + F + G4 model, a posterior mean site frequency (PMSF) approximation was used to calculate non-parametric bootstraps (Wang et al., 2018). Bayesian inference was performed with PhyloBayes MPI 1.4 (Lartillot et al., 2009) using the GTR matrix in combination with the CAT infinite mixture model and four gamma rate categories. We ran four MCMC chains for at least 10,000 generations and saved

every second tree. We tested for convergence (maxdiff < 0.1) using bpcomp, implemented in PhyloBayes, with default parameters (burn-in of 20%). Two chains of the four converged with a maxdiff = 0.0550384. Approximately unbiased tests were conducted using CONSEL v0.2 with per-site log-likelihood values calculated in IQ-Tree v1.6.6 using either the LG+C40+F+G4 or LG+F+R6 models (Nguyen et al., 2015; Shimodaira and Hasegawa, 2001).

3. Results

3.1. Phylogenomic analysis

Incorporation of the *A. solis* (AF-21), *A. prolabens.*, and Lap-1 transcriptomes into our phylogenomic dataset resulted in the addition of three new taxa with moderate to high proportions of data coverage. Of the 44,786 amino acids in the concatenated alignment, *A. solis, A. prolabens*, and Lap-1 had 83% (37,023 amino acids (aa)), 26% (11,789 aa), and 86% (38,637 aa) of sites present, respectively. Moreover, new genomic and transcriptomic data from *P. brassicae, Spongospora subterranea*, and *P. chromatophora* increased the data representation of these taxa from 13% (6,471 aa), 7% (3,633 aa), and 17% (8,853 aa) in a recent study (Cavalier-Smith et al., 2018), to 97% (43,598 aa), 60% (26,856 aa), and 96% (43,011 aa) (Fig. 1).

With this substantially updated alignment, we performed



Fig. 1. Maximum likelihood tree generated from an alignment comprising 161 genes and 44,786 sites using the LG + C40 + F + G4 substitution model implemented in IQ-Tree. Support was obtained from 100 standard non-parametric bootstraps (NPB) and 1000 ultrafast bootstraps (UFB). Black circles represent 100NPB/100UFB and values below 75 are not shown. The percent of genes (orange) and sites (blue) present for each taxon is shown on the left. The Cercozoa and Retaria are highlighted in blue and orange, respectively. A; Alveolata, R; Rhizaria, S; Stramenopiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phylogenomic analyses. Maximum likelihood analyses using both the empirical profile mixture model (Fig. 1) and the free rate heterogeneity model (Fig. S2) recovered a monophyletic Cercozoa with nearly full support. In agreement with previous studies and current systematics, our ML analyses also recovered a well supported and monophyletic Rhizaria, Retaria, Reticulofilosa, Monadofilosa, and Endomyxa (Fig. 1, Fig. S2). Furthermore, the Endomyxa was sister to a monophyletic group comprising the Reticulofilosa and Monadofilosa (Fig. 1, Fig. S2). However, the topologies within some of these clades, particularly the Monadofilosa, lacked support and differed in each of the trees. Sampling of the Reticulofilosa was limited to the chlorarachniophytes, so little can be said about the branching order within this group. Furthermore, A. solis and A. prolabens clustered within the Monadofilosa. whereas Lap-1 was sister to the Endomyxa, consistent with previously generated SSU rRNA phylogenies (Bass et al., 2018; Shiratori et al., 2014) (Fig. 1, Fig. S2).

In contrast to the ML trees, Bayesian analyses produced mixed results in regards to the monophyly of the Cercozoa. Of the four Markov chains, two converged chains and a third chain revealed a paraphyletic topology with the Retaria branching within the Endomyxa, whereas a single chain generated the monophyletic tree (Fig. S3). However, approximately-unbiased (AU) topology tests rejected the topology of the paraphyletic trees, and these trees had lower log-likelihood values when compared to those of both the best trees and the monophyletic Bayesian tree when using either of the ML models (Fig. S3). Given the mixed results of the Bayesian analysis, we also sought to test whether the monophyly of the Endomyxa, Lap-1, the formanifera, and the radiolarians could be statistically rejected. To this end we performed AU tests on ML trees generated using topology constraints and both substitution models (Fig. 2). As with the Bayesian trees, the constrained ML trees were rejected by AU test (Fig. 2).

3.2. Description of clone Lap-1

Given the unique phylogenetic position of clone Lap-1, we sought to morphologically characterize and provide a description of the species, which has only previously been reported as an environmental isolate (Bass et al., 2018). The cells of Lap-1 (Fig. 3A-L, Video S1) are generally flattened but highly metabolic, taking on various forms. Moving cells are usually obovate, $21-34 \mu m$ long, $15-22 \mu m$ wide, and have a broader anterior end (Fig. 3A-C) (n = 20). Cells are sometimes observed shrunken and rounded down to 12 µm in diameter (Fig. 3K). A notable lateral depression is situated in the middle lateral point of the cell body (Fig. 3B), and two unequal flagella are present (Fig. 3A, D, F, K). The anterior flagellum originates on the anterior part of the cell from a small pit, is about half the length of the cell, and points to the side. The posterior flagellum is visible from the lateral depression, is about 1.5 times longer than the cell body (Fig. 3B), points to the side opposite to the anterior flagellum, and turns backwards. Cells usually swim (sometimes quickly) near the surface by rapidly beating both flagella, but gliding on the flagella was not observed. Both flagella can also beat very fast when the cell is stopped, which creates a visual artifact of the presence of four flagella (Fig. 3D). One (rarely two) large contractile vacuole is usually situated at the anterior end of the cell (Fig. 3E). A nucleus is located medially and slightly closer to the anterior end of the cell (Fig. 3A,F) and the cytoplasm also contains light-refracting granules. Lap-1 cells tend to form one large 'tail' pseudopodium (Fig. 3G-J) and can attach to substrates using it (Fig. 3I). In some instances, however, short, wide or thin, pseudopodia were observed all over the cell surface (Fig. 3F). The organism was observed travelling short distances by amoeboid movement.



Video S1. Lapot gusevi moving.



Fig. 2. Maximum likelihood trees generated using the 161 gene concatenation in IQ-Tree with the monophyly of the Endomyxa, *L. gusevi*, and the Retaria constrained. Trees were estimated using either LG + C40 + F + G4 (A) or LG + F + R6 (B) substitution models. *P*-values were calculated using approximately unbiased tests and the difference in log-likelihood relative to the unconstrained tree is shown below (Δ LogL). Cercozoa and Retaria are highlighted in blue and orange, respectively. The branch leading to *Guttulinopsis vulgaris* has been reduced by half for simplicity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



Fig. 3. Morphology of *Lapot gusevi* n. gen. n. sp (Lap-1). A-D. general cell view with flagella; the lateral depression with the outgoing posterior flagellum is visible. E. Cell with a large contractile vacuole on the anterior end. F-J. Diversity of forms of the pseudopodia. K. Shrunk and rounded cell. L. Cyst. Scale bar: A ,20 µm; B, 10 µm; C-K, 15 µm; L, 12 µm.

Lap-1 cells are predatory and fed on other flagellates (e.g., bodonids and stramenopiles) by consuming prey cells in their entirety. Lap-1 was dependent on feeding as cultures perished in the absence of supplemented eukaryotic prey. A large food vacuole is formed at the posterior end of the cell body following feeding (Fig. 3C). Division was not observed but Lap-1 was noted to form roundish cysts, 21 μ m in diameter (Fig. 3L).

4. Discussion

4.1. The monophyly of the Cercozoa

Here we generated new transcriptomes from three cercozoan taxa, including the novel species Lap-1 (hereby referred to as *Lapot gusevi* n. gen. n. sp.), and integrated these and other new publically available data from an additional six Cercozoa (two endomyxans, three reticulofilosans, and one monadofilosan) into a previously developed phylogenomic framework (Burki et al., 2016, 2013, 2012). In contrast to previous phylogenomic studies, our analyses recovered a robustly supported cercozoan clade, a result which challenges the recent

proposal to transfer the Endomyxa to the Retaria (Cavalier-Smith, 2018). A previous report attempted to reconcile this transfer using morphological characters, particularly the predominance of filose (thread-like) pseudopodia in the Reticulofilosa and Monadofilosa, as opposed to the reticulose (net-like) pseudopodia in the Endoymyxa and Retaria (Cavalier-Smith, 2018). However, these characters are not synapomorphic and other characters, such as the absence of cortical alveoli in cercozoans, endomyxans, and formanifera are not consistent with molecular phylogenies, and clearly have a complex history (Cavalier-Smith et al., 2018). Evidently, the morphological diversity observed amongst the rhizarians makes morphology-based taxonomy within this group challenging, thus increasing the importance of molecular markers. However, besides insertions in ribosomal protein L1 and the deletion in the SSU rRNA, few molecular traits have been identified that provide insights into cercozoan phylogeny (Burki et al., 2010; Cavalier-Smith et al., 2018; Cavalier-Smith and Chao, 2003b). A recent transcriptomic study examining rhizarian cytoskeletal proteins proposed that there was an actin duplication in the last common ancestor of the Endomyxa and Retaria, supporting cercozoan paraphyly (Krabberød et al., 2017). However, actin phylogenies revealed only a

single endomyxan paralog and suggest that this is more likely a Retariaspecific duplication (Krabberød et al., 2017). Overall, no set of morphological or molecular traits has been identified that either supports or rejects cercozoan monophyly, emphasizing the importance of accurate phylogenomics in inferring the evolutionary history of this group.

Although our ML analyses recovered a monophyletic cercozoa, mixed signals were observed in the data as exemplified by the Bayesian analysis which failed to converge and also produced a number of paraphyletic topologies. However, these paraphyletic trees had consistently lower log-likelihoods relative to the ML tree and were rejected by AU-test when assessed with ML models. This suggests that these topologies may represent local maxima in tree space which may have hindered Markov chain progress and ultimately convergence. Consistent with this, constrained ML trees showing a paraphyletic cercozoa were also rejected by AU-test and had lower log-likelihoods compared to the non-constrained tree.

It is possible that the addition of data and new taxa and the use of sophisticated substitution models helped avoid artifactual topologies. Missing data and partial genes can bias phylogenomics and lead to erroneous results (Hosner et al., 2016; Roure et al., 2013). This may be particularly relevant to rhizarian analyses as many cercozoan and retarian taxa are challenging to culture and as a result, transcriptomic experiments are often dependent on single cell methods (Burki et al., 2010; Krabberød et al., 2017). These datasets can be limited due to the scarcity of the starting material, making the inclusion of deeply-sequenced samples pertinent. Therefore, we suspect that additional sampling of diverse cercozoan clades and the generation of deeply sequenced transcriptomic datasets and sequenced rhizarian genomes will help in resolving the rhizarian tree and will further consolidate the monophyly of the Cercozoa.

4.2. Morphology, phylogeny and description of Lapot gusevi n. gen. n. sp.

The phylogenetic position of *L. gusevi* within Rhizaria was assessed previously using SSU rRNA gene sequences (Bass et al., 2018). In Bayesian phylogenetic analysis, *L. gusevi* falls within the strongly supported 'Novel Clade 10' (order Aquavolonida Bass and Berney, 2018) which forms a clade with the Tremulida, which together is sister to the environmental 'Novel Clade 12'. This whole group (Aquavolonida, Tremulida, and 'Novel Clade 12') was shown to be sister to the Endomyxa, albeit without strong support (Bass et al., 2018). Our analyses supported this conclusion by consistently placing *L. gusevi* sister to the Endomyxa, but again with low statistical support. Further transcriptomic sampling of the Aquavolonidae, Tremulida, and 'Novel Clade 12' will likely resolve this topology.

Lapot gusevi shares some characteristics with related lineages, but can be clearly distinguished morphologically. Like Aquavolon spp. and Tremula longifila, L. gusevi possesses a pronounced medium notch. In contrast to Aquavolon spp. and T. longifila, L. gusevi is flattened and not elongated or spindle shaped and is much more metabolic (amoeboid). Also unlike T. longifila, L. gusevi does not glide on its flagella while in contact with substratum, and in contrast to Aquavolon spp. it does not perpetually rotate along its longitudinal axis. Both recognized genera of the Aquavolonida (Lapot and Aquavolon) are eukaryovorous protists which capture whole prey cells. Indeed, eukaryovory is not common amongst cercozoan protists and mostly know in the metromonads, Metromonas simplex and Metopion fluens, and the pansomonad A. solis. Furthermore, the presence of flagellated cells is very rare within the Endomyxa, a group of largely amoeboid protists producing filopodia, rhizopodia, and reticulopodia (e.g., the reticulopodial amoeba Filoreta marina), many of which are endoparasites. However, some of the endomyxids produce flagellated dispersal cells or zoospores, such as Gromia oviformis, plasmodiophorids (Phytomyxea), and paradinids (Ascetosporea) (Barr and Allan, 1982; Chatton, 1910; Hedley and Bertaud, 1962).

combination with previous SSU rRNA gene analyses (Bass et al., 2018), allow us to classify the clone Lap-1 to a new genus and new species of rhizarian.

Taxonomy: Eukaryota; SAR; Rhizaria Cavalier-Smith 2002; Cercozoa Cavalier-Smith 1998, emend. Adl et al. 2005, Aquavolonida Bass and Berney, 2018

Lapot n. gen. Tikhonenkov, Mylnikov, Irwin and Keeling

Diagnosis: Unicellular protist with two smooth flagella. Cells are very metabolic and flattened, with a notable lateral depression in the middle lateral point of the cell body. The nucleus is located medially, slightly closer to anterior end of the cell. Cells are observed rapidly swimming near the substrate by beating both flagella. Eukaryovorous.

Etymology: Derived from 'Lapot', a traditional Russian bastshoe, due to the resemblance in widened and flattened shape.

Zoobank Registration. urn:lsid:zoobank.org:act:7214A621-6004-4722-930E-3700A0FD23A5

Type species. Lapot gusevi.

Lapot gusevi n. sp. Tikhonenkov, Mylnikov, Irwin and Keeling

Diagnosis: Cells obovate with lateral depression, $21-34 \mu m$ long, and $15-22 \mu m$ wide with broader anterior end. Anterior flagellum originated in the anterior part and is about half of the cell length, directs to the side. Posterior flagellum becomes visible from the lateral depression and is 1.5 times longer than cell body, directs to the side opposite to the anterior flagellum and backwards. Cells swim near the surface by rapid beating of both flagella and not gliding. One to two large contractile vacuoles. Cells able to produce large pseudopodium posteriorly and short wide or thin pseudopodia all over the cell surface. Forms roundish cysts about 21 μm in diameter.

Type Figure: Fig. 3A illustrates a live cell of strain Lap-1.

Gene sequence: The SSU rRNA gene sequence has the GenBank Accession Number 818165.

Type locality: Sediment of a shallow pool, Cát Tiên National Park, Dong Nai Province, S.R. Vietnam.

Etymology: Named after Dr. Evgeny Gusev, Russian algologist, who significantly contributed to sample collection and field trip organization in S.R. Vietnam.

Zoobank Registration: urn:lsid:zoobank.org:act:7801EDE4-B580-4AA4-922E-8C904F804A34

Notes: Due to the presumably temporary character of the waterbody where the species was found and low values of water conductivity (likely rainwater), the possibility of habitation in the soil is not excluded.

Acknowledgements

We thank Dr. Viktor Komov, Larysa Pakhomova, Dr. Evgeny Gusev, Dr. Vladimir Gusakov, and Tran Duc Dien for help with sample collection in the Canary Islands, Strait of Georgia (British Columbia), and Vietnam, as well as the staff of the Vietnam-Russian Tropical Centre, Coastal Branch (Nha Trang, Vietnam), especially Nguyen Thị Hai Thanh and Tran Thanh Quang, for their assistance with trip management and sampling in Vietnam.

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) (grant no. 227301) awarded to PJK and from the Russian Foundation for Basic Research (grant no. 17-04-00565) awarded to APM. The work of DVT was supported by the Russian Science Foundation (grant no. 18-14-00239). NATI was supported by an NSERC Canadian Graduate Scholarship.

Competing interests

The authors declare no competing interests.

References

Ultimately, the phylogenomic and morphological data, in

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local

Alignment Search Tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

- Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
- Archibald, J.M., Longet, D., Pawlowski, J., Keeling, P.J., 2003. A novel polyubiquitin structure in Cercozoa and Foraminifera: Evidence for a new eukaryotic supergroup. Mol. Biol. Evol. 20, 62–66. https://doi.org/10.1093/molbev/msg006.
- Barr, D.J.S., Allan, P.M.E., 1982. Zoospore ultrastructure of *Polymyxa graminis* (Plasmodiophoromycetes). Can. J. Bot. 60, 2496–2504. https://doi.org/10.1139/ b82-302.
- Bass, D., Moreira, D., López-García, P., Polet, S., Chao, E.E., Von Der Heyden, S., Pawlowski, J., Cavalier-Smith, T., 2005. Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. Protist 156, 149–161. https://doi.org/10.1016/j.protis. 2005.03.001.
- Bass, D., Tikhonenkov, D.V., Foster, R., Dyal, P., Janouskovec, J., Keeling, P.J., Gardner, M., Neuhauser, S., Hartikainen, H., Mylnikov, A.P., Berney, C., 2018. Rhizarian 'Novel Clade 10' Revealed as Abundant and Diverse Planktonic and Terrestrial Flagellates, including Aquavolon n. gen. J. Eukaryot. Microbiol. https://doi.org/10. 1111/jeu.12524.
- Bhattacharya, D., Helmchen, T., Melkonian, M., 1995. Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent rhizopod lineage containing the euglyphina and the chlorarachniophyta. J. Eukaryot. Microbiol. 42, 65–69. https://doi.org/10.1111/j.1550-7408.1995.tb01541.x.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/ bioinformatics/btu170.
- Brown, M.W., Kolisko, M., Silberman, J.D., Roger, A.J., 2012. Aggregative multicellularity evolved independently in the eukaryotic supergroup Rhizaria. Curr. Biol. 22, 1123–1127. https://doi.org/10.1016/j.cub.2012.04.021.
- Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., Keeling, P.J., 2013. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. Curr. Biol. 23, 1541–1547. https://doi.org/10.1016/j. cub.2013.06.033.
- Burki, F., Kaplan, M., Tikhonenkov, D.V., Zlatogursky, V., Minh, B.Q., Radaykina, L.V., Smirnov, A., Mylnikov, A.P., Keeling, P.J., 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. https://doi.org/ 10.1098/rspb.2015.2802.
- Burki, F., Kudryavtsev, A., Matz, M.V., Aglyamova, G.V., Bulman, S., Fiers, M., Keeling, P.J., Pawlowski, J., 2010. Evolution of Rhizaria: New insights from phylogenomic analysis of uncultivated protists. BMC Evol. Biol. 10. https://doi.org/10.1186/1471-2148-10-377.
- Burki, F., Okamoto, N., Pombert, J.-F., Keeling, P.J., 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc. R. Soc. B Biol. Sci. 279, 2246–2254. https://doi.org/10.1098/rspb.2011.2301.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
- Cavalier-Smith, T., 2018. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. Protoplasma 255, 297–357. https://doi.org/10.1007/s00709-017-1147-3.
- Cavalier-Smith, T., 1998. A revised six-kingdom system of life. Biol. Rev. 73, 203–266. https://doi.org/10.1111/j.1469-185X.1998.tb00030.x.
- Cavalier-Smith, T., Chao, E.E., Lewis, R., 2018. Multigene phylogeny and cell evolution of chromist infrakingdom Rhizaria: contrasting cell organisation of sister phyla Cercozoa and Retaria. Protoplasma 255, 1517–1574. https://doi.org/10.1007/ s00709-018-1241-1.
- Cavalier-Smith, T., Chao, E.E.Y., 2003a. Phylogeny and Classification of Phylum Cercozoa (Protozoa). Protist 154, 341–358. https://doi.org/10.1078/143446103322454112.
- Cavalier-Smith, T., Chao, E.E.Y., 2003b. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. J. Mol. Evol. 56, 540–563. https://doi. org/10.1007/s00239-002-2424-z.
- Chatton, É., 1910. Paradinium poucheti n. g., n. sp., flagellé parasite d'Acartia clausi Giesbrecht (Copépode pélagique) (Note préliminaire). C. R. Soc. Biol. 69, 341–343.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084.
- Hedley, R.H., Bertaud, W.S., 1962. Electron-Microscopic Observations of Gromia oviformis (Sarcodina). J. Protozool. 9, 79–87.
- Hehenberger, E., Tikhonenkov, D.V., Kolisko, M., del Campo, J., Esaulov, A.S., Mylnikov, A.P., Keeling, P.J., 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. https://doi.org/10.1016/j.cub.2017.06.006.
- Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2016. Avoiding

missing data biases in phylogenomic inference: An empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. 33, 1110–1125. https://doi.org/10.1093/molbev/msv347.

- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https:// doi.org/10.1093/molbev/mst010.
- Keeling, P.J., 2002. Molecular phylogenetic position of *Trichomitopsis termopsidis* (Parabasalia) and evidence for the Trichomitopsiinae. Eur. J. Protistol. 38, 279–286. https://doi.org/10.1078/0932-4739-00874.
- Keeling, P.J., 2001. Foraminifera and Cercozoa are related in actin phylogeny: Two orphans find a home? Mol. Biol. Evol. 18, 1551–1557. https://doi.org/10.1093/ oxfordjournals.molbev.a003941.
- Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biol. 12. https://doi.org/10.1371/journal. pbio.1001889.
- Keeling, P.J., Deane, J.A., McFadden, G.I., 1998. The phylogenetic position of alpha- and beta-tubulins from the *Chlorarachnion* host and Cercomonas (Cercozoa). J. Eukaryot. Microbiol. 45, 561–570.
- Krabberød, A.K., Orr, R.J.S., Bråte, J., Kristensen, T., Bjørklund, K.R., Shalchian-Tabrizi, K., 2017. Single cell transcriptomics, mega-phylogeny, and the genetic basis of morphological innovations in Rhizaria. Mol. Biol. Evol. 34, 1557–1573. https://doi. org/10.1093/molbev/msx075.
- Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. F1000Research 6, 1287. https://doi.org/10.12688/f1000research.12232.1.
- Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25, 2286–2288. https://doi.org/10.1093/bioinformatics/btp368.
- Longet, D., Archibald, J.M., Keeling, P.J., Pawlowski, J., 2003. Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. Int. J. Syst. Evol. Microbiol. 53, 1735–1739. https://doi.org/10.1099/ijs.0.02597-0.
- Medlin, L., Elwood, H.J., Stickel, S., Sogin, M.L., 1988. The characterization of enzymatically amplified eukaryotes 16S-like ribosomal RNA coding regions. Gene 71, 491–500.
- Moreira, D., von der Heyden, S., Bass, D., López-García, P., Chao, E., Cavalier-Smith, T., 2007. Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excava ta. Mol. Phylogenet. Evol. 44, 255–266. https://doi.org/10.1016/j.ympev.2006.11.001.
- Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.
- Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181. https://doi.org/10.1038/nprot.2014.006.
- Quang, L.S., Gascuel, O., Lartillot, N., 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24, 2317–2323. https://doi.org/10.1093/ bioinformatics/btn445.
- Rambaut, A., 2012. FigTree, a graphical viewer of phylogenetic trees. http//tree.bio.ed. ac.uk/software/figtree.
- Roure, B., Baurain, D., Philippe, H., 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol. Biol. Evol. 30, 197–214. https://doi. org/10.1093/molbev/mss208.
- Roure, B., Rodriguez-Ezpeleta, N., Philippe, H., 2007. SCaFoS: A tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. 7. https://doi. org/10.1186/1471-2148-7-S1-S2.
- Schweim, A., Fogelqvist, J., Knaust, A., Jülke, S., Lilja, T., Bonilla-Rosso, G., Karlsson, M., Shevchenko, A., Dhandapani, V., Choi, S.R., Kim, H.G., Park, J.Y., Lim, Y.P., Ludwig-Müller, J., Dixelius, C., 2015. The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. Sci. Rep. 5, 1–12. https://doi.org/10. 1038/srep11153.
- Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247. https://doi.org/10.1093/ bioinformatics/17.12.1246.
- Shiratori, T., Yokoyama, A., Ishida, K.I., 2014. Phylogeny, Ultrastructure, and Flagellar Apparatus of a New Marimonad Flagellate *Abollifer globosa* sp. nov. (Imbricatea, Cercozoa). Protist 165, 808–824. https://doi.org/10.1016/j.protis.2014.10.003.
- Sierra, R., Canas-Duarte, S.J., Burki, F., Schwelm, A., Fogelqvist, J., Dixelius, C., González-García, L.N., Gile, G.H., Slamovits, C.H., Klopp, C., Restrepo, S., Arzul, I., Pawlowski, J., 2015. Evolutionary origins of rhizarian parasites. Mol. Biol. Evol. 33, 980–983. https://doi.org/10.1093/molbev/msv340.

- Sierra, R., Matz, M.V., Aglyamova, G., Pillet, L., Decelle, J., Not, F., de Vargas, C., Pawlowski, J., 2013. Deep relationships of Rhizaria revealed by phylogenomics: A farewell to Haeckel's Radiolaria. Mol. Phylogenet. Evol. 67, 53–59. https://doi.org/ 10.1016/j.ympev.2012.12.011.
- Takishita, K., Inagaki, Y., Tsuchiya, M., Sakaguchi, M., Maruyama, T., 2005. A close relationship between Cercozoa and Foraminifera supported by phylogenetic analyses based on combined amino acid sequences of three cytoskeletal proteins (actin, αtubulin, and β-tubulin). Gene 362, 153–160. https://doi.org/10.1016/j.gene.2005. 08.013.
- Tikhonenkov, D.V., Janouškovec, J., Keeling, P.J., Mylnikov, A.P., 2016. The Morphology, Ultrastructure and SSU rRNA Gene Sequence of a New Freshwater Flagellate, *Neobodo borokensis* n. sp. (Kinetoplastea, Excavata). J. Eukaryot. Microbiol. 63, 220–232. https://doi.org/10.1111/jeu.12271.
- Tikhonenkov, D.V., Janouškovec, J., Mylnikov, A.P., Mikhailov, K.V., Simdyanov, T.G., Aleoshin, V.V., Keeling, P.J., 2014. Description of *Colponema vietnamica* sp.n. and *Acavomonas peruviana* n. gen. n. sp., Two New Alveolate Phyla (Colponemidia nom.

nov. and Acavomonidia nom.nov.) and Their Contributions to Reconstructing the Ancestral State of Alveolates and Eukaryotes. PLoS One 9. https://doi.org/10.1371/journal.pone.0095467.

- Tikhonenkov, D.V., Mazei, Y.A., Embulaeva, E.A., 2008. Degradation succession of heterotrophic flagellate communities in microcosms. Zh Obs. Biol 69, 57–64.
- 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, 216–235. https://doi.org/10.1093/sysbio/syx068.
- Wylezich, C., Meisterfeld, R., Meisterfeld, S., Schlegel, M., 2002. Phylogenetic analysis of small subunit ribosomal RNA coding regions reveal a monophyletic lineage of Euglyphid Testate Amoebae (Order Euglyphida). J. Eukaryot. Microbiol. 49, 108–118.
- Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620. https://doi.org/10.1093/ bioinformatics/btt593.