
Evolutionary Origins of Rhizarian Parasites

Sierra Roberto ^{1,*}, Canas-Duarte Silvia J. ^{2,8}, Burki Fabien ³, Schwelm Arne ⁴, Fogelqvist Johan ⁴,
Dixelius Christina ⁴, Gonzalez-Garcia Laura N. ², Gile Gillian H. ⁵, Slamovits Claudio H. ⁵,
Klopp Christophe ⁶, Restrepo Silvia ², Arzul Isabelle ⁷, Pawlowski Jan ¹

¹ Univ Geneva, Dept Genet & Evolut, Geneva, Switzerland.

² Univ Los Andes, Dept Biol Sci, Bogota, Colombia.

³ Univ British Columbia, Dept Bot, Vancouver, BC, Canada.

⁴ Swedish Univ Agr Sci, Linnean Ctr Plant Biol, Uppsala BioCtr, Dept Plant Biol, Uppsala, Sweden.

⁵ Dalhousie Univ, Dept Biochem & Mol Biol, Halifax, NS, Canada.

⁶ INRA, GenoToul Bioinformat Facil, F-31931 Toulouse, France.

⁷ IFREMER, La Tremblade, France.

⁸ Harvard Univ, Sch Med, Dept Syst Biol, Boston, MA USA.

* Corresponding author : Roberto Sierra, email address : roberto.sierra@me.com

Abstract :

The SAR group (Stramenopila, Alveolata, Rhizaria) is one of the largest clades in the tree of eukaryotes and includes a great number of parasitic lineages. Rhizarian parasites are obligate and have devastating effects on commercially important plants and animals but despite this fact, our knowledge of their biology and evolution is limited. Here, we present rhizarian transcriptomes from all major parasitic lineages in order to elucidate their evolutionary relationships using a phylogenomic approach. Our results suggest that Ascetosporea, parasites of marine invertebrates, are sister to the novel clade Apofilosa. The phytomyxean plant parasites branch sister to the vampyrellid algal ectoparasites in the novel clade Phytorhiza. They also show that Ascetosporea + Apofilosa + Retaria + Filosa + Phytorhiza form a monophyletic clade, although the branching pattern within this clade is difficult to resolve and appears to be model-dependent. Our study does not support the monophyly of the rhizarian parasitic lineages (Endomyxa), suggesting independent origins for rhizarian animal and plant parasites.

Keywords : Rhizaria, parasites, phylogenomics

The SAR group includes a large diversity of eukaryotic parasites, including economically important plant parasites such as oomycetes (Stramenopila) and plasmodiophorids (Rhizaria), as well as the human malaria parasite *Plasmodium* (Alveolata), along with many other animal parasites that infect commercially important invertebrates. Parasitic rhizarians comprise poorly studied organisms that are split into two major assemblages based on their hosts, either animals or plants. The Ascetosporea, composed of Haplosporida (e.g. *Bonamia*) and Paramyxida (e.g. *Marteilia*, no phylogenomic data available for the group), are intracellular parasites of marine invertebrates such as Turbellaria, annelids, molluscs and crustaceans (Hartikainen, Ashford, et al. 2014; Hartikainen, Stentiford, et al. 2014). The Phytomyxea, composed of Plasmodiophorida (e.g. *Plasmodiophora*) and Phagomyxida (e.g. *Phagomyxa*, no phylogenomic data available for the group), are obligate biotrophic parasites of plants, diatoms, oomycetes and brown algae (Neuhauser, et al. 2014). Additionally, the naked vampyrellid amoebae, most commonly found in freshwater habitats, are ectoparasites of algae, fungi, small metazoans, and other protists (Hess, et al. 2012; Berney, et al. 2013). Despite being very diverse, vampyrellids have remained enigmatic but these parasites will likely gain attention as the commercial algae industry grows (Carney and Lane 2014).

All rhizarian parasites have been classified as Endomyxa based on 18S rRNA phylogenies together with two non-parasitic lineages *Gromia* and *Filoreta* (Cavalier-Smith 2002). However, multiple gene phylogenetic analysis did not support their common origin (Sierra, et al. 2013), but also did not resolve their evolutionary relationships, mainly because Ascetosporea were represented only by the extremely fast-evolving *Mikrocytos mackini* (Burki, et al. 2013). Here, we generated extensive transcriptomic data from representative species of all major rhizarian clades to elucidate the rhizarian parasites' evolutionary relationships by means of phylogenomic analyses. Most notably, we obtained and analyzed Illumina sequenced transcriptomic data for two ascetosporeans (*Bonamia* and *Minchinia*), a vampyrellid (*Leptophrys*), and the plasmodiophorids *Spongospora* and the complete genome data from *Plasmodiophora* (Schwelm, et al. 2015). In addition, 454 GS-FLX Titanium reads were generated for 12 additional rhizarian species, including *Gromia*, *Aulacantha scolymantha* and 10 retarians (fig. 1, species in bold). These data were used to assemble a super-matrix of 229 proteins and 56 species containing

64,107 amino acid positions. The average missing sites per taxon in the full matrix is 45% (supplementary table S2).

We recovered a sister relationship of Rhizaria and Alveolata with Stramenopila at their base using green plants as outgroup (fig. 1). This topology is strongly supported (fig. 1, branch support 1/0.5/98/98/100) by most analyses with the exception of the Bayesian inference calculated on PhyloBayes (Lartillot, et al. 2009), where the posterior probability (PP) for the branch is 0.5 (fig. 1 and supplementary fig. S3). This result is also congruent with the previously obtained phylogeny based on 109 protein-coding genes (Sierra, et al. 2013). However, different topologies have been proposed for the SAR group. Burki, et al. (2013) suggested a sister relationship of Alveolata and Stramenopila with Rhizaria at their base while Parfrey, et al. (2010) proposed a sister relationship of Rhizaria and Stramenopila with Alveolata at their base. Hence the branching order within the group remains unresolved.

Our results show that the ascetoporean parasites form a monophyletic group of shellfish parasites including a strongly supported (fig. 1, branch support .99/1/93/98/93) clade of *Bonamia* and *Minchinia*. The best topologies obtained from our analyses place the intracellular oyster parasite, *M. mackini*, in the Ascetosporea clade. However, due to the fast evolutionary rates, its phylogenetic position is not strongly supported (fig. 1, branch support .57/-/53/65/100), except for the maximum support obtained with the non-parametric SH-*alrt* single branch test. The Bayesian analysis obtained on PhyloBayes (Lartillot, et al. 2009) does not resolve the position of *M. mackini* and thus results in a trichotomy of Retaria, Ascetosporea + *Gromia* + *Filoreta* and *M. mackini* (supplementary fig. S3). As previously suggested by Burki, et al. (2013), the Ascetosporea form a sister group to the clade *Gromia* + *Filoreta*, called here Apofilosa. The support for this sister group relationship is not very strong (fig. 1, branch support .60/-/53/65/100), yet in the absence of *M. mackini*, the Ascetosporea + Apofilosa form a robust clade (supplementary fig. S1, branch support .92/.97/100/100). On the other hand, the rhizarian plant parasites, Phytomyxea and Vampyrellida were recovered in a strongly supported (fig. 1 branch support 1/1/100/100/100) monophyletic clade named here Phytorhiza. In the absence of *M. mackini*, the Phytorhiza clade is still recovered with maximum branch support values, but its position within Rhizaria is weakly supported (supplementary fig. S1, branch support .92/-/64/70) compared to the full data set analyses (fig. 1 branch support .98/1/89/90/100). A monophyletic

Endomyxa (rhizarian parasites, *Gromia* and *Filoreta*) was only recovered by Bayesian inference implemented on PhyloBayes in the absence of *M. mackini*, though not supported (0.66 PP, supplementary fig. S8). Additionally, the sister relationship of this assemblage with Retaria is not resolved (0.58 PP).

To conclude, the highly supported deep nodes of our phylogenomic tree suggest that there are four major groups of Rhizaria: Filosa, Phytorhiza, Ascetosporea + Apofilosa, and Retaria (fig. 1); this greatly clarifies rhizarian phylogeny. Nonetheless, the diversity of these groups is extremely high and further sampling and data is needed to obtain better phylogenetic resolution. In particular, the addition of two ascetosporean taxa to the data matrix resulted in a sister relationship between ascetosporeans and Apofilosa, together with the strongly supported clade of Vampyrellida and Phytomyxea, allowed confirming the independent evolution of parasitism in Phytorhiza and Ascetosporea.

Materials and Methods

Species and alignments. A phylogenomic supermatrix of 229 genes (64107 positions, Supplementary table S1) was constructed using full-length proteins based on a pool of genes from (Burki, et al. 2010; Burki, et al. 2012; Burki, et al. 2013). Rhizarian sequences were assigned and added to the individual alignments based on homology searches using blastp. The alignments included representative species for all major groups of eukaryotes. New, unpublished sequences from *L. vorax*, *S. subterranea*, *P. brassicae*, *M. chitonis*, *B. ostreae*, *M. mackini*, *G. sphaerica*, six species of Foraminifera (*Elphidium margaritaceum*, *Globobulimina turgida*, *Brizalina* sp., *Bulimina marginata*, *Nonionellina* sp., and *Ammonia* sp.), 5 radiolarians (*Phyllostaurus siculus*, *Amphilonche elongata*, *Astroilonche serrata*, *Collozoum* sp., and *Spongospaera streptacantha*) and *A. scolymantha* were generated for this study. See supplementary methods for details of library preparation, sequencing, and assembly. All new sequences were added and automatically aligned to the data set using MAFFT v.6.847b (Katoh et al., 2002). Ambiguously aligned positions were removed using BMGE (Criscuolo and Gribaldo 2010).

Phylogenetic analyses. Single-gene maximum likelihood trees were performed on RAxML v.8.0.26 (Stamatakis 2006) using the PROTGAMMALGF setting with 1000 bootstrap replicates. All trees were carefully examined to discard any sequence from the SAR group that branched with plants or animals. The 229 alignments were concatenated using SCAFoS (Roure, et al. 2007). The ML analysis of the phylogenomic matrix was performed using RAxML with PROTGAMMALGF setting and statistical support was evaluated with 1000 standard bootstrap replicates. The IQ-Tree algorithm was used to calculate a second ML analysis with LG+G parameters, 1000 Ultrafast bootstrap replicates and the non parametric SH-*alrt* single branch test (Nguyen, et al. 2015). Bayesian Inferences was carried out using ExaBayes using *revmat*=(0) setting (Aberer, et al. 2014) and PhyloBayes v.3.3f (Lartillot, et al. 2009) under the CAT-GTR model with two independent chains until convergence. For post-analysis of the independent chains a 20% burnin was used.

Data availability. The transcriptomes used for this study were deposited under: SRX1014535, SRX1014576, SRX1014585, SRX1014586, SRX1014587, SRX1014589, SRX1016224, SRS621308, SRS412271, PRJEB8376, PRJEB9159, SRS621308, SRS621524.

Acknowledgements. Most computations were performed at the University of Geneva on the Baobab cluster (<http://baobab.unige.ch>), PhyloBayes was run on the CIPRES portal (phylo.org), sequence assemblies were performed at the Vital-IT High Performance Computing Center (<http://www.vital-it.ch>) at the Swiss Institute of Bioinformatics. We are indebted to Bryony A. P. Williams for providing chitons infected with *Minchinia chitonis*. GHG is supported by an NSERC postdoctoral fellowship. CHS is a Fellow of the Canadian Institute for Advanced Research and is also supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 386345). AS and CD are supported by BioSoM and the Swedish University of Agricultural Sciences, SLU. This work was supported by the Swiss National Science Foundation Grant No. 140766 and 159709 (RS, JP).

References

- Aberer AJ, Kobert K, Stamatakis A. 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol Biol Evol* 31:2553-2556.
- Berney C, Romac S, Mahe F, Santini S, Siano R, Bass D. 2013. Vampires in the oceans: predatory cercozoan amoebae in marine habitats. *ISME J* 7:2387-2399.
- Burki F, Corradi N, Sierra R, Pawlowski J, Meyer GR, Abbott CL, Keeling PJ. 2013. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. *Curr Biol* 23:1541-1547.
- Burki F, Kudryavtsev A, Matz MV, Aglyamova GV, Bulman S, Fiers M, Keeling PJ, Pawlowski J. 2010. Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. *BMC Evol Biol* 10:377.
- Burki F, Okamoto N, Pombert JF, Keeling PJ. 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc Biol Sci* 279:2246-2254.
- Carney LT, Lane TW. 2014. Parasites in algae mass culture. *Front Microbiol* 5:278.
- Cavalier-Smith T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol* 52:297-354.
- 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.
- Hartikainen H, Ashford OS, Berney C, Okamura B, Feist SW, Baker-Austin C, Stentiford GD, Bass D. 2014. Lineage-specific molecular probing reveals novel diversity and ecological partitioning of haplosporidians. *ISME J* 8:177-186.
- Hartikainen H, Stentiford GD, Bateman KS, Berney C, Feist SW, Longshaw M, Okamura B, Stone D, Ward G, Wood C, et al. 2014. Mikrocytids are a broadly distributed and divergent radiation of parasites in aquatic invertebrates. *Curr Biol* 24:807-812.
- Hess S, Sausen N, Melkonian M. 2012. Shedding light on vampires: the phylogeny of vampyrellid amoebae revisited. *PLoS ONE* 7:e31165.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286-2288.
- Neuhauser S, Kirchmair M, Bulman S, Bass D. 2014. Cross-kingdom host shifts of phytomyxid parasites. *BMC Evol Biol* 14:33.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268-274.

Parfrey LW, Grant J, Tekle YI, Lasek-Nesselquist E, Morrison HG, Sogin ML, Patterson DJ, Katz LA. 2010. Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst Biol* 59:518-533.

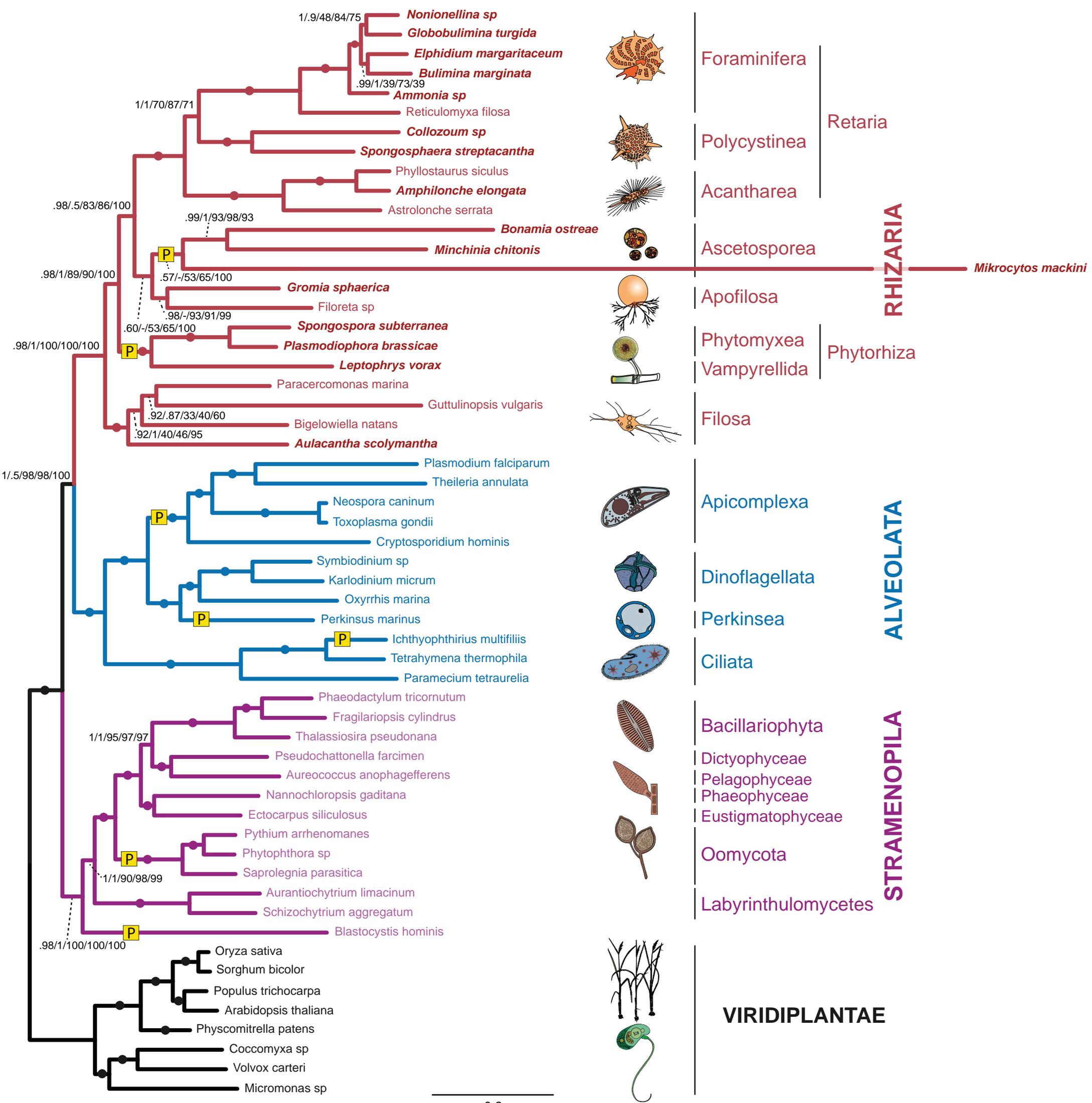
Roure B, Rodriguez-Ezpeleta N, Philippe H. 2007. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol Biol* 7 Suppl 1:S2.

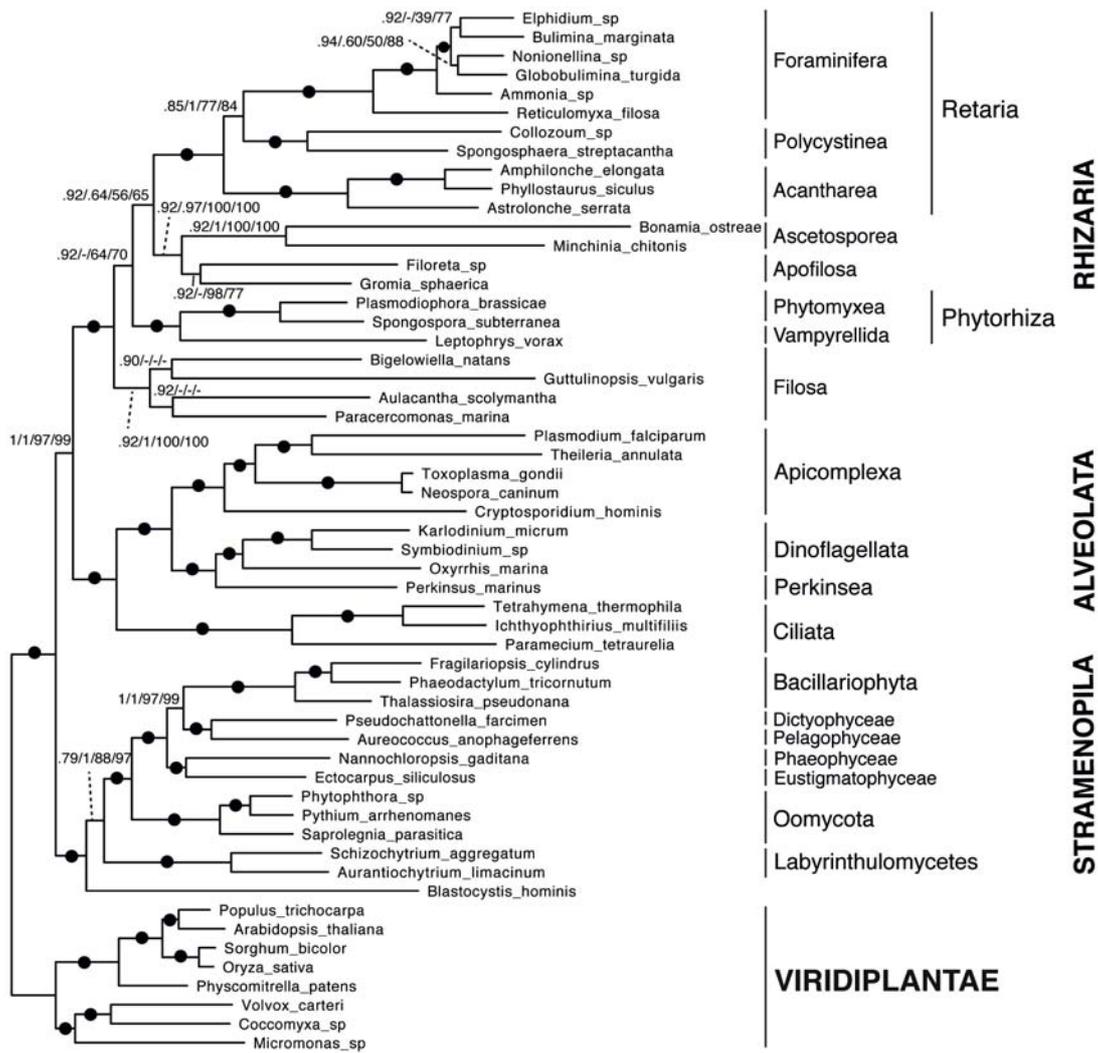
Schwelm A, Fogelqvist J, Knaust A, Jülke S, Lilja TB-R, G., Karlsson M, Shevchenko A, Dhandapani V, Choi SR, Kim HG, et al. 2015. The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. *Sci Rep* 5:11153.

Sierra R, Matz MV, Aglyamova G, Pillet L, Decelle J, Not F, de Vargas C, Pawlowski J. 2013. Deep relationships of Rhizaria revealed by phylogenomics: a farewell to Haeckel's Radiolaria. *Mol Phylogenet Evol* 67:53-59.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690.

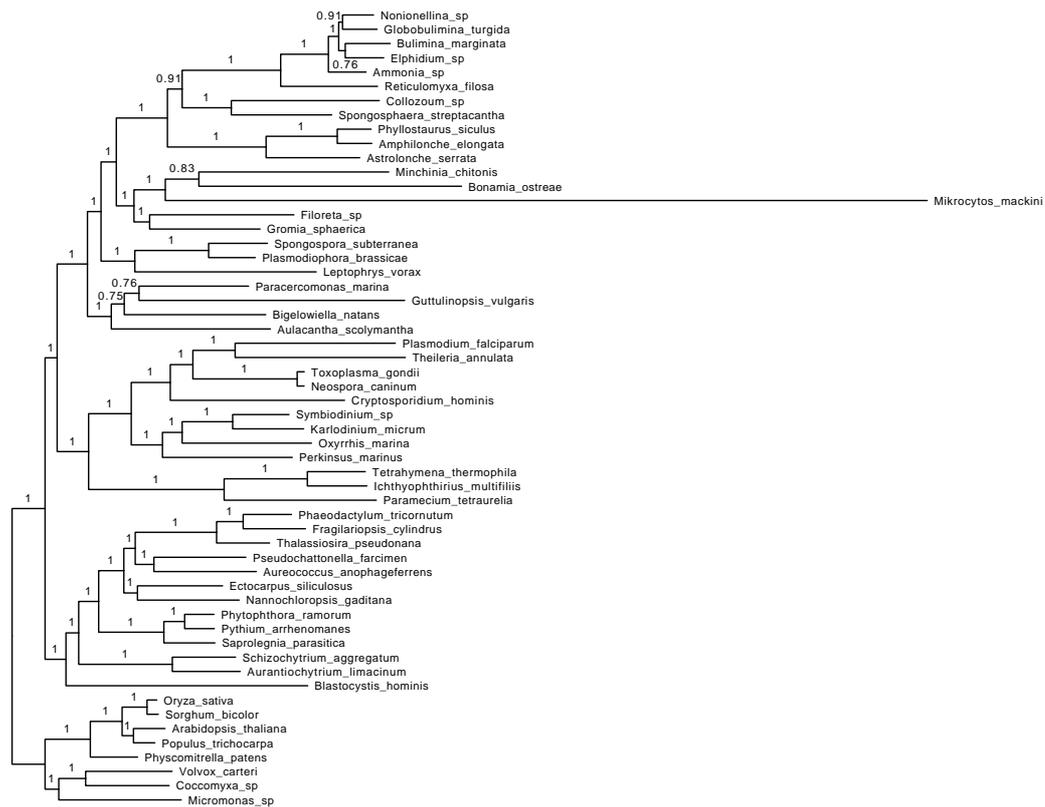
Figure 1. Phylogeny of the SAR group. The consensus tree shown was obtained by the majority rule of the Bayesian tree pool after 385 000 generations inferred with 229 genes using the GTR matrix (revMat with Dirichlet process). The statistical support on branches was estimated by Bayesian posterior probabilities of post-burnin bipartitions using ExaBayes, PhyloBayes under CAT-GTR model, bootstrap replicates from maximum likelihood analysis on RAxML and ultrafast bootstrap values and the single branch SH-like approximate likelihood ratio test (SH-aLRT) implemented on IQ-Tree, respectively. Maximum statistical support was obtained under the five methods where solid dots and no values are shown (i.e. 1/1/100/100/100). The parasitic lineages are marked with a capital "P" on the branches. New data added in this study is shown for the rhizarian species in bold italic fonts.





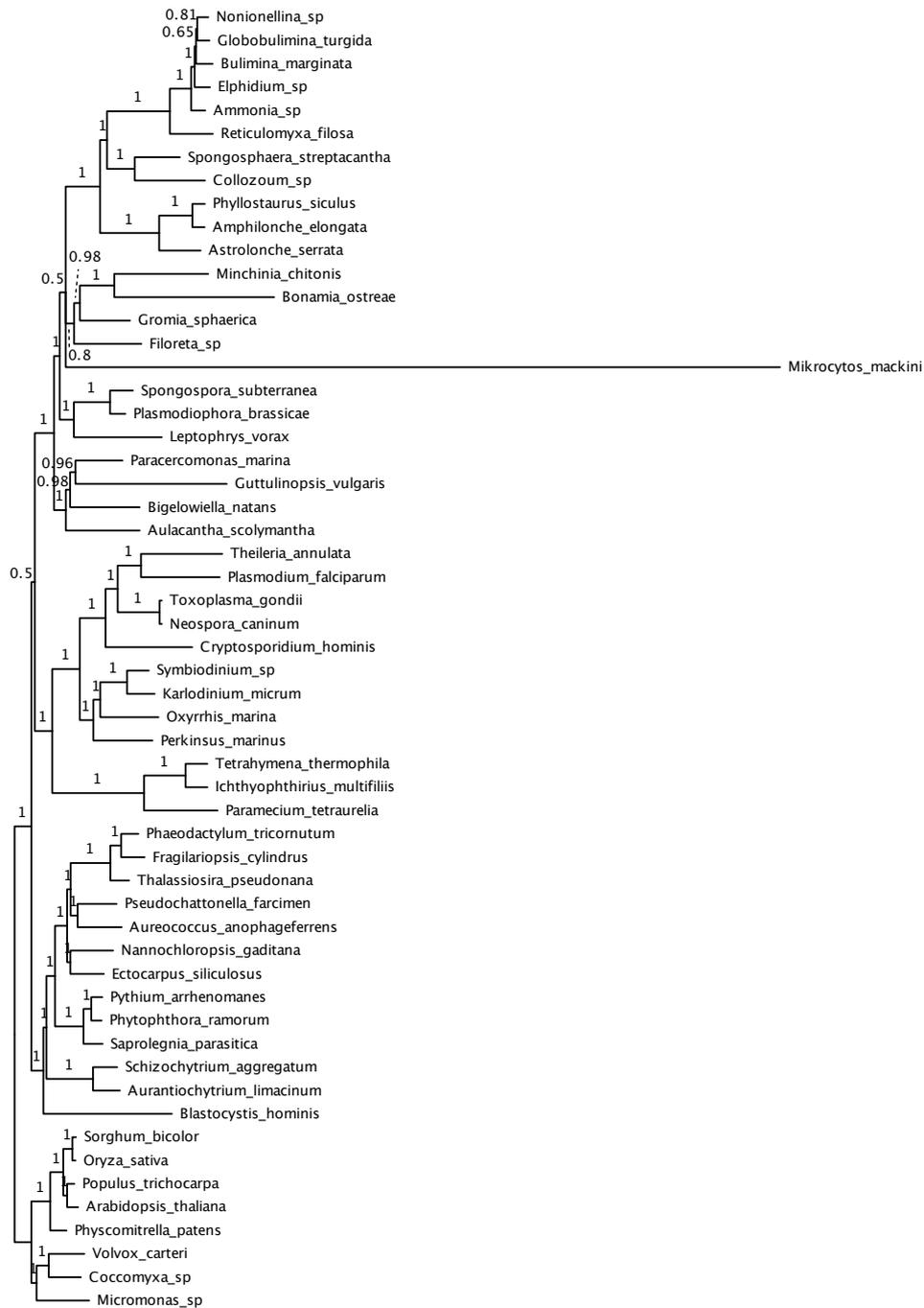
0.2

Supplementary figure S1. Phylogeny of the SAR group excluding *Mykrocytos mackini*. Summary tree of supplementary figures 7-10. The consensus tree shown was obtained by the majority rule of the Bayesian tree pool after 15 000 generations inferred with 229 genes using the GTR matrix (revMat with Dirichlet process). The statistical support on branches was estimated by Bayesian posterior probabilities of post-burnin bipartitions using ExaBayes, PhyloBayes under CAT-GTR model, bootstrap replicates from maximum likelihood analysis on RAxML and ultrafast bootstrap values implemented on IQ-Tree, respectively. Maximum statistical support was obtained under the four methods where solid dots and no values are shown (i.e. 1/1/100/100).

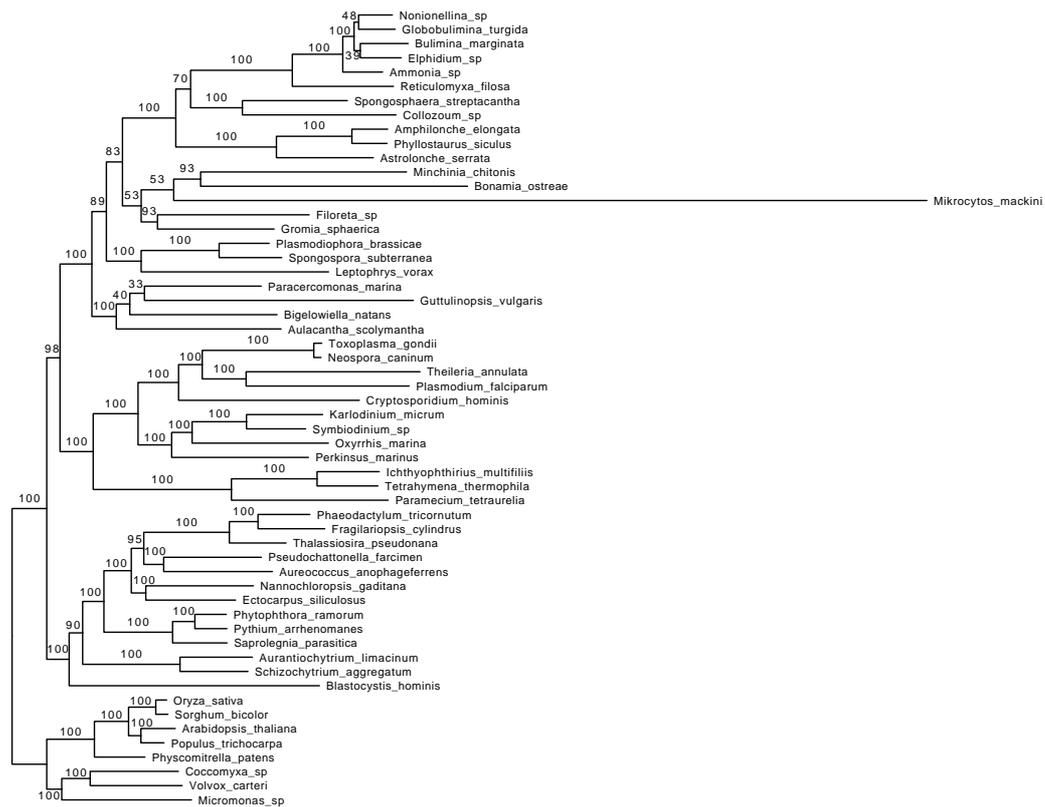


0.3

Supplementary figure S2. Phylogeny of the SAR group. The consensus tree shown was obtained by the majority rule of the Bayesian tree pool after 15,000 generations inferred with 229 genes using the GTR matrix (revMat with Dirichlet process) as implemented in ExaBayes. The statistical support on branches was estimated by Bayesian posterior probabilities of post-burnin bipartitions.

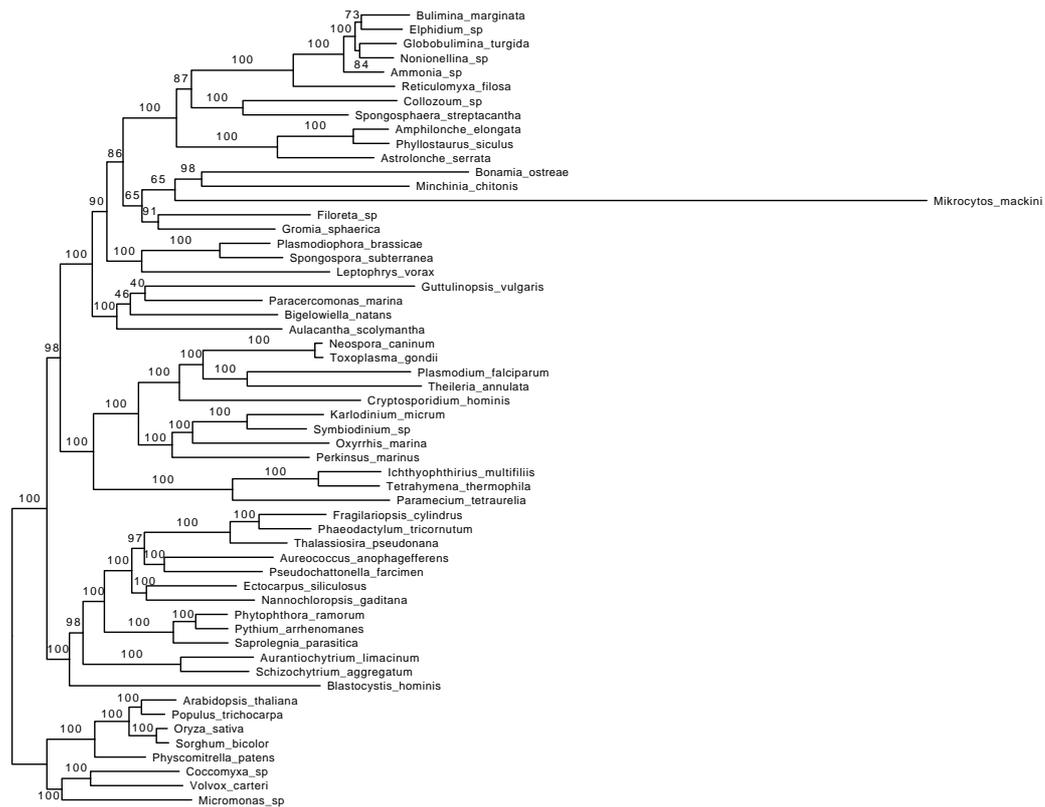


Supplementary figure S3. Phylogeny of the SAR group. The strict consensus tree shown was obtained by the majority rule of the Bayesian tree pool after 4,000 generations inferred with 229 genes using the CAT+GTR setting as implemented in PhyloBayes. The statistical support on branches was estimated by Bayesian posterior probabilities of post-burnin bipartitions.



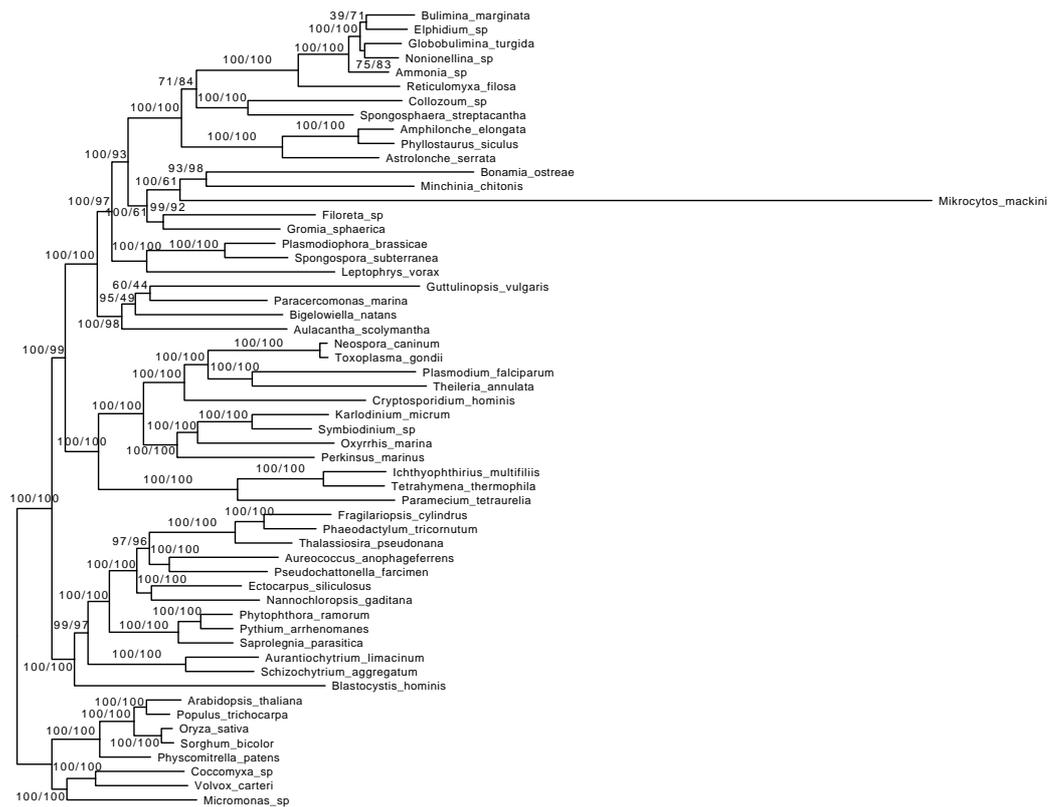
0.3

Supplementary figure S4. Phylogeny of the SAR group. The best maximum likelihood tree shown was obtained using RAxML inferred with 229 genes using the PROTGAMMALG setting. The statistical support on branches was estimated by 1000 bootstrap replicates.



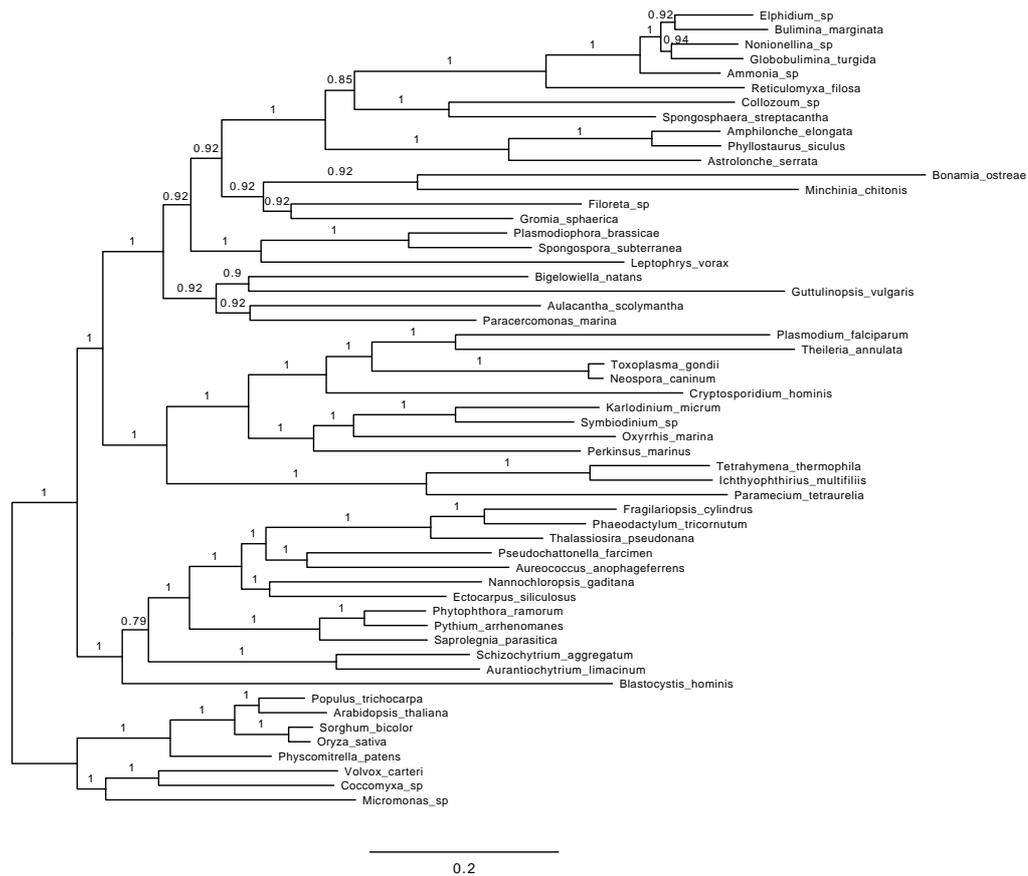
0.3

Supplementary figure S5. Phylogeny of the SAR group. The best maximum likelihood tree shown was obtained using IQ-Tree inferred with 229 genes using the LG+GAMMA setting. The statistical support on branches was estimated by 1000 ultrafast bootstrap replicates.

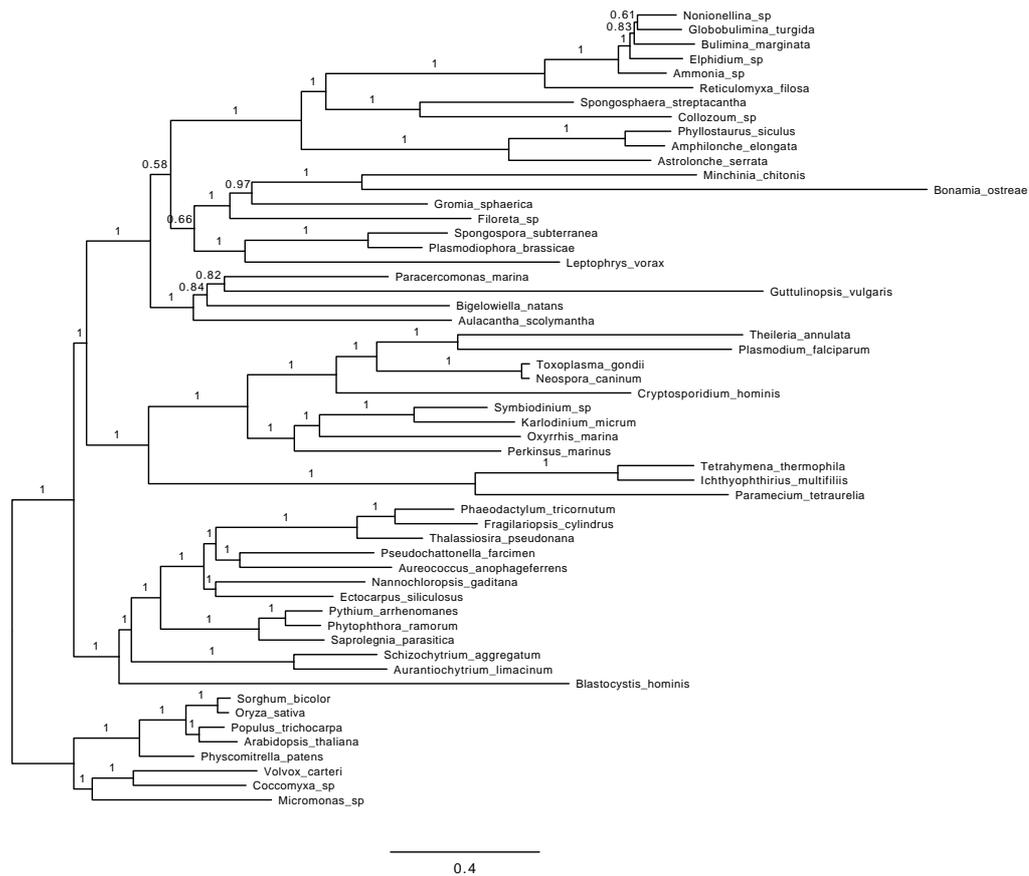


0.3

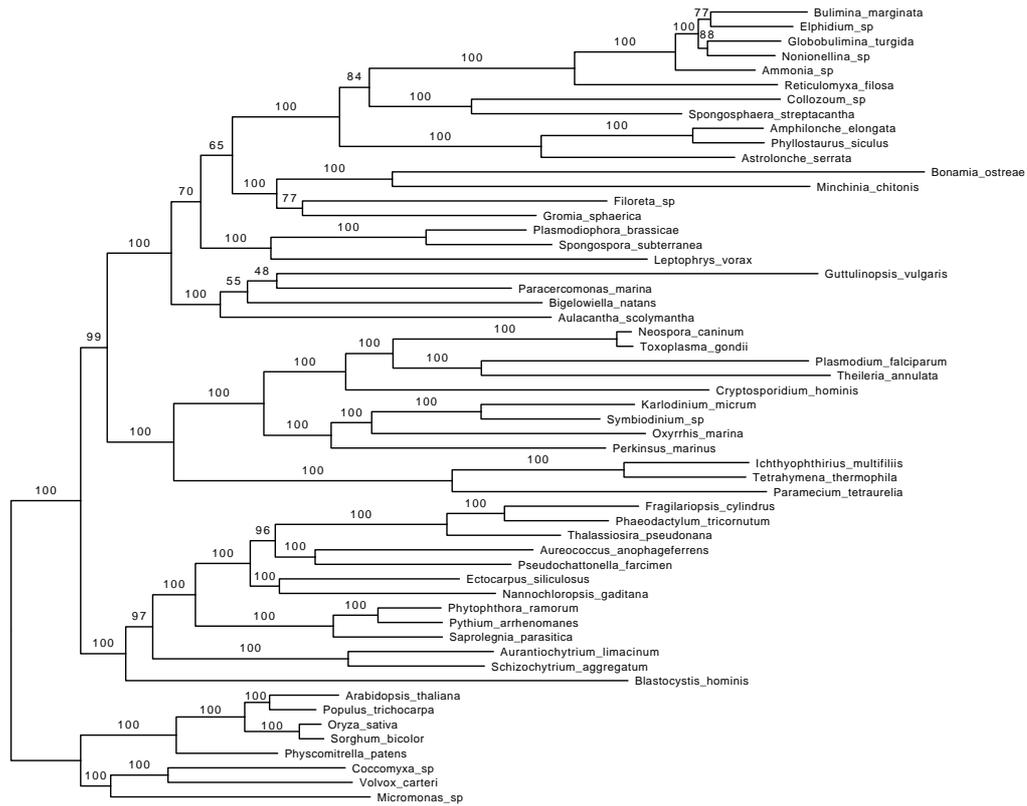
Supplementary figure S6. Phylogeny of the SAR group. The best maximum likelihood tree shown was obtained using IQ-Tree inferred with 229 genes using the LG+GAMMA setting. The statistical support on branches was estimated by the non-parametric SH-*alrt* and 1000 ultrafast bootstrap replicates, respectively.



Supplementary figure S7. Phylogeny of the SAR group excluding *Mikrocytos mackini*. The consensus tree shown was obtained by the majority rule of the Bayesian tree pool after 15,000 generations inferred with 229 genes using the GTR matrix (revMat with Dirichlet process) as implemented in ExaBayes. The statistical support on branches was estimated by Bayesian posterior probabilities of post-burnin bipartitions.



Supplementary figure S8. Phylogeny of the SAR group excluding *Mikrocytos mackini*. The strict consensus tree shown was obtained by the majority rule of the Bayesian tree pool after 4,000 generations inferred with 229 genes using the CAT+GTR setting as implemented in PhyloBayes. The statistical support on branches was estimated by Bayesian posterior probabilities of post-burnin bipartitions.



0.2

Supplementary figure S10. Phylogeny of the SAR group excluding *Mikrocytos mackini*. The best maximum likelihood tree shown was obtained using IQ-Tree inferred with 229 genes using the LG+GAMMA setting. The statistical support on branches was estimated by 1000 ultrafast bootstrap replicates.

Evolutionary origins of rhizarian parasites

Supplementary Materials and Methods

Isolation of *Leptophrys vorax*, library and preliminary analyses.

Leptophrys vorax CCAC 3422B was co-cultured in tall petri dishes with *Closterium* sp. CCAC 2697B in Waris-H culture medium at 18°C. After the algae were consumed, the vampyrellid cells were collected and total RNA was extracted using the NucleoSpin RNA XS (Macherey–Nagel, Germany). The sample was prepared using the TruSeq RNA sample kit (Illumina) and sequenced on the Illumina HiSeq2000. Low quality reads were discarded and the raw reads were assembled using the Velvet/oases software (Zerbino and Birney 2008).

Isolation of *Plasmodiophora brassicae* and *Spongospora subterranean*, library and preliminary analyses.

The DNA and RNA isolations and gene prediction and annotation we performed as described by Schwelm, et al. (2015).

Isolation of *Bonamia ostreae*, library preparation and preliminary analysis

Sample collection. *B. ostreae* was purified according to the protocol developed by Mialhe et al. (1988) from one highly infected flat oyster *Ostrea edulis* collected in Quiberon Bay (Morbihan, France), an endemic bay regarding bonamiosis. Briefly, the heavily infected oyster was selected by examination of heart tissue imprints using light microscopy. After homogenisation of all the organs except the adductor muscle, parasites were concentrated by differential centrifugation on sucrose gradients. Finally, the purified *B. ostreae* cells were suspended in filtered seawater (FSW) and counted using a Malassez-cell. About 2.108 parasites were obtained.

RNA extraction and library preparation. After removing, parasite cells were placed in TRIzol® (Invitrogen) at -80°C until RNA isolation was done following the manufacturer's protocol. Total RNA concentration was measured on a NanoDrop 2000c spectrophotometer (Thermo Scientific). The quality of RNA sample was then assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies). The cDNA libraries were

prepared from 3 µg of total RNA with Illumina Tru-Seq RNA Sample Prep Kits (Illumina) for 100 bp paired-end reads according to the manufacturer's instructions. Libraries were sequenced on a HiSeq 2000 (Illumina).

Transcriptome analyses. The read pairs were checked and stored in ng6 environment (Mariette, et al. 2012). They were cleaned from remaining sequencing adapters using trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The over-represented reads were filtered out using the normalize_by_kmer_coverage.pl script from the Trinity software package (Haas, et al. 2013). The longest sub-sequence without Ns were extracted from each read and kept if their length exceeded half of the initial read length. This step was performed using an in-house script. Nine Oases (Schulz, et al. 2012) assemblies using k-mers ranging from 25 to 69 (pace 6) were performed on the cleaned read pairs. Only the longest contig per locus provided by the OasesV0.2.04OutputToCsvDataBase.py script (<http://code.google.com/p/oases-to-csv/>) were retained. All contigs files were merged. Finally, anti-sense chimeras (accidentally produced by the assembly step) were found and split with an in-house script. Cd-hit-est (Li and Godzik 2006) was used to cluster and remove similar contigs produced by different k-mers, based on their sequence similarities. TGICL (Pertea, et al. 2003) was used to assemble contigs having significant overlaps. The contigs were filtered on a minimum length of 200 base pairs. Input reads were then mapped back to the contigs using bwa aln (Li and Durbin 2009). The resulting alignment files were used to correct the contig sequence from spurious insertions and deletions using an in-house script and to filter out those with very low coverage. The filter excluded contigs with less than two mapped reads per million.

Isolation of *Minchinia chitonis*, library and preliminary analysis.

Sample collection. Two heavily infected individuals of *Lepidochitona cinerea* were collected from Torre Abbey Sands beach, Torquay, Devon, on November 3, 2009. The presence of *Minchinia chitonis* spores was verified by light microscopy. Each chiton was allowed to rot in seawater in the lab for 2 weeks, after which time the plates of each chiton were removed, and remaining clumps of chiton tissue were diced with a razor

blade. The tissue and spore slurry from each chiton was concentrated by centrifugation and resuspended in 1 mL dH₂O.

Density gradient centrifugation. Samples were layered on 90% Percoll™ in 0.15M NaCl and spun at 30 000 rpm in an SW 41 rotor on a Beckman-Coulter Optima 190-K ultracentrifuge (Beckman-Coulter, Brea, CA, USA) for 1 hour at 12 °C. Each sample separated into a pale layer near the top of the tube, a small, dark pellet at the bottom of the tube, and a brown layer slightly above and incompletely separated from the pellet. The pellet consisted of spores and sand particles, the brown layer was nearly pure spores, and the top pale layer was mainly cellular debris with a few spores. The spore layers were extracted and the spin procedure repeated twice more. Clean spore fractions from both chitons were pooled for subsequent wash steps. In order to remove Percoll™, then surface bacteria, then salts, the spore sample was washed twice in 0.15M NaCl, once in 10% SDS, and twice in dH₂O, with spores collected after each wash by benchtop centrifugation at 12 000 g for 10 minutes (Eppendorf, Hamburg, Germany).

RNA extraction and whole transcriptome amplification. Clean spores were resuspended in 1 mL DNase buffer and treated with 2µL DNase I (Invitrogen, Carlsbad, CA, USA) at room temperature for 25 minutes. Spores were collected by centrifugation and resuspended in 0.5 mL lysis buffer (10mM Tris HcL pH 8, 0.1M EDTA pH 8, 0.5% w/v SDS) and incubated with 5µL of 20 mg/mL proteinase K (Bioshop, Burlington, ON, Canada) overnight at 50°C. Pellet density was estimated to be 4000 spores per µL. For RNA extraction using the Nucleospin RNA XS kit (Macherey-Nagel, Düren, Germany) approximately 800,000 spores were repeatedly frozen in liquid nitrogen and ground by hand in buffer RA1 with 4µL TCEP in a small tube-shaped glass mortar for 1 hour. Subsequent steps were performed according to the manufacturer's protocol, without use of carrier RNA, and including the filtration step. Total RNA was eluted in 10 µL elution buffer. Agarose gel electrophoresis of 2µL of the RNA sample did not reveal detectable bands after staining with SYBR green II for RNA, though both NEB ssRNA ladder (New England Biolabs, Ipswich, MA, USA) and Ambion RNA 6000 ladder (Life Technologies, Grand Road, NY, USA) were clearly visualized. The remaining 8µL of RNA sample were subjected to whole transcriptome amplification using the Sigma WTA kit (Sigma, St. Louis, MO, USA). 40µL of the amplified cDNA sample was ethanol-

precipitated (100 μ L 100% EtOH and 4 μ L 3M NaOAc pH 5.2) and resuspended in dH₂O. 10 μ L of 100 ng/ μ L amplified cDNA was submitted to NCGR for sequencing. Project ID: MMETSP0186 (<http://data.imicrobe.us/>).

The full materials and methods for the library construction and preliminary analyses of the transcriptome from *Mikrocytos mackini* can be found in Burki *et al.*, 2013; and for *Gromia sphaerica*, *Nonionellina* sp., *Globobulimina turgida*, *Elphidium margaritaceum*, *Bulimina marginata*, *Ammonia* sp., *Collozoum* sp., *Spongospaera streptacantha*, *Phyllostaurus siculus*, *Amphilonche elongata*, *Astrolonche serrata* and *Aulacantha scolymantha* in Sierra *et al.*, 2012.

References

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

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658-1659.

Mariette J, Escudie F, Allias N, Salin G, Noirot C, Thomas S, Klopp C. 2012. NG6: Integrated next generation sequencing storage and processing environment. *BMC Genomics* 13:462.

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, et al. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19:651-652.

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28:1086-1092.

Schwelm A, Fogelqvist J, Knaust A, Jülke S, Lilja TB-R, G., Karlsson M, Shevchenko A, Dhandapani V, Choi SR, Kim HG, et al. 2015. The *Plasmodiophora brassicae* genome reveals insights in its life cycle and ancestry of chitin synthases. *Sci Rep* 5.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821-829.