

Article (refereed) - postprint

Schuler, Gabriel A.; Tice, Alexander K.; Pearce, Rebecca A.; Foreman, Emily; Stone, Jared; Gammill, Sarah; Wilson, John D.; Reading, Chris; Silberman, Jeffrey D.; Brown, Matthew W. 2018. **Phylogeny and classification of novel diversity in Sainouroidea (Cercozoa, Rhizaria) sheds light on a highly diverse and divergent clade.**

© 2018 Elsevier GmbH.

This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>



This version available <http://nora.nerc.ac.uk/520809/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

NOTICE: this is the author's version of a work that was accepted for publication in *Protist*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in ***Protist* (2018), 169 (6), 853-874.**

<https://doi.org/10.1016/j.protis.2018.08.002>

www.elsevier.com/

Contact CEH NORA team at
noraceh@ceh.ac.uk

ORIGINAL PAPER

Phylogeny and Classification of Novel Diversity in Sainouroidea (Cercozoa, Rhizaria) Sheds Light on a Highly Diverse and Divergent Clade

Running title: Evolution of the Highly Diverse Sainouroidea Amoebae

Gabriel A. Schuler^a, Alexander K. Tice^a, Rebecca A. Pearce^b, Emily Foreman^a, Jared Stone^b, Sarah Gammill^b, John D. Wilson^b, Chris Reading^c, Jeffrey D. Silberman^b, and Matthew W. Brown^{a,1}

^aDepartment of Biological Sciences, Mississippi State University, MS State 39762, MS, USA

^bDepartment of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

^cCentre for Ecology and Hydrology, Wallingford, Oxon, OX10 8BB, UK

Submitted March 8, 2018; Accepted August 10, 2018

Monitoring Editor: Alastair Simpson

¹Corresponding author;

e-mail Matthew.Brown@msstate.edu (M. W. Brown).

Sainouroidea is a molecularly diverse clade of cercozoan flagellates and amoebae in the eukaryotic supergroup Rhizaria. Previous 18S rDNA environmental sequencing of globally collected fecal and soil samples revealed great diversity and high sequence divergence in the Sainouroidea. However, a very limited amount of this diversity has been observed or described. The two described genera of amoebae in this clade are *Guttulinopsis*, which displays aggregative multicellularity, and *Rosculus*, which does not. Although the identity of *Guttulinopsis* is straightforward due to the multicellular

fruiting bodies they form, the same is not true for *Rosculus*, and the actual identity of the original isolate is unclear. Here we isolated amoebae with morphologies like that of *Guttulinopsis* and *Rosculus* from many environments and analyzed them using 18S rDNA sequencing, light microscopy, and transmission electron microscopy. We define a molecular species concept for Sainouroidea that resulted in the description of 4 novel genera and 12 novel species of naked amoebae. Aggregative fruiting is restricted to the genus *Guttulinopsis*, but other than this there is little morphological variation amongst these taxa. Taken together, simple identification of these amoebae is problematic and potentially unresolvable without the 18S rDNA sequence.

Keywords: Amoebae; sorocarpic multicellularity; evolution; zoonotic amoebae; morphometrics.

Introduction

The rhizarian group Cercozoa, Cavalier-Smith 1998, is a morphologically diverse group of eukaryotes that does not seem to have distinctive unifying morphological characteristics and was formed based on molecular phylogenetics (Adl et al. 2012; Cavalier-Smith 1998). Environmental sequencing of 18S rDNAs continually reveals high sequence diversity and taxon diversity within Cercozoa (Bass and Cavalier-Smith 2004; Bass et al. 2016; Fiore-Donno et al. 2017). In particular, members of the cercozoan group Sainouroidea, Cavalier-Smith et al. 2009, have largely eluded detection due to their highly divergent 18S rDNA, which makes amplification using “Universal Eukaryotic” PCR primers problematic (Bass et al. 2005; Bass et al. 2016; Brown et al. 2012a). Currently Sainouroidea contains five genera: *Cholamonas*, *Sainouron*, *Helkesimastix*, *Guttulinopsis*, and *Rosculus* (Bass et al. 2016; Cavalier-Smith et al. 2009). However, many sainouroid 18S rDNA OTUs found in environmental samples may represent unclassified clades (Bass et al. 2016).

Sainouroidea branch among a group of ancestrally amoeboid bi-flagellates that typically lack an outer cell coat (scales or theca) within a group referred to as core-cercozoans (Monadofilosa subdivision of Filosa) (Cavalier-Smith et al. 2009). It is common for these organisms to have a gliding motility, in which cells glide on their posterior flagellum, tubular mitochondrial cristae, and a microbody attached to the nucleus (Cavalier-Smith and Chao 2003). In the Sainouroidea, the three genera *Cholamonas* (from the gut of a diopsid fly), *Sainoureon* (from soils), and *Helkesimastix* (from marine sediments and goat dung) each have a flagellated life stage with some of the features of typical heterotrophic cercozoan flagellates (Cavalier-Smith et al. 2008, 2009; Flavin et al. 2000; Sandon 1924; Woodcock and Lapage 1915). Unlike most cercozoans, *Sainoureon* and *Helkesimastix* have flat mitochondrial cristae (Cavalier-Smith et al. 2009, Dumack et al. 2017). Sainouroidea was created after sequencing of the 18S rDNA from *C. cytrodiopsidis*, *S. acronematica*, and *H. marina*, which revealed a highly divergent clade within the Cercozoa (Cavalier-Smith et al. 2009).

The classification of the naked amoeba genera in Sainouroidea has changed multiple times. The first described genus currently included in this group is the aggregatively multicellular (sorocarpic) *Guttulinopsis*, found primarily on herbivore dung (Olive 1901). There are four described *Guttulinopsis* species: *G. vulgaris*, *G. stipitata*, *G. clavata*, *G. nivea* (Olive 1901; Raper et al. 1977). Naked amoebae have been placed in this genus primarily based on their ability to form sorocarpic fruiting bodies with a round white sorus (Olive 1902; Raper et al. 1977). The ability to form fruiting bodies in this manner initially led *Guttulinopsis* to be classified in the order Acrasieae, Olive 1901, which at the time contained all the known sorocarpic amoebae (Olive 1901, reviewed in Brown et al. 2012b). In 1988, *Guttulinopsis* was assigned to the group Heterolobosea based on amoeba morphology and flat mitochondrial cristae (Page 1988). As molecular phylogenies using the 18S rDNA sequence revised phylogenies and classifications of eukaryotes, researchers were unable to amplify the 18S rDNA sequence of *Guttulinopsis* (Bass et al. 2016; Brown et al. 2012a) and it was not until 2012 that a phylogenomic analysis of 159 proteins surprisingly placed *Guttulinopsis vulgaris* in the supergroup Rhizaria, contrary to previous classifications (Brown et al. 2012a). Subsequently, the 18S rDNA sequence was extracted from the

transcriptome and incorporated into a taxon-rich data set to more precisely place *Guttulinopsis vulgaris* in the cercozoan group Sainouroidea (Bass et al. 2016). Out of the four described species, *G. vulgaris* has remained the only one with published molecular data.

The other genus of naked amoebae in Sainouroidea is *Rosculus*, Hawes 1963, which was originally isolated from the rectum of a European grass snake, *Natrix natrix* (Hawes 1963). The morphology, ultrastructure and movement of amoebae in this genus is indistinguishable from that of *Guttulinopsis* (Page 1988). The principal difference between *Guttulinopsis* and *Rosculus* amoebae is the observed fruiting of *Guttulinopsis* amoebae (Page 1988). *Rosculus ithacus* is a fast-growing amoeba that can survive in both aerobic and anaerobic conditions (Hawes 1963). Amoebae morphologically identified as '*Rosculus*' are found free-living and infecting various animal hosts including snakes, fish, and the human nasal cavity (Dykova et al. 1996; Hawes 1963; Visvesvara et al. 1982). This suggested to Page (1974) that *Rosculus* is an animal-associated (amphizoic) protist genus, "which not only occur but also feed and multiply well in both the free-living (exozoic) and endozoic conditions". The genus *Rosculus* is currently represented by three species *R. ithacus*, *R. terrestris*, *R. elongata* (Bass et al. 2016; Hawes 1963).

In 2016, partial 18S rDNAs were amplified from a variety of fecal environments using universal eukaryotic and sainouroid clade-specific 18S rDNA PCR primers (Bass et al. 2016). This study demonstrated that previous eukaryotic rDNA environmental sampling excluded sequences from sainouroids due to their highly divergent SSU sequences (Bass et al. 2016). Bass et al. (2016) also revealed previously unknown diversity within Sainouroidea and found cercozoans (containing Sainouroidea) to be the most diverse group of eukaryotes in fecal environments.

In an effort to re-isolate the type species of *Rosculus* for inclusion in molecular phylogenetic analyses and to better characterize the diversity of sainouroid amoebae, we purchased from culture collections every culture accessioned as '*Rosculus*' and isolated numerous other amoebae with a similar morphology to *Rosculus* and *Guttulinopsis* from many different environments. These environments

included feces from the European grass snake (*N. natrix*), the type host of *R. ithacus*, North American relatives of *N. natrix* plus other snake species and selected prey items of these snakes. In addition to soils and freshwater, we sampled feces or intestinal contents from cows (*Bos taurus*), chickens (*Gallus gallus*), wild turkeys (*Meleagris gallopavo*), and camel crickets (*Ceuthophilus* sp.). The 18S rRNA gene from each strain was sequenced for inclusion in molecular phylogenetic analyses. We characterized all strains by light microscopy, and some using transmission electron microscopy. Qualitative morphological differences were observed among amoeba isolates. Using a molecular species concept based on 18S rDNA phylogenetic tree topology plus percent sequence divergence, our data revealed 4 novel genera (in addition to the already described genera *Rosculus* and *Guttulinopsis*) and 12 novel species of sainouroid amoebae. Some life history characteristics are discussed.

Results

Molecular Phylogeny

A total of 36 monoculture amoeba strains morphologically similar to *Guttulinopsis* and *Rosculus* were isolated from the environment or purchased from culture collections and their 18S rDNAs were sequenced (Fig. 1, Supplementary Material Table S1). Partial 18S rDNA sequence from an organism accessioned at the American Type Culture Collection (ATCC) as '*Rosculus* sp.' ATCC PRA-134 revealed that this isolate was mis-identified and mis-accessioned. It is actually closely related to *Micriamoeba* in the supergroup Amoebozoa, with an almost identical sequence to *Micriamoeba tesseri* (data not shown, SSU rDNA: GenBank MH643883). Previous analyses based on phylogenomics and light microscopy demonstrated that ATCC PRA-134 is an amoebozoan affiliated with *Micriamoeba* (Kang et al. 2017), thus it was excluded from further analyses within. The 18S rDNA phylogenetic tree showed that all other strains analyzed in this study branch within a highly supported Sainouroidea clade within Rhizaria (based on a 1,356 bp masked alignment containing 129 sequences, Fig. 2). Sainouroidea robustly branches within Filosa; specifically within Monadofilosa. However, the precise position of Sainouroidea within

Monadofilosa is unclear due to low support values in both Bayesian inference (BI) and Maximum Likelihood (ML) analysis (Fig. 2). Figure 2 shows *Cholamonas*, followed by *Sainouron*, as the basal branching genera in Sainouroidea, as seen in previous studies (Cavalier-Smith et al. 2008, 2009). A number of the new isolates formed a clade with either the previously published *Guttulinopsis* or *Rosculus* 18S rDNA sequences, but quite a few did not. Instead they delineated several novel sainouroid lineages (Fig. 2).

Genus and Species Delineation

To determine genus and species-level nodes within Sainouroidea in a reproducible manner without human induced bias, an uncorrected pairwise distance matrix of aligned 18S rDNA sequences trimmed of ambiguously aligned sites was used. This matrix was inferred in an automated fashion using a pipeline described below. Only full or nearly full length (i.e., greater than 1,500 bp) sainouroid 18S rDNA sequences were used in an uncorrected pairwise distance matrix generated from unambiguously aligned and masked positions (1,393 bp). The strains without a full length 18S rDNA sequence were not classified past the genus level; these included *R. elongata* and *R. terrestris* (from Bass et al. 2016), RA (*Puppisaman* sp.), and STA (*Guttulinopsis* sp.) (Supplementary Material Table S1, Fig. 2). This uncorrected pairwise distance matrix is shown with a BI phylogeny of the same alignment (Fig. 3). Based on the tree, the extent of pairwise distances suggested discrete cutoff values for genus and species delineation: we propose a difference of less than 1.7% to determine a species-level designation and taxa within 12% to belong to the same genus (Fig. 3 and Supplementary Material Table S2). These analyses delineate a total of 10 genera and 18 species in Sainouroidea, resulting in 5 novel genus-level clades and 12 novel species-level designations (Fig. 3). Genus and species-level sequence similarity seen in the uncorrected pairwise distance matrix corresponded directly to fully or highly supported clades in both the BI and ML analyses (Fig. 3) and are congruent with the tree topology in our taxon-rich data set (Fig. 2). The same genus clades seen in Figure 2 and Figure 3 were recovered in the reduced dataset, which contained only the V5 region of the 18S rDNA (V5 + eukaryotic equivalent to the V6 region)

(Supplementary Material Fig. S1). However, this shortened dataset did not have enough sites to recover the same species clades seen in Figures 2 and 3 (Supplementary Material Fig. S1). These results lead us to describe 4 novel genera: *Olivorum* n. gen., *Puppisaman* n. gen., *Homocognata* n. gen., and *Acantholus* n. gen. (see Taxonomy Summary section) and the recognition of a fifth genus-level clade from a presumed endobiont detected in the genome sequencing efforts of the western tarnished plant bug, *Ligus hesperus* (Pánek et al. 2017, GenBank KY201455). These results also lead us to describe 12 novel species: *R. liberus* n. sp., *R. incognitus* n. sp., *R. piscicus* n. sp., *R. vulgaris* n. sp., *R. philanguis* n. sp., *R. hawesi* n. sp., *G. erdosi* n. sp., *G. rogosa* n. sp., *O. cimiterus* n. sp., *P. gallanis* n. sp., *H. vulgaris* n. sp., and *A. ambiguus* n. sp. (see Taxonomy Summary section).

To further analyze the species delineations within a genus, a ML phylogeny and corresponding uncorrected pairwise distance matrix were created for the genera represented by more than one species (i.e., *Rosculus* and *Guttulinopsis*) (Supplementary Material Figs S2, S3). Reducing the alignments to contain only full or nearly full length (i.e. greater than 1500 bp) intra-genus sequences resulted in longer masked alignments (*Rosculus*: 1830 bp and *Guttulinopsis*: 2080 bp) and more resolution. All *Guttulinopsis* species were monophyletic and had a sequence difference of less than 1.7% as seen in Figure 3 (Supplementary Material Fig. S3). All *Rosculus* species, except *R. hawesi* n. sp., were monophyletic and had a sequence difference of less than 1.7% as seen in Figure 3 (Supplementary Material Fig. S2). In Figure 3, using a difference of less than 1.7%, *R. hawesi* is represented by the three isolates: C1C, MSUPP161R, and CSA. However, when only *Rosculus* sequences are used in the alignment CSA branches sister to C1C and MSUPP161R and has a sequence difference greater than 1.7% (Supplementary Material Fig. S2). These results lead us to only identify CSA as *Rosculus* sp., leaving a monophyletic *R. hawesi* represented by C1C and MSUPP161R (Fig. 2, Supplementary Material Fig. S2). A phylogeny and uncorrected pairwise distance matrix of *Homocognata* was not created because the polytomy at the base of the genus would make rooting the tree problematic (Figs 2, 3). Also, genera only represented by three sequences or less were not further analyzed (Fig. 2). We tentatively identify species,

other than the type species, with *cf.* in genera that have not been analyzed in detail (Supplementary Material Table S1).

Light Microscopy

The morphology of each strain was studied in detail under light microscopy. The length, width, nucleus and nucleolus diameters of trophic amoebae and (where present) cyst diameter from each strain were measured. The averages and standard deviations (SD) of these measurements for each strain are listed in Table S1. The average length of trophic amoebae among strains ranged from 6.36 μm to 15.53 μm . The largest amoeba strain was a *Guttulinopsis* sp. with an average length of 15.5 μm (SD= 2.5) and an average width of 9.7 μm (SD= 1.7) (STA Supplementary Material Table S1). The smallest amoeba strain was a *Rosculus erdosi* n. sp. with an average length of 6.1 μm (SD= 1.1) and an average width of 4.2 μm (SD= 0.7) (C1C Supplementary Material Table S1).

A principal component analysis (PCA) of trophic amoeba dimensions showing each genus grouped by the measurements of cell length, cell width, nucleus, and nucleolus is presented in Figure 4. At least 30 amoebae from each culture were measured for the PCA analysis. Cyst diameter was not included in the PCA analysis because some strains did not form or have lost the ability to form cysts in culture. The cultures ATCC 50577 (*Rosculus incognitus*), *R. terrestris*, *R. elongata*, T1 (*Guttulinopsis rogosa*), *Guttulinopsis vulgaris* KU738571, GvTice (*Guttulinopsis vulgaris*), and RA (*Puppisaman* sp.) were lost before the measurements were taken and are not included in the PCA analysis (Supplementary Material Table S1). The results of the PCA illustrate that there are no clearly separated groups, based on these measurements (Fig. 4). Thus, these taxa overlap in their form and size making a morphometric identification of amoebae infeasible.

Although measurements alone are not sufficient for the identification of these sainouroid amoebae, some qualitative morphological characteristics can be seen by careful observation of these strains. In general, trophic amoebae assigned to *Acantholus* n. gen. and *Homocognata* n. gen. share more qualitative traits with each other, while trophic amoebae in the *Puppisaman* n. gen., *Olivorum* n. gen.,

Guttulinopsis, and *Rosculus* clade share more qualitative traits with one another. *Acantholus* n. gen. and *Homocognata* n. gen. amoebae multiply slowly compared to other sainouroid amoebae, which multiply and consume their food source rapidly. The pseudopodia of *Acantholus* n. gen. and *Homocognata* n. gen. amoebae usually consist of tongue-like extensions that occasionally possess filament-like sub-pseudopodia that resemble acanthopodia (Tice et al. 2016). *Acantholus* n. gen. and *Homocognata* n. gen. amoebae usually have multiple contractile vacuoles and often move in a gliding fashion (Supplementary Material Video S1). Pseudopodia in the strains assigned to the genera *Puppisaman* n. gen., *Olivorum* n. gen., *Guttulinopsis*, and *Rosculus* commonly move in a rippling or wave-like fashion (Supplementary Material Video S1). Additionally, the hyaloplasm of *Acantholus* n. gen. and *Homocognata* n. gen. appears to be thicker and more opaque than in *Puppisaman* n. gen., *Olivorum* n. gen., *Guttulinopsis*, and *Rosculus* (Fig. 1). The granuloplasm of *Rosculus* amoebae tends to consist of small particles, which often result in a sandy appearance (Fig. 1 CC-UU). The trophic amoebae in the *Guttulinopsis* clade are generally larger than amoebae in the *Rosculus* clade with a dense granuloplasm made of many distinct spheres (Fig. 1 T, Supplementary Material Table S1). *Guttulinopsis* amoebae often exhibit a locomotion that consists of extending a single broad pseudopodium forward, causing the rest of cell body to be pulled forwards until the elongated cell shape is perpendicular to the direction of the original pseudopod (Supplementary Material Video S1). However, the most distinguishing character among all genera was that sorocarpic fruiting was only seen within the *Guttulinopsis* clade (Fig. 1 AA and BB).

Transmission Electron Microscopy

The ultrastructure of a representative from each newly described genus was studied. The ultrastructure of *Guttulinopsis vulgaris* was previously studied in detail (Erdos and Raper 1978), therefore we did not perform any further TEM analyses on *Guttulinopsis*. Here we studied the ultrastructure of the five sainouroid amoeba genera: *Acantholus* n. gen. (ATCC 50888), *Rosculus* (RSA), *Olivorum* n. gen. (UACEM), *Homocognata* n. gen. (EuroGSA), and *Puppisaman* n. gen. (CP16-1) (Fig. 5). The ultrastructure of each strain was very similar to that previously described for *Guttulinopsis* (Erdos and

Raper 1978). The ultrastructure similarities include: no observed MTOC (Microtubular Organizing Center), multiple mitochondria of variable size, endoplasmic reticulum that is vesicular and lamellate throughout the cytoplasm, lipid inclusions, and food vacuoles (Erdos and Raper 1978). The only clear difference seen among genera was in the morphology of mitochondrial cristae. The genus *Acantholus* n. gen. had mitochondria with tubular cristae (Fig. 5C). Both tubular mitochondrial cristae and somewhat flat mitochondrial cristae were seen in *Homocognata* n. gen. (Fig. 5A, B). The genera *Puppisaman* n. gen., *Olivorum* n. gen., and *Rosculus* contained mitochondria with flat mitochondrial cristae (Fig. 5 D, E, F) as does *Guttulinopsis* (Erdos and Raper 1978).

Discussion

We used culturing techniques in association with morphological and molecular phylogenetic analyses to reveal the genus and species-level diversity within Sainouroidea. The highly divergent nature of sainouroid 18S rDNA sequences has made the placement of Sainouroidea within Monadofilosa difficult to determine (Bass et al. 2016; Cavalier-Smith et al. 2008, 2009). Even with our greatly expanded sainouroid 18S rDNA taxon sampling, the group remains relatively unresolved within the ‘core-cercozoa’. Ultrastructural comparison has strengthened Sainouroidea’s relationship with core-cercozoans (Cavalier-Smith et al. 2008, 2009), but to fully resolve the position of Sainouroidea in Rhizaria a multigene phylogenomic approach is needed and is the focus of another project currently in progress. In previous work, *Cholamonas* branched outside of the Sainouroidea clade with no support (Bass et al. 2016); however, our phylogeny shows a highly supported monophyletic Sainouroidea with *Cholamonas* basally branching within the clade (Fig. 2). The combination of this phylogenetic support and previously published ultrastructural evidence strongly suggests that *Cholamonas* is within a monophyletic Sainouroidea (Fig. 2; Cavalier-Smith et al. 2008, 2009).

Genus and Species Delineation

As a method to reduce bias during classification of these organisms, an uncorrected pairwise distance matrix was used in conjunction with an 18S rDNA molecular phylogeny (Fig. 3). Clear boundaries in the matrix are seen when the genus-level similarity was set to less than 12% and this corresponded directly to 10 distinct and highly supported groups in the tree (Fig. 3). This was also true when the species-level sequence difference was set to less than 1.7% (Fig. 3). However, assigning the species-level variation from these data was slightly more problematic because nine of the ten genera were represented by a single species and some of the known taxonomic diversity, comprising partial 18S rDNA sequences, were excluded from our taxonomic assignment analyses. As a conservative approach, excluding type isolates, we tentatively identified species in genera represented by a single species with cf. to ensure that species remain monophyletic as more isolates are discovered and described.

Morphometric criteria alone are not sufficient to discriminate many of the amoeboid sainouroid taxa from one another (Fig. 4). A molecular phylogenetic species concept based on a congruence of sequence divergence and monophyletic groups is most appropriate for Sainouroidea given the high level of molecular variation, but results in a high degree of morphological similarity among species (Figs 1, 3). We understand that as more sainouroid amoebae are discovered and have their 18S rDNA gene sequenced the topology of the Sainouroidea tree will change and the uncorrected pairwise distance matrix may also change. The original sequence difference values of less than 12% for a genus and less than 1.7% for a species may not always result in monophyletic groups, but this is a viable working taxonomic scheme of generic and specific delineations to which future data can be appended. The problem of a non-monophyletic species was originally seen in *R. hawesi* n. sp. In the Sainouroidea distance matrix, CSA had a sequence difference of less than 1.7% from C1C and MSUPP161R (Fig. 3 and Supplementary Material Table S2). However, the previously described *R. elongata* and *R. terrestris*, which were excluded from the species delineation analysis as they are partial sequences, branched within this clade (Fig. 2). To solve this problem, we inferred an intra-genus ML phylogeny and uncorrected pairwise distance matrix (using aligned and trimmed sequences) of *Rosculus* (Supplementary Material Fig. S2).

By increasing the number of sites in the alignment we found that CSA is sister to *R. hawesi* with a sequence difference of greater than 1.7%, while all other *Rosculus* species remained the same (Supplementary Material Fig. S2). We chose to not describe CSA as a new species because of the ambiguity surrounding the partial sequences of *R. elongata* and *R. terrestris*. Even after further intra-genus analysis, future cases could arise in which genus or species cutoffs are not consistent with the phylogeny; in such cases, we recommend that the monophyly of the genus or species take precedence over the species sequence difference cutoff.

Assignment of *Guttulinopsis*

A highly supported genus-level clade contains the type species of *Guttulinopsis*. The evolution of multicellularity in Sainouroidea appears to be restricted to this genus (Supplementary Material Table S1 and Fig. 3). In our study, all strains isolated from sorocarpic fruiting bodies found on cow feces were assigned to the genus *Guttulinopsis*. We identified 3 species of *Guttulinopsis* by molecular criteria (Figs 2, 3, Supplementary Material Fig. S3) that morphology appears to corroborate. Traditionally *Guttulinopsis* species have been classified by the morphology of the fruiting body (Olive 1901; Raper et al. 1977). The *Guttulinopsis* strain GvTice was isolated from a typical *G. vulgaris* fruiting body, identical in morphology to the fruiting body of *G. vulgaris* KU738571 (Fig. 1, AA). GvTice and *G. vulgaris* KU738571 form a clade with a less than 1.7% sequence difference that can confidently be classified as *G. vulgaris* (Fig. 3). The *Guttulinopsis* strain FoldedA was isolated from a fruiting body that differed from the rounded shape of a typical *G. vulgaris* fruiting body and all other described species of *Guttulinopsis* (Olive 1902; Raper et al. 1977, Fig. 1, BB). The fruiting body of FoldedA appeared thinner and had a sorus with a wrinkled or “folded” appearance unlike that of previously described *Guttulinopsis* species (Fig. 1, BB). FoldedA formed a species-level clade with the isolates ATCC 50030, and T1, which were originally isolated as amoebae and not observed to fruit (Fig. 3). It is worth noting that fruiting is not observed in any *Guttulinopsis* strains after the culture was grown without sterile cow dung and fruiting has yet to be recovered in any strains. The FoldedA clade was classified as *G. rogosa* nov. sp. GS4C is

the third *Guttulinopsis* species recognized in our analysis and was classified as *G. erdosi* nov. sp. (Fig. 3). This species was originally isolated as amoebae and has not been observed to fruit; however, it can be distinguished by its significantly smaller average length and width (8.0 μm length SD= 2.0/ 4.8 μm width SD= 1.0) compared to other *Guttulinopsis* amoebae (10-20 μm length / 8–12 μm width) (Raper et al. 1977 and Supplementary Material Table S1). Interestingly *G. erdosi* marks the first verified occurrence of a *Guttulinopsis* amoeba isolated from a host living in an aquatic environment, suggesting that this genus may be more widespread than previously thought or perhaps indicating a rather flexible life-history/life-cycle (to be discussed in more depth later).

Assignment of *Rosculus*

Determining which genus-level clade to designate as *Rosculus* proved more difficult than anticipated. Utilizing our molecular species concept criteria, four cultures accessioned as ‘*Rosculus*’ in culture collections, (ATCC 50030, ATCC 50888, ATCC 50577 and Culture Collection of Algae and Protozoa (CCAP) 1571/3) belong to three separate sarracoid genera (Figs 2, 3). Our analyses indicate that ATCC 50030 is a genuine *Guttulinopsis* species (*G. rogosa*; as discussed above) and partial 18S rDNA sequence unequivocally identified the remaining accessioned ‘*Rosculus*’, ATCC PRA-135, as a *Micriamoeba* in another supergroup altogether, Amoebozoa (SSU rDNA accessioned here as MH643883; also examined phylogenomically by Kang et al. 2017). To help guide our taxonomic assignments, we closely compared morphologies and life-history characteristics observed within the other relevant genus-level groups to the original description of the type species of *Rosculus*, *R. ithacus*.

We attempted to re-isolate *R. ithacus* from its type host species and isolated amoeba strain EuroGSA from *N. natrix* captured from the same locale studied by Hawes (1955). Instead of possessing a broad, thin, fan-shaped pseudopodium that moves in a rippling manner as described for *R. ithacus* (Hawes 1963), EuroGSA amoebae have a motility that usually involves multiple thicker pseudopodia moving in a “tongue-like” fashion (Figure 1, N). The EuroGSA isolate was found to contain both tubular and flat mitochondrial cristae (Fig. 5A, B). These differences in morphology and mitochondrial cristae led us to

reject the designation of *Rosculus* to the clade with EuroGSA and instead assign it to the new genus *Homocognata* n. gen. Likewise, ATCC 50888 was not designated as *Rosculus* because of its uncharacteristic morphology. It possesses tubular mitochondrial cristae (Fig. 5), while *Rosculus sensu* Page has flat mitochondrial cristae (Page 1988). ATCC 50888 has pseudopodia that are commonly filose as opposed to the rounded pseudopodia described in *Rosculus* by Hawes (1963) and therefore was re-assigned to the new genus *Acantholus* n. gen.

On the other hand, the morphology of motile amoebae of strains CCAP 1571/3, ATCC 50577 and all the other amoeba isolates comprising this genus-level clade (Figs 2, 3) are very similar to the original description of *Rosculus ithacus* (Hawes 1963). Isolate RSA (in the same genus-level clade as CCAP 1571/3 and ATCC 50577) was found to have flat mitochondrial cristae as described in *Rosculus* (Page 1988). However, based on internal cell morphology, locomotion and placement in the phylogenetic tree, either the strain UACEM or the clade containing CP16-1 are equally viable options for designation as the genus *Rosculus* (Figs 1, 2, Supplementary Material Video S1). Yet, we prefer to assign the genus to the latter clade because it better maintains continuity with other published amoeba isolates placed into *Rosculus* after Hawes' original description (Figs 1, 2; Bass et al. 2016). Furthermore, this taxonomic assignment for the genus *Rosculus* is consistent with the 18S rDNA sequences previously designated as *Rosculus* (Bass et al. 2016; Tysl and Dykova 2018) and for the moment, alleviates confusion that may arise from unwarranted or insufficiently justified taxonomic name changes. Hence the genus-level classification of *Rosculus* was assigned to the clade containing RSA, CCAP 1571/3 and ATCC 50577 (Fig. 3). Here UACEM was assigned to the new genus *Olivorum* n. gen. and the clade containing CP1-16 was assigned to the new genus *Puppisaman* n. gen (Fig. 3).

It is unclear if the type species of *Rosculus*, *R. ithacus*, has been re-isolated and investigated since the original isolate of Hawes (Hawes 1955, 1963). Even though we cultured *Rosculus* strains from North American relatives of *N. natrix* (Figs 2,3, Supplementary Material Table S1), we are hesitant to assign them to *R. ithacus*. Hawes (1963) noted that no contractile vacuoles were observed in *R. ithacus*, while all of our *Rosculus* isolates possess them. Page (1974) explicitly noted the presence of contractile vacuoles

in his soil isolate of *Rosculus* and felt at the time that this character was not “sufficient ground” to warrant a different taxonomic assignment. We now have multiple *Rosculus* strains representing multiple species to compare, and this morphological character may be of taxonomic significance. Therefore, we prefer to reserve the taxonomic designation of *R. ithacus* to a species-level molecular clade that is represented by an amoeba isolate that fits the morphological description of the type species isolated from the type host. Thus, we reclassified CCAP 1571/3 (accessioned as ‘*Rosculus ithacus*’) to *Rosculus liberus* nov. sp. because it possesses contractile vacuoles unlike that of the type species and was isolated from soil and not the type host of *R. ithacus*, the European grass snake (*N. natrix*) (Hawes 1963).

An area of confusion arose while sequencing the 18S rDNA from the CCA culture. PCR, cloning, and Sanger sequencing of 18S rDNA shortly after a mono-amoeba culture was established revealed two variant 18S rDNAs represented by the 18S rDNA clones labeled CCAC1 (clone 1) and CCAC3 (clone 3) (Table S1). Both CCAC1 and CCAC3 sequences branched within the *Rosculus* clade, however they were divergent enough (> 1.7%) to be classified as two separate species (Figs 2, 3). There are two possible explanations for this heterogeneity; 1) the original culture was not clonal and the 18S rRNA genes belong to different species or 2) a mono-typic amoeba has two divergent classes of 18S rRNA genes in its genome. Intragenomic polymorphism of the 18S rDNA has been discovered in other eukaryotic lineages (Gunderson et al. 1987; Weber and Pawlowski 2014), but not in any of the other sariouroid taxa examined to date. To resolve this problem, a transcriptome was sequenced from the mass culture of CCA after multiple passages (Supplementary Material Table S1). With CCAC1 and CCAC3 as query sequences and the CCA transcriptome as a reference we used BLASTn to retrieve all assembled contigs of 18S rRNA (data not shown). The BLASTn output contained a full length 18S rRNA contig along with a few partial length 18S rRNA contigs that were not fully assembled. Each contig had a higher percent identity to CCAC1 than to CCAC3 (data not shown). CCA (the full length assembled 18S rDNA taken from the camel cricket amoeba transcriptome) was included in this analysis and it branched very closely with CCAC1 (Figs 2, 3). The uncorrected pairwise distance matrix showed a sequence difference of 0.2% between the transcriptome generated 18S rRNA gene sequence (CCA) and CCAC1

18S rDNA, while there was a 9.9% sequence difference between CCA and CCAC3 (Fig. 3). Since the CCAC3 18S rRNA gene sequence was not found in the transcriptome, it is most likely that there was not intragenomic polymorphism, but rather two species of *Rosculus* harbored in the same camel cricket and that after multiple rounds of passages the *Rosculus* species containing the CCAC3 rRNA gene was lost from the culture. Since this amoeba strain was lost without detailed morphological observations, we will not assign a species name to the CCAC3 rDNA sequence, but will await analyses of future isolates that form a species-level clade with this sequence.

Character Traits of Sainouroidea

By mapping character traits to the tree of Sainouroidea we can begin to speculate about evolution in the group as a whole (Fig. 2). Perhaps the most unique character trait is the sorocarpic multicellularity that independently evolved within *Guttulinopsis* (Fig. 2). This represents the only incidence of sorocarpic multicellularity in the supergroup Rhizaria and it appears restricted to the genus *Guttulinopsis* (Fig. 2; Brown et al. 2012a; Seb -Pedr s et al. 2017). We also see at least two separate losses of the flagellar apparatus, one in the *Guttulinopsis*, *Rosculus*, *Puppisaman*, *Olivorum* clade and one in the *Homocognata*, *Acantholus* clade (Fig. 2). This is especially interesting considering that the flagellar apparatus in *Cholamonas*, the basal branching Sainouroid, was the result of a doubling event (Flavin et al. 2000). The changes between tubular and flat mitochondrial cristae in Sainouroidea are unusual for a group that is related and morphologically very similar (Figs 2, 4). A change from tubular mitochondrial cristae to flattened mitochondrial cristae has been observed in a few other protistan lineages, such as *Kraken* in Cercozoa and *Stygamoeba* and *Vermistella* in Amoebozoa (Dumack et al. 2017; Moran et al. 2007; Smirnov 1996). Comparative genomic studies of Sainouroidea will likely reveal evolutionary histories that can help us better understand these character changes in Sainouroidea.

Life History of Sainouroid Amoebae

With the previous knowledge and the findings presented here we can begin to speculate on the life-history of sainouroid amoebae. These amoebae appear to be animal-associated but they can also be found free-living, most often associated with fecal environments (Bass et al. 2016, Supplementary Material Table S1). These bacterivores thrive in aerobic environments and in the microaerophilic gut environment of many animals. Indeed, *Rosculus* cannot only survive, but can thrive in anaerobic culture (Hawes 1955). It is currently unknown if the preferred habitat of sainouroid amoebae is free-living or endozoic or if they are equally at home in both types of environments.

In the only attempt to elucidate the mode of transmission to animal hosts Hawes (1955) attempted to reinfect a small number of *N. natrix* snakes with the original *R. ithacus* isolate. No infections were established by oral inoculation of cysts or trophic amoebae (0 infections, N=3) but direct inoculation into the rectum was successful (3 infections, N= 4) (Hawes 1955). This experimental infection route is analogous to passing a culture into fresh medium and it seems unlikely that all of the hosts that we have now identified are rectally infected in nature. Thus, we strategically sampled the preferred prey items of the plain-bellied water snake (*Nerodia erythrogaster*) and Graham's crawfish snake (*Regina grahamii*) captured from the same locations as the snakes, to assess if they could directly vector infection by ingestion. *Rosculus hawesi* was found in crawfish (*Procambarus* sp.), the prey of Graham's crawfish snake in which was found a closely related *Rosculus* sp (Fig. 2, Supplementary Material Table S1). It is plausible that ingestion of infected crawfish could directly transmit *Rosculus* amoebae to snakes, however it is just as plausible that oral ingestion of cyst can result in the infection of both hosts since *R. hawesi* has also been isolated from pond water (Fig. 2, Supplementary Material Table S1). Different sainouroid amoeba species were found infecting a long-term captive common water snake (*Thamnophis sirtalis* ssp.) and its green sunfish (*Lepomis cyanellus*) food. Different sainouroid amoeba species were likewise found in a field captured plain-bellied water snake (which feeds on amphibians and fish) and green sunfish captured from the same locale (Supplementary Material Table S1).

Through our sampling efforts, we serendipitously uncovered a wealth of hitherto unknown sainouroid amoeba diversity. We observed that some host species could harbor different sainouroid

amoeba genera and species (e.g., *R. piscicus*, *R. vulgaris*, *G. erdosi*, were isolated from green sunfish; *R. hawesi* and *H. vulgaris* were isolated from crawfish; *G. rogosa* and *P. gallanis* were isolated from turkey feces). A single animal may also be infected with multiple amoeba species (*R. vulgaris* and *Rosculus* sp. were isolated from a single camel cricket). A single amoeba species can also infect different animal hosts (e.g., *R. vulgaris* was isolated from green sunfish and a camel cricket; *R. philanguis* was only found in 3 distantly related snake species; *R. hawesi* was found free-living and infecting crawfish and Graham's crawfish snake; *G. rogosa* was found in humans, turkey, tortoise (*Centrochelys sulcata*) and cows; *P. gallanis* was isolated from chickens and turkey; *H. vulgaris* was isolated from crawfish and 3 snake species). Our limited sampling highlights the need for controlled laboratory experiments to tease apart the life-cycle(s) of sainouroid amoebae in conjunction with a greatly expanded sampling of hosts, amoebae and gene sequences to better understand the life history and ecology of these organisms.

Conclusions

In conjunction with gross morphology and locomotive forms, our results show that at this time, positive identification of amoebae within Sainouroidea requires molecular data. Here we have discovered four novel sainouroid genera *Olivorum*, *Puppisaman*, *Homocognata*, and *Acantholus* and erected twelve novel species *R. liberus*, *R. incognitus*, *R. piscicus*, *R. vulgaris*, *R. philanguis*, *R. hawesi*, *G. erdosi*, *G. rogosa*, *O. cimiterus*, *P. gallanis*, *H. vulgaris*, and *A. ambiguus*. This study attempted to provide cultured representatives to the uncharacterized genetic diversity seen previously in environmental sequencing (Supplementary Material Fig. S1) (Bass et al. 2016), but our molecular species concept (which requires full length 18S rRNA sequence) prevents us from confidently assigning species names to any short 18S rRNA amplicon. Yet even with this uncertainty it appears that at least four more lineages of putative genus level within Sainouroidea are without cultured representatives (i. e., Rhizarian exLh KY201455, KU738554_goat_dung18 plus KU738553_cow_sheep_dung17, KU738546_banteng_dung9, and KU738545_banteng_dung8) (Supplementary Material Fig. S1).

Sainouroid amoebae occupy many environments and are often associated with the fecal material of animals. These organisms do not seem to be restricted to any particular group of animals as they have been isolated from the guts or fecal material of mammals (cattle, dogs, horses, humans, mice, pigs, rabbits), birds (turkey, chicken and penguin), reptiles (snakes and turtles), fishes (sunfish), and arthropods (crawfish and crickets) as well as directly from marine, freshwater, and soil environments (Supplementary Material Table S1; Olive 1902). Sainouroid amoebae are fast growing bacterivores and many possess the ability to grow in anaerobic and aerobic conditions (Hawes 1963). Our work supports the suggestion that sainouroid amoebae are amphizoic (Dyková et al. 1996; Page 1974), but more molecular data are needed to understand if the ability to infect animals is specific to particular sainouroid species or not. Regardless, these traits and the apparent ubiquitous nature of these eukaryotes suggest an important and overlooked role in metazoan microbiomes and microbial ecosystems.

Taxonomic Summary

Rhizaria Cavalier-Smith 2002

Cercozoa Cavalier-Smith 1998

Filosa Cavalier-Smith and Chao 2003

Monadofilosa Cavalier-Smith and Chao 2003

Sainouroidea Cavalier-Smith et al. 2009

Taxonomy of Novel Genera and Species

***Rosculus* Hawes, 1963**

Rosculus incognitus Schuler et Brown n. sp.

Diagnosis: Average length in locomotion 8.6 μm , average width in locomotion 6.7 μm , average nucleus diameter 1.7 μm , average nucleolus diameter 1.2 μm , and cyst not observed. Rapidly dividing and

growing in population size. Moves rapidly with broad lobose pseudopodia in hyaline area at the anterior of the motile cell. Primarily a bacterivore.

Type Strain: ATCC 50577. Grown and routinely kept on *E. coli* MG1655 at 20 °C.

Type Location: Unknown

Type Material: The type culture (ATCC 50577) has been deposited with the ATCC under accession 50577. This culture is considered the hapantotype (name-bearing type) of the species (see Art. 73.3 of the International Code for Zoological Nomenclature, 4th Edition).

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488769

ZooBank ID: urn:lsid:zoobank.org:act:F00CB27B-0DAE-4515-B0EE-53643223D400

Etymology: *incognitus*; Latin (incognitus – unknown); masculine

Rosculus liberus Schuler et Brown n. sp.

Diagnosis: Minimally inhabiting soil. Average length in locomotion 9.7 µm, average width in locomotion 6.4 µm, average nucleus diameter 1.8 µm, average nucleolus diameter 0.9 µm, and average cyst diameter 5.8 µm. Moves rapidly with wide hyaline area at the anterior of the motile cell. Primarily a bacterivore.

Type strain: CCAP 1571/3. Grown and routinely kept on *E. coli* MG1655 at 20 °C.

Type location: Isolated from soil in Southern Scotland (55°28'12.0"N 2°13'57.0"W).

Type Material: The type culture (CCAP 1571/3) has been deposited with the CCAP under accession 1571/3. This culture is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number KU738570.

ZooBank ID: urn:lsid:zoobank.org:act:68BFA660-9283-47CC-8769-9C45B1D61D97

Etymology: *liberus*; Latin (liber - free), pertaining to free-living amoeba isolated from soil; masculine

Rosculus piscicus Schuler, Silberman, et Brown n. sp.

Diagnosis: Minimally inhabiting fish guts. Average length in locomotion 6.8 μm , average width in locomotion 4.1 μm , average nucleus diameter 1.5 μm , average nucleolus diameter 0.9 μm , and cyst not observed. Moves rapidly with wide hyaline area at the anterior of the motile cell. Primarily a bacterivore.

Type strain: green Sunfish 5 Creek (GS5C) amoeba. Grown and routinely kept on *E. coli* MG1655 at 20 °C.

Type location: Isolated from the gut material of a green sunfish (*Lepomis cyanellus*) captured in Owl Creek, Fayetteville, AR (36°04'16"N 94°13'53"W).

Type Material: The type culture (GS5C) has been deposited with the CCAP under accession CCAP 1571/6. This culture is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488758.

ZooBank ID: urn:lsid:zoobank.org:act:776CE9EC-5F7A-44F2-83D5-8A6B705034D2

Etymology: *piscicus*; Latin (piscus - fish), pertaining to fish; masculine

Rosculus vulgaris Schuler, Silberman, et Brown n. sp.

Diagnosis: Originally found inhabiting fish guts and camel cricket guts. Average length in locomotion 7.7 μm , average width in locomotion 4.4 μm , average nucleus diameter 1.3 μm , average nucleolus diameter 0.8 μm , and cyst not observed. Moves rapidly with hyaline area at the anterior of the motile cell. Primarily a bacterivore.

Type strain: green sunfish 10 creek amoeba (GS10C). Grown and routinely kept on *E. coli* MG1655 at 20 °C.

Type location: Isolated from the gut content of a green sunfish (*Lepomis cyanellus*) captured in Owl Creek, Fayetteville, AR (36°04'16"N 94°13'53"W).

Type Material: The type culture (GS10C) has been deposited with the CCAP under accession CCAP 1571/7. This culture is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488757.

ZooBank ID: urn:lsid:zoobank.org:act:4F2645F1-2F40-4008-A545-F7B898C2A957

Etymology: *vulgaris*; Latin (vulgus – common); masculine

Rosculus philanguis Schuler, Silberman, et Brown n. sp.

Diagnosis: Originally found inhabiting snake dung. Average length in locomotion 9.7 μm , average width in locomotion 6.1 μm , average nucleus diameter 1.5 μm , average nucleolus diameter 1.0 μm , and average cyst diameter 4.6 μm . Moves rapidly with broad hyaline area at the anterior of the motile cell. Primarily a bacterivore.

Type strain: Ratsnake amoeba (RSA). Grown and routinely kept on *E. coli* MG1655 at 20 °C. The type culture (RSA) has been deposited with the CCAP under accession CCAP 1571/5.

Type location: Isolated from the feces of a western ratsnake (*Pantherophis obsoletus*) captured in Fayetteville, AR (35°08'00"N 92°22'00"W).

Type Material: A fixed and embedded resin TEM block of the type isolate RSA was deposited in the Smithsonian Museum under accession USNM 1493890. This permanent physical specimen is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488767.

ZooBank ID: urn:lsid:zoobank.org:act:4F2522C1-E1AD-4FFC-968E-D7A9D93E7555

Etymology: *philanguis*; mixed etymological origin, Greek (phila – love) Latin (anguis - snake), love of snakes; masculine

Rosculus hawesi Schuler, Tice, et Brown n. sp.

Diagnosis: Minimally inhabiting snake dung, a crawfish carapace, and free-living in fresh water. Average length in locomotion 9.1 μm , average width in locomotion 5.7 μm , average nucleus diameter 1.6 μm , average nucleolus diameter 1.0 μm , and average cyst diameter 5.1 μm . Moves rapidly with lobose pseudopodia in hyaline area at the anterior of the motile cell. Primarily a bacterivore.

Type strain: MSUPP16R. Grown and routinely kept on *E. coli* MG1655 at 20 °C.

Type location: isolated from water in a small ephemeral puddle in a parking lot on the Mississippi State University campus in Starkville, MS (33°27'39.05"N 88°47'10.45"W).

Type Material: The type culture (MSUPP16R) has been deposited with the CCAP under accession CCAP 1571/4. This culture is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488764.

ZooBank ID: urn:lsid:zoobank.org:act:EF185703-E63F-4120-8AA2-5463D23E174C

Etymology: *hawesi*; named after R. S. J. Hawes, the describer of *Rosculus ithacus*; masculine

Guttulinopsis Olive, 1901

Guttulinopsis erdosi Schuler, Silberman, et Brown n. sp.

Diagnosis: Minimally inhabiting fish guts. Average length in locomotion 8.0 µm, average width in locomotion 4.8 µm, average nucleus diameter 1.4 µm, average nucleolus diameter 0.9 µm, and cyst not observed. Motile amoeba smaller in dimension than other species of *Guttulinopsis*. Moves rapidly with hyaline area at the anterior of the motile cell. Primarily a bacterivore.

Type Strain: GS4C. Grown and routinely kept on *E. coli* MG1655 at 20 °C.

Type location: Isolated from the gut content of a green sunfish (*Lepomis cyanellus*) captured in Owl Creek, Fayetteville, AR (36°04'16"N 94°13'53"W).

Type Material: The type culture (GS4C) has been deposited with the CCAP under accession CCAP 1926/1. This culture is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488770.

ZooBank ID: urn:lsid:zoobank.org:act:ED07311F-2E21-457C-81FF-359B10198971

Etymology: *erdosi*; named after G. Erdos, an early researcher of *Guttulinopsis* ultrastructure; feminine

Guttulinopsis rogosa Schuler, Tice, Silberman, et Brown n. sp.

Diagnosis: Minimally inhabiting cow dung, turkey dung, and a human nasal cavity. Fruiting body morphology is atypical of the genus. Fruiting bodies found on cow dung are approximately 150 µm tall with a stalk characteristic of *G. vulgaris*, but the sorus is oval with vertical folds (compared to smooth sphere sorus characteristic of genus). Average length in locomotion 13.4 µm, average width in locomotion 8.8 µm, average nucleus diameter 2.5 µm, average nucleolus diameter 1.7 µm, and average cyst diameter of 7.6 µm. Moves rapidly with hyaline area at the anterior of the motile cell. Primarily a bacterivore.

Type Strain: FoldedA. Grown and routinely kept on *E. coli* MG1655 at 20 °C.

Type location: Isolated from a fruiting body on the feces of a cow (*Bos taurus*) collected from Maginot Farm, Winslow, AR.

Type Material: The type culture (FoldedA) has been deposited with the CCAP under accession CCAP 1926/2. This culture is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488774.

ZooBank ID: urn:lsid:zoobank.org:act:43F1096A-DBE4-4DB1-AA1B-B254F8D56E4D

Etymology: *rogosa*; Lain (ruga - wrinkle), for the wrinkled fruiting body shape; feminine

Olivorum Schuler, Tice, Silberman, et Brown n. gen.

Diagnosis. Small amoeba with rounded granuloplasm usually with one or two hyaline anterior lobose pseudopodia. Locomotive trophozoites lack distinct uroid. Mitochondria with flat cristae. Cysts round or oval.

Type Species. *Olivorum cimiterus* Schuler, Tice, Silberman, et Brown n. sp.

ZooBank ID: urn:lsid:zoobank.org:act:547A48C0-94B1-491C-A281-C37A90273519

Etymology: *Olivorum*; named after E. W. Olive, the discoverer of *Guttulinopsis*; masculine

Olivorum cimiterus Schuler, Tice, Silberman, et Brown n. sp.

Diagnosis: Minimally inhabiting soil. Average length in locomotion 10.9 μm , average width in locomotion 6.4 μm , average nucleus diameter 1.8 μm , average nucleolus diameter 0.9 μm , and average cyst diameter 5.8 μm . Moves rapidly with usually one to two lobose pseudopodia in the fan-shaped hyaline area at the anterior of motile cells. Has flat mitochondrial cristae. Primarily a bacterivore.

Type strain: UACEM. Grown and routinely kept on *E. coli* MG1655 at 20°C. The type culture (UACEM) has been deposited with the CCAP under accession CCAP 1942/1.

Type location: isolated from grassy soil in Fayetteville, AR (36°03'49.0"N 94°10'10.0"W).

Type Material: A fixed and embedded resin TEM block of the type isolate UACEM was deposited in the Smithsonian Museum under accession USNM 1493891. This permanent physical specimen is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488783.

ZooBank ID: urn:lsid:zoobank.org:act:868B09A5-A9AD-41A0-AB3F-AB4D865F3FFF

Etymology: *cimiterus*; Greek origin of the word cemetery, as it was found at the Evergreen cemetery abutting the University of Arkansas Fayetteville campus; masculine

Puppisaman Schuler et Brown n. gen.

Diagnosis. Small oval shaped amoeba usually seen with one broad lobose pseudopod. Locomotive trophozoites lack distinct uroid. Mitochondria with flat cristae. No cyst observed.

Type Species. *Puppisaman gallanis* Schuler et Brown n. sp.

ZooBank ID: urn:lsid:zoobank.org:act:03EEAA12-A23B-4E85-94E6-B87E821C1D85

Etymology: *Puppisaman*; Arbitrary collection of letters, but with the first part derived from the Latin "puppis" (stern of a ship, later the etymological origin of the naval term "poop deck") referring to the infantile informal English word for feces; neuter

Puppisaman gallanis Schuler et Brown n. sp.

Diagnosis: Minimally inhabiting bird dung and rabbit dung. Average length in locomotion 9.6 μm , average width in locomotion 6.1 μm , average nucleus diameter 1.8 μm , average nucleolus diameter 1.0 μm , and cysts not observed. Moves rapidly with usually one to two lobose pseudopodia in the anterior hyaline area of motile cells. Has flat mitochondrial cristae. Primarily a bacterivore.

Type strain: Chicken Poo 1 (CP16-1). Grown and routinely kept on *E. coli* Mg1655 at 20 °C.

Type location: Isolated from chicken dung in Starkville, MS (33°29'40.7"N 88°45'10.3"W).

Type Material: A fixed and embedded resin TEM block of the type isolate CP16-1 was deposited in the Smithsonian Museum under accession USNM 1493892. This permanent physical specimen is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488785.

ZooBank ID: urn:lsid:zoobank.org:act:6FAC653F-6834-437F-A6B7-8772E0EFBC6C

Etymology: *gallanis*; named after the scientific name of chicken (*Gallus gallus*); masculine

Homocognata Schuler, Silberman, et Brown n. gen.

Diagnosis. Small fast-moving round amoeba with tongue-like lobose pseudopodia. Locomotive trophozoites lack distinct uroid. Commonly seen with multiple contractile vacuoles. Mitochondria with flat and tubular cristae. Cysts round or oval.

Type Species. *Homocognata vulgaris* Schuler, Silberman, et Brown n. sp.

ZooBank ID: urn:lsid:zoobank.org:act:534A84D2-D7D0-4A9F-A5AD-8BC5FCF8F8D8

Etymology: *Homocognata*; Latin (homo - same | cognatus- kinsman); feminine

Homocognata vulgaris Schuler, Silberman, et Brown n. sp.

Diagnosis: Minimally inhabiting snake dung and crawfish guts. Possesses both tubular and flat mitochondrial cristae. Average length in locomotion 12.2 μm , average width in locomotion 8.1 μm , average nucleus diameter 2.4 μm , average nucleolus diameter 1.6 μm , and average cyst diameter of 4.9 μm . Generally multiplies and consumes food source slower than other amoebae in Sainouroidea. Moves

by using multiple small rounded pseudopodia. Generally, has multiple contractile vacuoles. Primarily a bacterivore.

Type strain: European grass snake (EuroGSA). Grown and routinely kept on *E. coli* Mg1655 at 20 °C.

Type location: Isolated from the feces of European grass snake (*Natrix natrix*) captured in Wallingford, England (50°43'33.63"N 2°09'25.35"W).

Type Material: A fixed and embedded resin TEM block of the type isolate EuroGSA was deposited in the Smithsonian Museum under accession USNM 1493894. This permanent physical specimen is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488779.

ZooBank ID: urn:lsid:zoobank.org:act:EB2B359F-1328-4092-9B29-A4E9B33B7005

Etymology: *vulgaris*; Latin (vulgus – common); masculine

Acantholus Schuler et Brown n. gen.

Diagnosis. Small fast-moving limax shaped amoeba with spiny or smooth lobose pseudopodia. Locomotive trophozoites often glide and lack a distinct uroid. Mitochondria with tubular cristae. Cyst round or oval.

Type Species. *Acantholus ambiguus* Schuler et Brown n. sp.

ZooBank ID: urn:lsid:zoobank.org:act:0AA882E9-CFFD-44F5-B156-1B52D349751B

Etymology: *Acantholus*; Arbitrary collection of letters, but with first part Latin (acanthi - thorn), after the spiny pseudopodia of ATCC 50888; masculine.

Acantholus ambiguus Schuler et Brown n. sp.

Diagnosis: Minimally inhabiting snake dung and marine sediment. Tubular mitochondrial cristae. Average length in locomotion 11.7 µm, average width in locomotion 4.7 µm, average nucleus diameter 1.9 µm, average nucleolus diameter 1.0 µm, and average cyst diameter of 4.9 µm. Generally multiplies

and consumes food source slower than other amoebae in Sainouroidea. Can have distinct spiny pseudopods and move in a gliding motion on outstretched pseudopod. Primarily a bacterivore.

Type strain: ATCC 50888. Grown on a saltwater wMY agar plate streaked with *E. coli* Mg1655 at 20 °C.

Type location: Isolated from salt marsh sediment in Hog Island, eastern shore of VA, USA (37°23'45.5"N 75°41'49.2"W).

Type Material: A fixed and embedded resin TEM block of the type isolate ATCC 50888 was deposited in the Smithsonian Museum under accession USNM 1493893. This permanent physical specimen is considered the hapantotype.

Gene Sequence Data: 18S rDNA of type strain: GenBank accession number MH488781.

ZooBank ID: urn:lsid:zoobank.org:act:3A464016-371F-4334-97DA-659507AE2EA6

Etymology: *ambigus*; Latin (ambiguus – doubtful), ambiguous - can have very different morphology within the culture; masculine

Methods

Acquisition, isolation, and maintenance of cultures: All accessioned '*Rosculus*' were purchased from the American Type Culture Collection (ATCC) and Culture Collection of Alga and Protozoa® (CCAP) and initially cultured according to the respective culture collection's instructions. New amoeba strains were isolated from a wide range of substrates that included spores from cellular slime molds fruiting on cow feces, feces collected from animals, soil samples, and gut-contents from dissected animals (Table 1). To assess possible modes of infection we strategically sampled the preferred prey items (crawfish and green sunfish) of the water snake and Graham's crawfish snake, captured from the same locations as the snakes, to assess if they could directly be infected by ingestion of the prey (Supplementary Material Table S1). Collection and handling of snakes and fish were conducted under an approved IACUC protocol and

appropriate Arkansas Game and Fish collection permits (JDW). Feces from European grass snakes were collected in the field in southern England and shipped overnight to Arkansas for processing.

Except where noted, amoebae were isolated by inoculating 0.1 – 0.5gm of substrate into the middle of weak malt yeast agar plates (wMY) (1 liter dH₂O, 0.75 g K₂PO₄, 0.002 g yeast extract, 0.002 g malt extract, 15 g Bacto agar, Spiegel et al. 2005) covered with a thin confluent lawn of either *Escherichia coli*, or an equal mixture of *E. coli*, *Klebsiella pneumonia*, and *Enterobacter aerogenes*. The primary isolation plates were incubated at room temperature (~21 °C) in ambient light and observed daily for evidence of an amoeba feeding front, which usually became evident between 3-7 days after plating. Amoebae were then transferred from the feeding front to wMY plates streaked with *E. coli* (strain MG1655 (ATCC 700926)) for the generation of mono-eukaryotic and clonal cultures.

Fresh feces were collected off the ground into sterile containers from cows, turkeys, rabbit, sheep, and goats. Feces from snakes were obtained by gently squeezing it out (like toothpaste from a tube) onto saran wrap, leaving the snakes physically unharmed. The uric acid fecal cap was avoided and only pieces of formed feces were aseptically inoculated onto primary isolation plates. The hindguts of sacrificed crawfish and green sunfish were aseptically dissected and intestinal contents were plated. An 1cm² piece of carapace from a large crawfish was plated. The contents of the fore- and hindgut of a camel cricket were examined under light microscopy and the foregut contents containing numerous small (~10µm) unidentified bi-flagellates was inoculated into 15ml polypropylene conical tubes containing a solid slant of 3ml inspissated horse serum with a 3ml overlay of ATCC medium 802. The tubes were tightly sealed to maintain a microaerophilic environment along with the native bacteria. The flagellates disappeared from the culture within one week concurrently with the appearance of small amoebae. The amoebae grew slowly through the first 4 weekly passages prior to their rapid expansion and increased growth rate when inoculated onto an aerobic wMY plate streaked with *E. coli*, where the culture was subsequently maintained. The strains FoldedA and GvTice were isolated from spores picked from sorocarpic fruiting bodies on cow dung and cultured on wMY plates streaked with *E. coli* MG1655.

ATCC 50888 and ATCC PRA-134 were cultured on wMY plates made with artificial saltwater (34 ppt, Instant® Ocean Sea Salt) instead of tissue culture flasks with the ATCC-stipulated liquid saltwater 802 medium (ATCC medium 1525). Food preference for most strains was assessed by inoculating amoebae into the middle of a 3-pronged bacterial spoke with each spoke comprising either *E. coli*, *Enterobacter aerogenes* or *Klebsiella pneumoniae*. All isolates preferred *E. coli* and *Enterobacter* over *Klebsiella* and for simplicity all isolates were maintained on wMY plates streaked with *E. coli* MG1655 with serial passage by placing an agar block cut from the feeding front and placing it upside down onto an *E. coli* streak on a new wMY plate. Clonal cultures were created by isolating single amoeba using a sterilized loop and then inoculating it onto a wMY plates streaked with *E. coli* MG1655. The strains used in this study are presented in Table 1.

Light microscopy: Trophic cells were cut from the feeding front of cultures and mounted on slides with PAS (Page's Amoeba Saline) for micrographs (Page 1988). Micrographs were taken at 400x and 1000x magnification using a Zeiss Axioskop2 Plus (Zeiss, Peabody, MA) under DIC (differential interference contrast) optics with a Canon 5DS camera. The measurements of length, breadth, nucleus, nucleolus, and cyst diameter of trophic amoebae from each culture were taken using ImageJ (Schindelin et al. 2012).

Morphology comparison: The morphological measurements of amoeba cultures were analyzed by running a PCA (principal component analysis). Variables included in the analysis were length and breadth of amoebae, as well as nucleus, and nucleolus diameters. The PCA was computed using R v3.2.3 with the *munsell*, *labeling*, and *ggbiplot* packages (R Core Team 2013). The PCA used an $n \geq 30$ for each strain. The ellipses used to group genera in the PCA use a probability of 95%.

Transmission electron microscopy: Amoebae cultures were suspended in liquid wMY, concentrated with 1000 x g centrifugation for 2 minutes forming a pellet, and the liquid wMY was removed. Concentrated cells were fixed using a simultaneous fixation of 25% glutaraldehyde (100 μ L) and 4% OsO₄ (250 μ L) buffered with sodium cacodylate 0.2M pH 7.2 (250 μ L) and liquid wMY media (400 μ L) for 30 minutes on ice. Fixed cells underwent a wash of liquid wMY, 2 washes of ddH₂O, and

then were enrobed in a 2% agarose gel. The gel containing fixed cells was dehydrated in graded washes of EtOH to 100% EtOH, followed by graded washes of acetone and EtOH to 100% acetone, and embedded in Spurr's resin (Spurr 1969). Thin sections (60nm) of the embedded cells were cut with a Diatome® diamond knife (Hatfield, PA) using a Reichert-Jung Ultracut E Ultramicrotome. Thin sections were then collected on formvar coated grids and left to dry overnight. Grids were stained with uranyl acetate for 20 minutes and lead citrate for 7 minutes. Stained sections were viewed using a JEOL 1230 120kV TEM (Institute for Imaging & Analytical Technologies, Starkville, MS).

18S rDNA sequences: The method used to determine the 18S rDNA sequence of each strain is listed in Supplementary Material Table S1 using primers listed in Table 1. For method 1 (PCR) the primers and annealing temperature used for each strain are listed in Supplementary Material Table S1. Each method is described below:

1. DNA extraction was performed using QuickExtract DNA Extraction Solution (Epicentre, Madison, WI) according to the manufacturer's protocol. The 18S rDNA sequences were amplified in a PCR reaction of 1-5µl of the extracted DNA, GoTaq Green Master Mix (Promega) and PCR primers (IDT, Coralville, IA) designed specifically for *Guttulinopsis* and *Rosculus* 18S rDNA (Table 1). The following PCR cycling conditions were used for each isolate: [heated lid at 105 °C] 1) 98 °C for 30 seconds 2) 98 °C for 10 seconds 3) annealing temperature (approximately 40-60 °C, depending on primers, see Table S1) for 45 seconds 4) 72 °C for 3 minutes 5) cycle back to step 2 30x and 6) hold at 4 °C indefinitely. SSU amplicons were visualized and subsequently purified after gel electrophoresis through a 1% agarose Tris-Acetate gel containing SYBR Safe (Life Technologies Corp.). SSU amplicons were cut from the gel and placed in a filter pipet tip inside of 1.5mL Eppendorf tubes. The tubes were centrifuged at 10,000x *g* for 10min and the filter pipet tip then removed. Remaining liquid was dried using a SpeedVac [Savant Refrigerated Condensor and Speed Vacuum]. Samples were sent to the Arizona State University DNA Lab (Tempe, AZ) for

Sanger sequencing. DNA chromatograms of the sequences were assembled and edited using Sequencher v5.4 (GeneCodes, Madison, WI).

Cloning: The CCA and ATCC 50030 isolates had sequence variations within the same isolate that prevented direct sequencing of portions of the amplification product. Therefore, the PCR product was cloned using the PCR Blunt TOPO II Vector cloning kit (Invitrogen) to sequence individual recombinant plasmids.

All cloning reactions were performed following the manufacturer's recommendation except as half-reactions (total of 3.0 μ l) in 1.5ml tubes. The entire ligation reactions were transfected into 25 μ l of chemically competent *E. coli* (DH5a or Top10) on ice for 30-60 min, heat-shocked at 42 °C for 30 sec and the bacterial cells metabolically revived by the addition of 250 μ l and incubation for 1hr at 37 °C with shaking. After the incubation period, the tubes were centrifuged at 8,000 x g for 1 minute to pellet the bacteria, remove half of the liquid followed by plating all the cells onto pre-warmed LB agar plates containing 50 μ g/ml Kanamycin plus 50 μ l of 2% X-gal spread evenly over each plate. The LB plates were incubated overnight at 37 °C to allow the colonies to grow. Twelve white colonies from each ligation were screened by PCR using the M13F/R vector primers to determine if the plasmids contained the correct sized insert following the manufacturer's suggested protocols. All twelve of the CCA colonies screened and seven of the twelve ATCC 50030 colonies contained appropriate sized inserts. These colonies were grown up overnight at 37 °C in 4 ml LB plus 50ug/ml kanamycin (final concentration) and the Zyppy Plasmid Miniprep kit (Zymo Research) was then used to extract the plasmid DNA from the recombinants using the manufacturer's recommended protocol. The recovered plasmid DNA was at a concentration of approximately 100 ng/ μ l.

2. Approximately 20 to 40 amoeba cells were picked using a platinum loop, placed into a cell lysis solution (Picelli et al. 2014), and then underwent multiple rounds of freeze (ethanol -80 °C) thaw (tap water approximately 23 °C). Total RNA was then extracted and double stranded cDNA was constructed using a modified version of Smart-Seq 2 (Kang et al. 2017; Picelli et al. 2014) that is described in detail in Tice et al. 2016. cDNA libraries were sequenced using Illumina HiSeq 2000 or HiSeq 2500 at Genome Quebec. The raw sequences from Illumina sequencing were trimmed and cleaned using Trimmomatic (Bolger et al. 2014). Trimmed RNA sequences were assembled using Trinity de novo assembly. The 18S rDNA was retrieved using BLASTn (NCBI) with a query sequence of a previously sequenced *Guttulinopsis* 18S rDNA and the assembled transcriptome as the database.
3. 2 mL of liquid wMY were poured onto a dense culture of amoebae. Amoebae were scraped off of the agar and collected in a conical tube. Total RNA was extracted from the culture using Trizol reagent (Sigma-Aldrich, St Louis MO) according to the manufacturer's protocol. Total RNA was sent to Novus Genomics Inc. (Philadelphia, PA) for paired-end sequencing on Illumina's HiSeq platform. 18S rDNAs were retrieved from processed transcriptomes as described above.

Phylogenetic analysis: The 18S rDNA sequences of amoebae were aligned with other Rhizarian 18S rDNA sequences and Stramenopile outgroup 18S rDNA sequences, retrieved from the NCBI GenBank database (Release 214), with MAFFT v7 (Katoh and Standley 2013), using MAFFT-LINSI with default parameters. Alignments were manually inspected using AliView v1.17.1 (Larsson 2014), but no manual editing was conducted. Masked alignments were created by removing ambiguously aligned sites in the sequence alignment using BMGE v1.1 (Criscuolo and Gribaldo 2010) using maximum gap rate allowed per character option set to “-g 0.6”. This methodology was used for all phylogenetic analyses and distance matrix calculations. In less inclusive taxon sampling datasets, for example those inferred for sainouroids only (Fig. 3, Supplementary Material Fig. S1) and individual genera

(Supplementary Material Figs S2, S3), alignments and trimming were recalculated due to differing taxon sampling leading to more or less unambiguous sites for analyses. ML trees were inferred using the masked alignment under a GTR + Γ model in RaxML v8.2. BI trees were inferred from the masked alignment in Mr.Bayes v3.2 under the GTR + Γ model. For the BI tree of Rhizaria (Fig. 2), 2 independent runs with 4 chains were run for 14,000,000 generations discarding a 25% burnin (3,500,000) at which the chains had converged. For the BI tree of Sainouroidea (Figure 3), 2 independent runs with 4 chains were run for 10,000,000 generations discarding a 25% burnin (2,500,000) by which the chains had converged. For the BI tree of the V5 region of Sainouroidea 18S rDNA sequences (Supplementary Material Fig. S1), 2 independent runs with 4 chains were run for 20,000,000 generations discarding a 5.8% burnin (1,165,000) by which the chains had converged. ML trees of *Rosculus* and *Guttulinopsis* (Supplementary Material Figs S2, S3) were inferred using a masked alignment under a GTR + Γ model in IQ-TREE and each was rooted at the deepest node seen in Figure 2 (Nguyen et al. 2014). Species and genera were delineated in the sainouroid clade by percent similarity over the masked 18S rDNA alignment (TreeBase: <http://purl.org/phylo/treebase/phyloids/study/TB2:S23022>) calculated in an uncorrected pairwise distance matrix using PAUP v4.0 (Swofford 2003) using option SHOWDIST with the following settings “DSET DISTANCE=P MISSDIST=INFER”.

Acknowledgements

This project was supported in part by the National Science Foundation (NSF) Division of Environmental Biology (DEB) grant 1456054 (<http://www.nsf.gov>), awarded to MWB, Arkansas Biosciences Institute awarded to JDS, Arkansas Undergraduate Research Fellowships awarded to JS, RAP and University of Arkansas Honors awards for partial support of JS, RAP, SG (in the lab of JDS). Mississippi State University's High Performance Computing Collaboratory provided some computational resources. Some data was collected while MWB was a postdoctoral scholar in Prof. Andrew J. Roger's lab in the Centre

for Comparative Genomics and Evolutionary Bioinformatics at Dalhousie University. We thank Prof. Roger for his support on this project. We thank Phil Vogrinc for handling snakes in the Wilson lab, the United Kingdom's Forestry commission for unlimited access to Wareham Forrest and Dr. Gabriela Jofré for aid in collecting European Grass snakes. We thank Eleni Gentekaki at the School of Science, Mae Fah Luang University, Chiang Rai, Thailand for aid in isolation of a strain. Electron microscopy images were acquired from the Institute for Imaging & Analytical Technologies (I²AT), which also provided technical support in particular from Amanda Lawrence.

References

- Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V., Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW** (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol*, **59**:429-514
- Bass D, Cavalier-Smith T** (2004) Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int J Syst Evol Microbiol* **54**:2393-2404
- Bass D, Moreira D, López-García P, Polet S, Chao EE, von der Heyden S, Pawlowski J, Cavalier-Smith T** (2005) Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. *Protist* **156**:149-161
- Bass D, Silberman J. D, Brown MW, Pearce RA, Tice AK, Jousset A, Geisen S, Hartikainen H** (2016) Coprophilic amoebae and flagellates, including *Guttulinopsis*, *Rosculus* and *Helkesimastix*, characterise a divergent and diverse rhizarian radiation and contribute to a large diversity of faecal-associated protists. *Environ Microbiol* **18**:1604-1619
- Bolger AM, Lohse M, Usadel B** (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114-2120
- Brown MW, Silberman JD, Spiegel FW** (2012b) A contemporary evaluation of the acrasids (Acrasidae, Heterolobosea, Excavata). *Europ J Protistol* **48**:103-123
- Brown MW, Kolisko M, Silberman JD, Roger AJ** (2012a) Aggregative multicellularity evolved independently in the eukaryotic supergroup Rhizaria. *Curr Biol* **22**:1123-1127
- Cavalier-Smith T** (1998) A revised six-kingdom system of life. *Biol Rev* **73**:203-266
- Cavalier-Smith T** (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol* **52**:297-354

- Cavalier-Smith T, Chao EEY** (2003) Phylogeny and classification of phylum Cercozoa (Protozoa). *Protist* **154**:341-358
- Cavalier-Smith T, Lewis R, Chao EE, Oates B, Bass D** (2008) Morphology and phylogeny of *Sainouiron acronematica* sp. n. and the ultrastructural unity of Cercozoa. *Protist* **159**:591-620
- Cavalier-Smith T, Lewis R, Chao EE, Oates B, Bass D** (2009) *Helkesimastix marina* n. sp. (Cercozoa: Sainouroidae superfam. n.) a gliding zooflagellate of novel ultrastructure and unusual ciliary behaviour. *Protist* **160**:452-479
- Criscuolo A, Gribaldo S** (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**:210
- Dumack K, Schuster J, Bass D, Bonkowski M** (2016) A novel lineage of 'naked filose amoebae'; *Kraken carinae* gen. nov. sp. nov. (Cercozoa) with a remarkable locomotion by disassembly of its cell body. *Protist* **167**:268-278
- Dykova I, Lom J, Machackova B, Sawyer TK** (1996) Amoebic infections in goldfishes and granulomatous lesions. *Folia Parasitol* **43**:81-90
- Erdos GW, Raper KB** (1978) Ultrastructural aspects of two species of *Guttulinopsis*. *Am J Bot* **65**:552-561
- Gunderson JH, Sogin ML, Wollett GM, Hollingdale M, de la Cruz VF, Waters AP, McCutchan TF** (1987) Structurally distinct, stage-specific ribosomes occur in *Plasmodium*. *Science* **238**:933-937
- Fiore-Donno AM, Rixen C, Rippin M., Glaser K, Samolov E, Karsten U, Becker B, Bonkowski M** (2017) New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput environmental diversity surveys, with emphasis on worldwide biological soil crusts. *Mol Ecol Resour* **18**:229-239

- Flavin M, O'Kelly CJ, Nerad TA, Wilkinson G** (2000) *Cholamonas cyrtodiopsis* gen. n., sp. n. (Cercomonadida), an endocommensal, mycophagous heterotrophic flagellate with doubled kinetid. *Acta Protozool* **39**:51-60
- Hawes RSJ** (1955) A limax-amoeba from the rectum of the grass snake, *Natrix natrix*, as a facultative aerobe *in vitro*. *Nature* **175**:779-780
- Hawes RSJ** (1963) On *Rosculus ithacus* gen. n., sp. n. (Protozoa, Amoebina), with special reference to its mitosis and phylogenetic relations. *J Morphol* **113**:139-149
- Kang S, Tice AK, Spiegel FW, Silberman JD, Pánek T, Čepička I, Kostka M, Kosakyan A, Alcântara DM, Roger AJ, Shadwick LL, Smirnov A, Kudryavstev A, Lahr DJG, Brown MW**(2017) Between a pod and a hard test: the deep evolution of amoebae. *Mol Biol Evol* **34**:2258-2270
- Katoh K, Standley DM** (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**:772-780
- Larsson A** (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**:3276-3278
- Moran DM, Anderson OR, Dennett MR, Caron DA, Gast RJ** (2007) A Description of seven antarctic marine Gymnamoebae including a new subspecies, two new species and a new genus: *Neoparamoeba aestuarina antarctica* n. subsp., *Platyamoeba oblongata* n. sp., *Platyamoeba contorta* n. sp. and *Vermistella antarctica* n. gen. n. sp. *J Eukaryot Microbiol* **54**:169-183
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ** (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**:268-274
- Olive EW** (1901) A preliminary enumeration of the Sorophoreae. *Proc Am Acad Art Sci* **37**:333-344
- Olive EW** (1902) Monograph of the Acrasieae. *Proc Boston Soci Nat History* **30**:451-510
- Page FC** (1974) *Rosculus ithacus* Hawes, 1963, Amoebida, Flabellulidea and the amphizoic tendency in amoebae. *Acta Protozool* **13**:143-154

- Page FC** (1988) A New Key to Freshwater and Soil Gymnamoebae: With Instructions for Culture. Freshwater Biological Association, Ambleside, Cumbria, 122 p
- Pánek T, Žihala D, Sokol M, Derelle R, Klimeš V, Hradilová M, Zadrobílková E, Susko E, Roger AJ, Čepička I, and Eliáš M** (2017) Nuclear genetic codes with a different meaning of the UAG and the UAA codon. *BMC Biology* **15**:8
- Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R** (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nature Protoc* **9**:171
- Raper KB, Worley AC, Kessler D** (1977) Observations on *Guttulinopsis vulgaris* and *Guttulinopsis nivea*. *Mycologia* **69**:1016-1030
- R Core Team** (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- Sandon H** (1924) Some protozoa from the soils and mosses of Spitsbergen. Results of the Oxford University Expedition to Spitsbergen, No. 27. *Zool J Linn Soc* **35**:449-475
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Tinevez J Y** (2012). Fiji: an open-source platform for biological-image analysis. *Nature Meth* **9**:676-682
- Sebé-Pedrós A, Degnan BM, Ruiz-Trillo I** (2017) The origin of Metazoa: a unicellular perspective. *Nature Rev Genet* **18**:498
- Smirnov AV** (1996) *Stygamoeba regulata* n. s p.(Rhizopoda)—a marine amoeba with an unusual combination of light-microscopical and ultrastructural features. *Arch Protistenkd* **146**:299-307
- Spiegel FW, Haskins EF, Cavender JC, Lindley-Settlemyre LA, Edwards SM, Nderitu G, Shadwick JD** (2005) A Beginner's Guide to Isolating and Culturing Eumycetozoans. Online at <http://slimemold.uark.edu/pdfs/isohandbook.pdf> (Consulted 15 Jun 2007)
- Spurr AR** (1969) A low-viscosity epoxy resin embedding medium for electron microscopy. *J Ultrastruct Res* **26**:31-43

Swofford DL (2003) PAUP*: Phylogenetic Analysis Using Parsimony, version 4.0 b10

Tice AK, Shadwick LL, Fiore-Donno AM, Geisen S, Kang ., Schuler GA, Spiegel FW, Wilkinson K,

Bonkowski M, Dumack K, Lahr DJG, Voelcker E, Clauß S, Zhang J, Brown MW (2016)

Expansion of the molecular and morphological diversity of Acanthamoebidae (Centramoebida, Amoebozoa) and identification of a novel life cycle type within the group. *Biol Direct* **11**:69

Tymł T, Dyková I (2018) *Sappinia* sp. (Amoebozoa: Thecamoebida) and *Rosculus* sp. (SAR: Cercozoa)

isolated from King Penguin guano collected in the Subantarctic (South Georgia, Salisbury Plain) and their coexistence in culture. *J Eukaryot Microbiol* **65**:544-555

Weber AAT, Pawlowski J (2014) Wide occurrence of SSU rDNA intragenomic polymorphism in

foraminifera and its implications for molecular species identification. *Protist* **165**:645-661

Woodcock HM, Lapage G (1915) Observations on the life-cycle of a new flagellate, *Helkesimastix*

faecicola, ng, n. sp.: Together with remarks on the question of syngamy in the trypanosomes. *Proc Roy Soc Lond B* **88**:353-370

Visvesvara G, Baxter P, Brandt F, Sawyer T (1982) Isolation of *Rosculus* sp. from a human nose and

demonstration of anti-*Rosculus* antibody in human-sera. *J Protozool* **29**:290

Figure Legends

Figure 1. Micrographs of amoebae strains used in this study. **A)** *Acantholus cf. ambiguus* CGSA (common garter snake amoeba). **B)** *A. cf. ambiguus* EGSA cyst. **C)** *A. ambiguus* ATCC 50888. **D)** *A. ambiguus* ATCC 50888 cysts. **E)** *Olivorum cimiterus* UACEM. **F)** *O. cimiterus* UACEM cysts. **G)** *Puppisaman cf. gallanis* T2 (turkey 2 Amoeba). **H)** *P. cf. gallanis* CP16-1 (chicken poo 1 Amoeba). **I)** *P. cf. gallanis* CP16-2 (chicken poo 2 amoeba) **J)** *Homocognata cf. vulgaris* C1PA (crawfish 1 prairie amoeba). **K)** *H. cf. vulgaris* C1PA cysts. **L)** *H. cf. vulgaris* PBWSA (plain-bellied water snake amoeba) **M)** *H. cf. vulgaris* PBWSA cysts. **N)** *H. vulgaris* EuroGSA (European grass snake amoeba) **O)** *H. vulgaris* EuroGSA cysts. **P)** *H. cf. vulgaris* CGS2A (common garter snake 2 amoeba) unusual morphology **Q)** *H. cf. vulgaris* CGS2A typical amoeba. **R)** *H. cf. vulgaris* CGS1A (common garter snake 1 amoeba). **S)** *H. cf. vulgaris* CGS1A cysts. **T)** *Guttulinopsis rogoza* FoldedA. **U)** *G. rogoza* FoldedA cysts. **V)** *G. erdosi* GS4C (green sunfish 4 creek). **W)** *Guttulinopsis* sp. STA (spurred tortoise amoeba). **X)** *G. vulgaris* GvTice. **Y)** *G. vulgaris* GvTice cyst. **Z)** *G. rogoza* ATCC 50030. **AA)** *G. vulgaris* KU738571 fruiting body on cow dung. **BB)** *G. rogoza* FoldedA fruiting body on cow dung. **CC)** *Rosculus* sp. CSA (crayfish snake amoeba). **DD)** *Rosculus* sp. CSA cyst. **EE)** *R. hawesi* C1C (crawfish 1 creek). **FF)** *R. hawesi* C1C cyst. **GG)** *R. incognitus* ATCC 50777. **HH)** *R. philanguis* CWSA (common water snake amoeba). **II)** *R. philanguis* CWSA cysts. **JJ)** *R. philanguis* RSA (ratsnake amoeba). **KK)** *R. philanguis* RSA cysts **LL)** *R. philanguis* HSA (hognose snake amoeba). **MM)** *R. philanguis* HSA cysts. **NN)** *R. vulgaris* GS7C (green sunfish 7 creek). **OO)** *R. vulgaris* CCA (camel cricket amoeba). **PP)** *R. piscicus* GS5C (green sunfish 5 creek). **QQ)** *R. liberus* CCAP 1571/3. **RR)** *R. liberus* CCAP 1571/3 cyst. **SS)** *R. hawesi* MSUPP16R. **TT)** *R. hawesi* MSUPP16R cyst. **UU)** *R. vulgaris* GS10C (green sunfish 10 creek). A-Z and CC-UU are proportional with the scale-bar = 10 μm in A. AA and BB scale-bars = 200 μm . A-W, Z, and CC-UU were imaged with DIC- microscopy, X and Y were imaged with Phase-Contrast microscopy, and AA and BB were imaged with transmitted light using a dissecting microscopy.

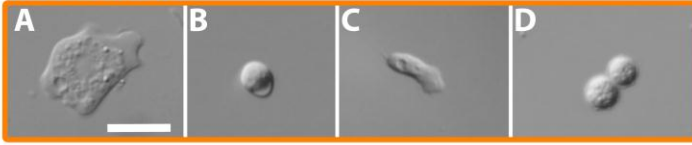
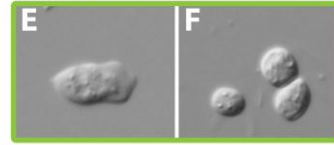
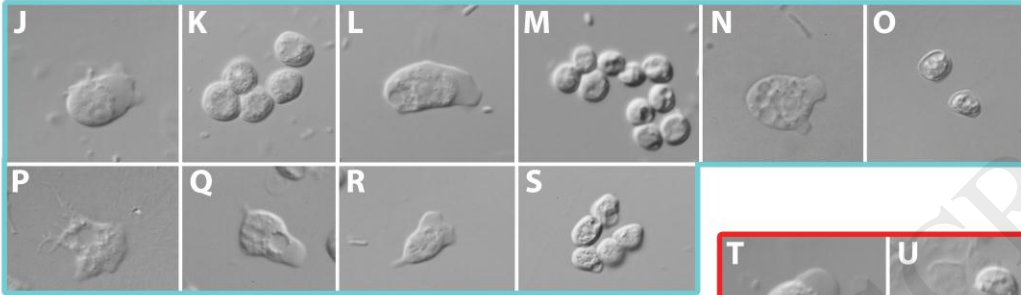
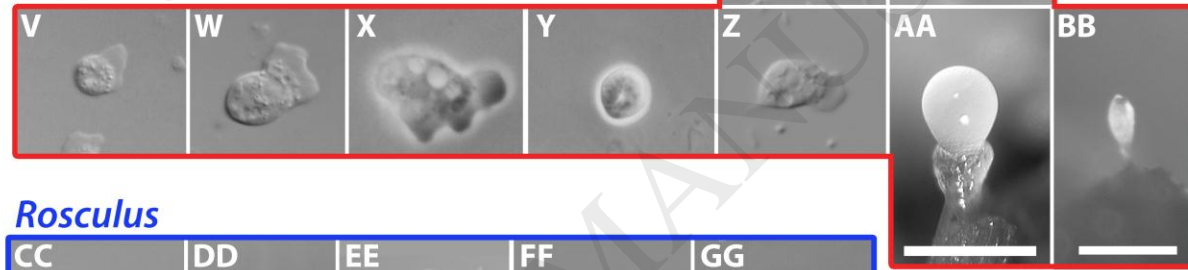
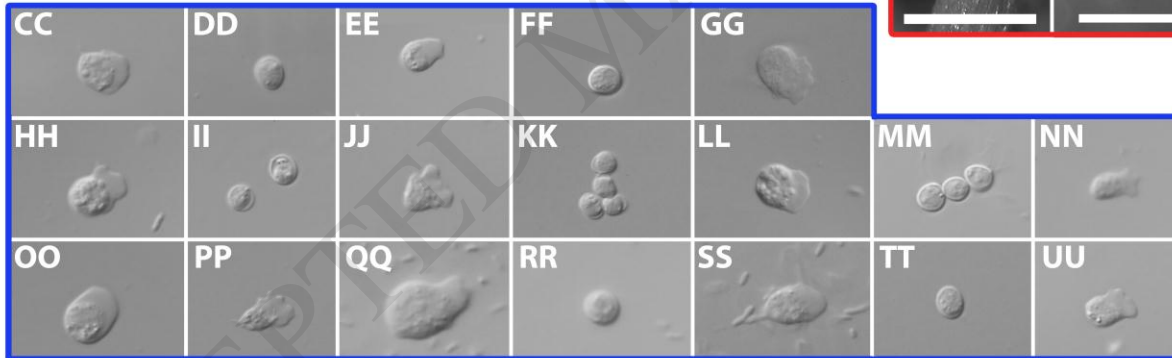
Figure 2. Bayesian phylogeny of a subset of Rhizaria rooted with two Stramenopiles as an outgroup based on 1,356 nucleotide positions of the 18S rRNA gene. The tree was constructed using Mr. Bayes (GTR + Γ model) and RaxML (GTR + Γ model). Values at nodes represent ML bootstrap and BI posterior probability values, respectively. Support values less than 60/0.60 are left blank and nodes not recovered in ML analysis are represented with a *. White circles and black circles represent fully supported nodes in ML bootstraps and BI posterior probability values respectively. Branches with // have a length divided by two and branches with //// have a length reduced by four. The genera of Sainouroidea are outlined with the following colors: Orange – *Acantholus*; green – *Olivorum*; Light Blue – *Homocognata*; Pink – *Puppisaman*; Red – *Guttulinopsis*; Navy Blue – *Rosculus*.

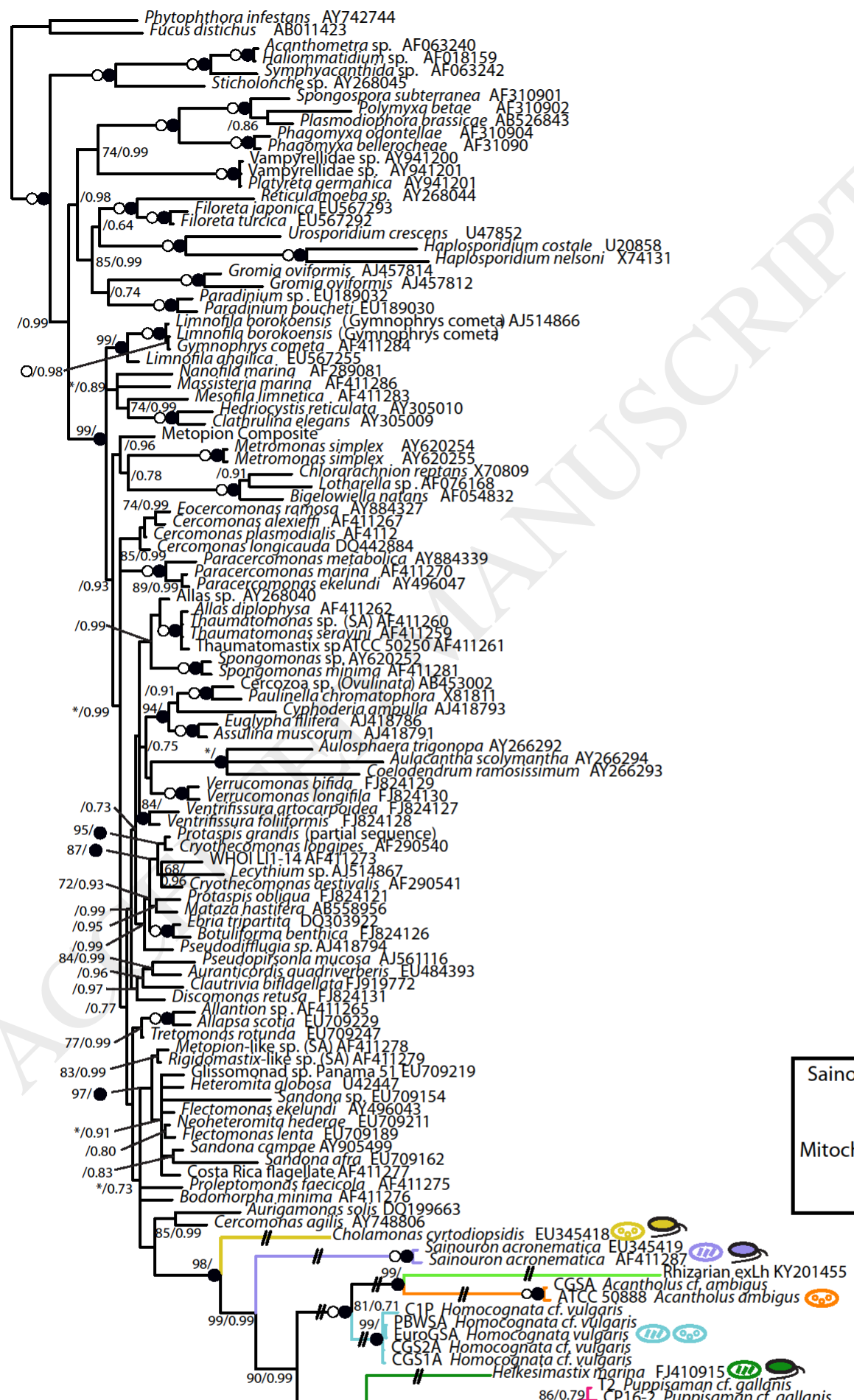
Figure 3. Above: Bayesian phylogeny of Sainouroidea rooted with *Cholamonas cyrtodiopsisidis* based on 1,393 aligned nucleotide positions of the 18S rRNA gene. The tree was constructed using Mr. Bayes (GTR + Γ model) and RaxML (GTR + Γ model). Values at nodes represent ML bootstrap and BI posterior probability values, respectively. White circles and black circles represent fully supported nodes in ML bootstraps and BI posterior probability values respectively. Genera are outlined with the following colors: Orange – *Acantholus*; green – *Olivorum*; Light Blue – *Homocognata*; Pink – *Puppisaman*; Red – *Guttulinopsis*; Navy Blue – *Rosculus*.

Below: Heat map of an uncorrected pairwise distance based on the 18S rRNA gene and 1,393 aligned nucleotide positions. The heat map was constructed using PAUP. Only 18S rDNA sequences > 1500 bp were used in this analysis.

Figure 4. PCA analysis of morphological characters (length, width, nucleus diameter, nucleolus diameter) from trophic amoebae. The ellipse for each genus represents 95% probability. Genera are designated with the following colors: Orange – *Acantholus*; Green – *Olivorum*; Light Blue – *Homocognata*; Pink – *Puppisaman*; Red – *Guttulinopsis*; Navy Blue – *Rosculus*.

Figure 5. TEM micrographs of **A)** *Homocognata vulgaris* n. gen. n. sp., EuroGSA (European grass snake amoeba) mostly distorted tubes/ampuliform (tubular-like) mitochondrial cristae. **B)** *Homocognata vulgaris* n. gen. n. sp., EuroGSA flat mitochondrial cristae. **C)** *Acantholus ambiguus* n. gen. n. sp., ATCC 50888 tubular mitochondrial cristae. **D)** *Puppisaman gallanis* n. gen. n. sp., CP16-1 (chicken poo 1) flat mitochondrial cristae. **E)** *Olivorum cimiterus* n. gen. n. sp., UACEM flat mitochondrial cristae. **F)** *Rosculus philangulis* n. sp. RSA (ratsnake amoeba) flat mitochondrial cristae. All Scale Bars = 200nm. Genera are outlined with the following colors: Orange – *Acantholus*; Green – *Olivorum*; Light Blue – *Homocognata*; Pink – *Puppisaman*; Red – *Guttulinopsis*; Navy Blue – *Rosculus*.

Acantholus*Olivorum**Puppisaman**Homocognata**Guttulinopsis**Rosculus*



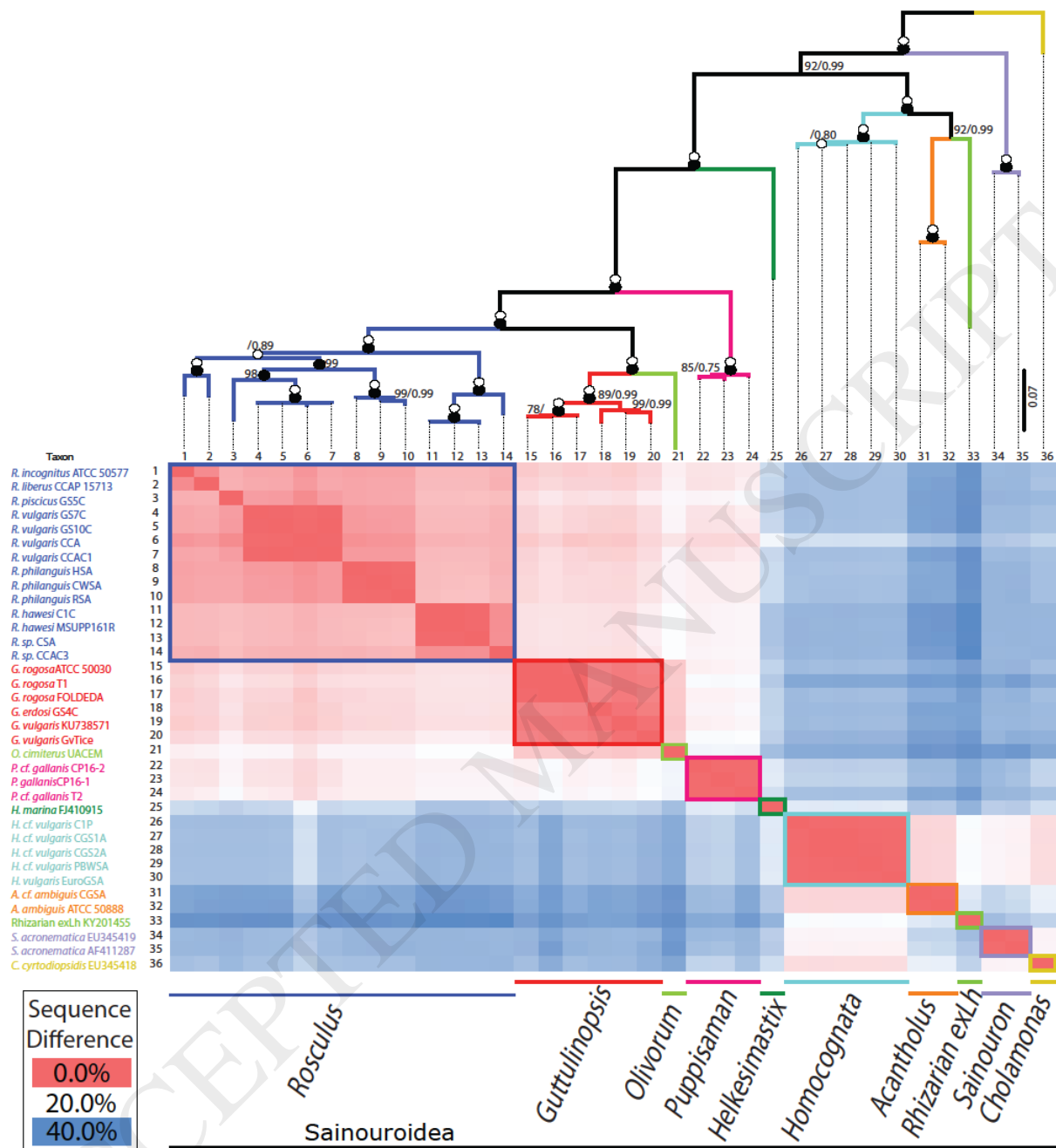
Stram
R

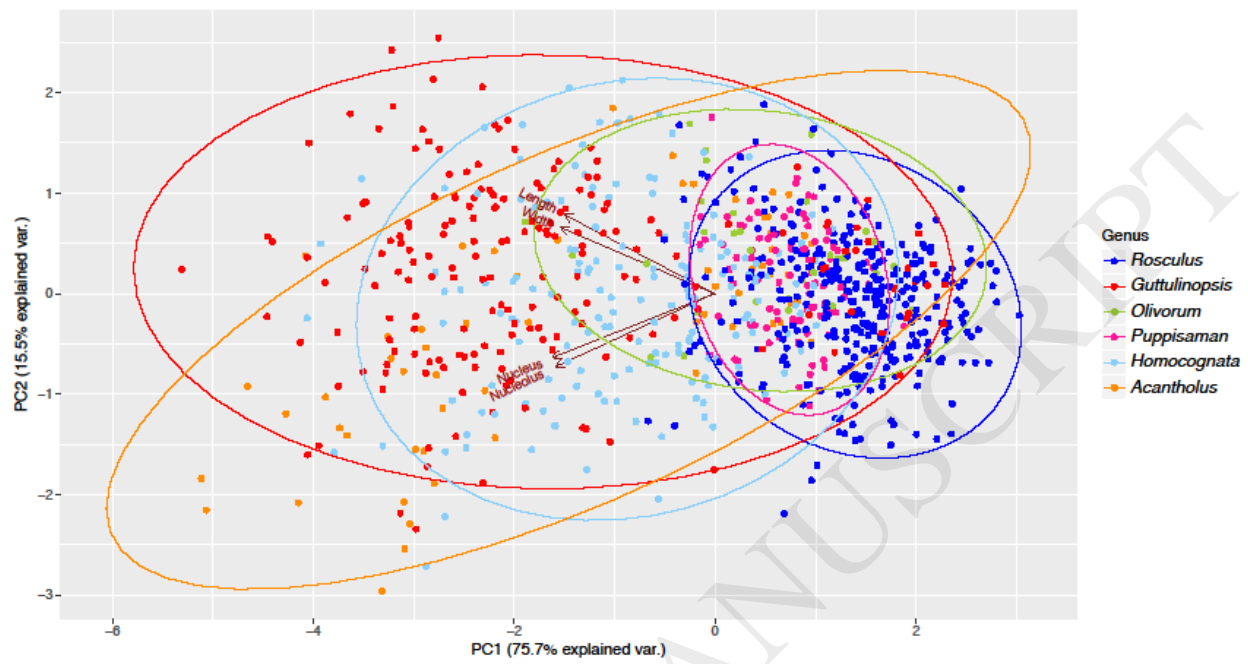
Sainouroid Features

- Flagellate
- Sorocarp

Mitochondrial Cristae

- Flat
- Tubular





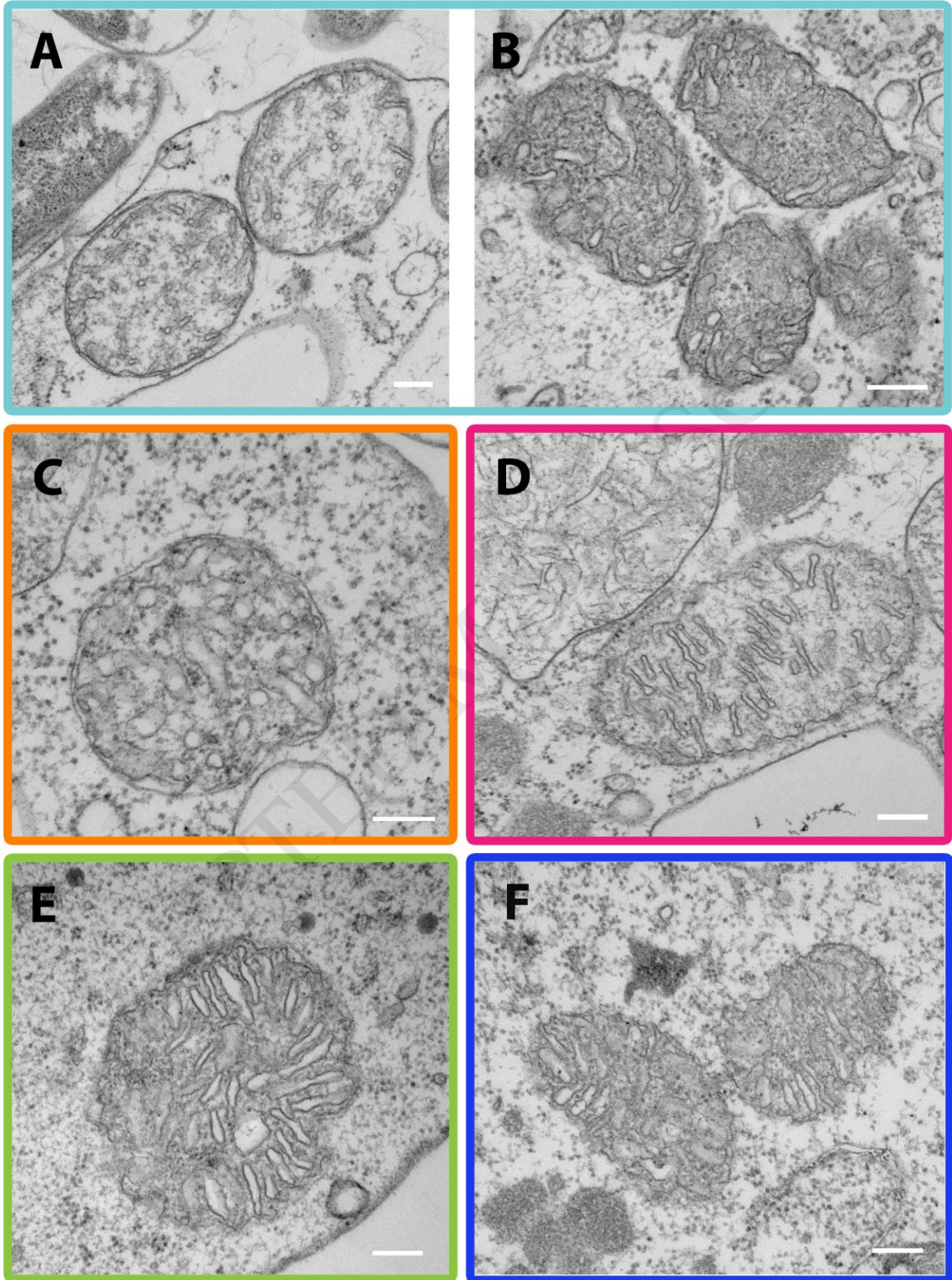


Table 1. 18S rDNA PCR primers used for amplification and those designed for sequencing *Rosculus*, *Guttulinopsis* and *Olivorum* 18S rRNA genes.

ID	Primer Name	Primer Sequence (5'-3')	Specificity of Primer
A	5' ssu 18!	CTGGTTGATCCTGCCAGT	Universal Eukaryotic
B	RibBr	GATCCTTCTGCAGGTTACC	Universal Eukaryotic
C	5'	CCGAATTCGTCGACAACCTGGTGGATCCTGCCAGT	Universal Eukaryotic
D	3'	CCCGGGATCCAAGCTTGATCCTTCTGCAGGTTACCTAC	Universal Eukaryotic
E	30F	AAAGATTAAGCCATGCAT	Universal Eukaryotic
F	GV5'	GAATTCACATTTGATCTTGATGT	Rosc/Gutt-specific
G	HGR419F	GCAGCAGGSRGMAAATT	Rosc/Gutt-specific
H	HGR2037R	ACCTTGTACGACTTTGGCTTCCTCTA	Rosc/Gutt-specific
I	1200F	ACAGGTCTGTGATGCCC	Universal Eukaryotic
J	514F	GGTGCCAGCASC CGCGGTAA	Universal Eukaryotic
K	GR514F	TGCCAGCAGCAGCGGTAAT	Rosc/Gutt-specific
L	GR514R	TATTACCGCTGCTGCTGGCA	Rosc/Gutt-specific
M	GR1500F	GTAGTGATGGCCTTTGGGAAG	Rosc/Gutt-specific
N	GR1500R	CACTTCCAAAGGCCATGCAC	Rosc/Gutt-specific