Molecular Phylogenetics and Evolution

Physiological diversity and adaptation of Rhizaria revealed by phylogenomics and comparative transcriptomics

--Manuscript Draft--

Manuscript Number:	MPE-D-25-00347
Article Type:	Research Paper
Keywords:	protists; functional traits; microbial ecology; multi-gene phylogeny; parasites; metabolism
Corresponding Author:	Jule Freudenthal, Ph.D. University of Koblenz Koblenz, GERMANY
First Author:	Jule Freudenthal, Ph.D.
Order of Authors:	Jule Freudenthal, Ph.D.
	Michael Bonkowski, Ph.D.
	Nele-Joelle Maria Engelmann, B.Sc.
	Martin Schlegel, Ph.D.
	Kenneth Dumack, Ph.D.
Abstract:	Protists are vastly diverse, forming over 20 supergroups of the eukaryotic diversity and fulfilling plentiful functions. Rhizaria is a widespread and highly abundant supergroup comprising important parasites and a huge diversity of free-living heterotrophic predators. Despite the diversity and biogeochemical importance of Rhizaria, our understanding of their physiology and metabolic capabilities remains limited, mainly due to a general lack of data and bioinformatic tools for cross-species comparisons of physiological traits. In this study, we assembled a total of 15 transcriptomes of the parasite-related bacterivorous Rhogostoma and their eukaryvorous relatives. By phylogenomic analyses and whole transcriptome comparison, we established an evolutionary framework to which we relate physiological traits. The morphologically highly similar Rhogostoma strains branch in two distinct clusters differing in orthogroups and gene expression patterns related to cell adhesion and biofilm formation. Furthermore, we reveal considerable intra-genus variation in amino acid and lipid metabolism, which might be explained by an ancient streamlining through gradual specialization to parasitism, bearing the potential for subsequent metabolic radiation. We conclude that even closely related and morphologically similar species in Rhizaria may differ distinctly in their functional repertoire. With the here established and showcased analyses, we create a basis for future characterization of the physiological traits of microeukaryotes.
Opposed Reviewers:	Micah Dunthorn, PhD Professor, University of Oslo Natural History Museum micah.dunthorn@nhm.uio.no He is an expert in SAR parasites.
	Fabien Burki, PhD Associate Professor, Uppsala University Department of Organismal Biology fabien.burki@ebc.uu.se He is an expert in Rhizaria evolution.
	Guillaume Lentendu, PhD University of Neuchâtel guillaume.lentendu@unine.ch He is an expert in protist ecology.
	David Singer, PhD HES-SO of Changins david.singer@changins.ch He is an expert in protist ecology.

	Jürgen Strassert, PhD Leibniz Institute of Freshwater Ecology and Inland Fisheries juergen.strassert@igb-berlin.de He is an expert in protist evolution.
	Lindsay Triplett, PhD Professor, The Connecticut Agricultural Experiment Station Lindsay.Triplett@ct.gov She is an expert in protist ecology.

Dr. Jule Freudenthal
Aquatic Ecosystem Analyses
Institute for Integrated Natural Sciences
University of Koblenz
Universitätsstraße 1
56070 Koblenz, Germany
jule.freudenthal@uni-koblenz.de

Molecular Phylogenetics and Evolution
Editorial Office

Koblenz, 12th of August, 2025

Dear Editor,

I am pleased to submit our manuscript entitled "*Physiological diversity and adaptation of Rhizaria revealed by phylogenomics and comparative transcriptomics*" for publication in Molecular Phylogenetics and Evolution.

Rhizaria are among the most diverse but least understood groups of microbial eukaryotes (protists), comprising a variety of free-living and heterotrophic taxa, such as the genus *Rhogostoma*, which is highly abundant across terrestrial ecosystems. Despite their ecological importance, Rhizaria remain poorly understood and underrepresented in molecular studies.

In this study, we address this knowledge gap by providing 15 novel, high-quality transcriptomes of bacterivorous *Rhogostoma* and their eukaryvorous relatives. In addition, we introduce a combined approach of phylogenomic analysis and cross-species transcriptome comparison to create an evolutionary framework to which we relate physiological traits. We show that morphologically highly similar *Rhogostoma* strains form two evolutionarily distinct clusters, differing in orthogroups and gene expression patterns related to cell adhesion and biofilm formation, as well as intra-genus variation in amino acid and lipid metabolism, indicating metabolic radiation. These findings revealed distinct functional diversity, even within closely related Rhizaria, which emphasises the need for future studies. The analyses established and showcased in this study pave the way for future research on the physiology of microeukaryotes.

We confirm that this manuscript has not been published or submitted elsewhere, except as part of my PhD thesis, which is publicly available in accordance with the institutional requirements of the University of Cologne (https://kups.ub.uni-koeln.de/78069/). We would greatly appreciate having the manuscript considered for publication in Molecular Phylogenetics and Evolution.

Sincerely, on behalf of the authors,

Jule Freudenthal

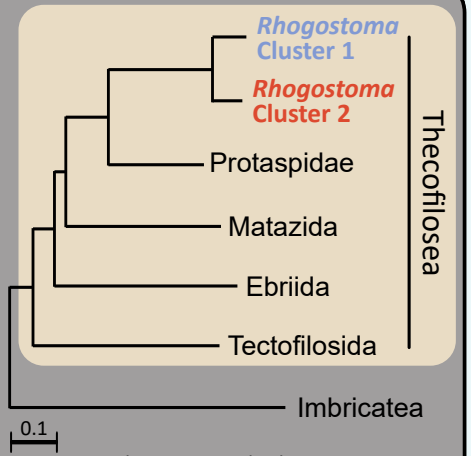
Highlights

- Phylogenomics resolves deep evolutionary relationships of Thecofilosea
- Physiological divergence suggests ecological flexibility via gene loss and gain
- Comparative transcriptomics reveal niche shifts in morphologically conserved genus
- Adhesion, sensing, and biofilm genes indicate prey- and habitat-driven selection
- High intra-genus variation may reflect streamlining and reversion from parasitism

Graphical Abstract

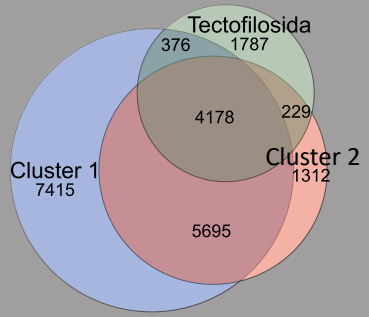
Physiological diversity and adaptation of Rhizaria

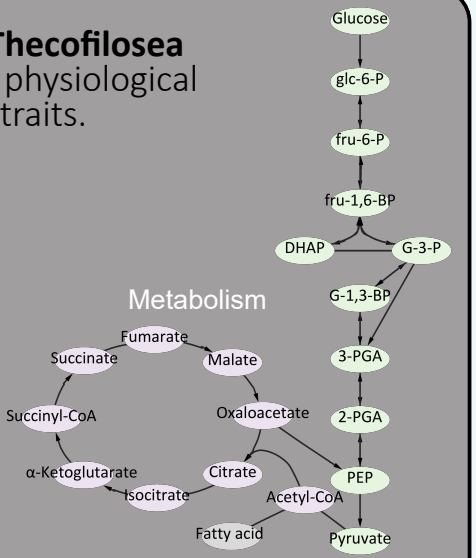
revealed by phylogenomics and comparative transcriptomics

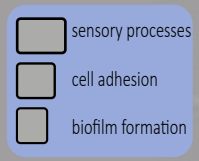


Multi-gene phylogeny resolves deep evolutionary relationships within **Thecofilosea**

Even closely related **Thecofilosea** strains exhibit distinct physiological and metabolic traits.







Cross-species differential gene expression & enrichment analysis indicate niche functional differentiation within morphologically conserved ***Rhogostoma***

Title: Physiological diversity and adaptation of Rhizaria revealed by phylogenomics and comparative transcriptomics

Authors :

Jule Freudenthal[a], Michael Bonkowski[b], Nele-Joelle Maria Engelmann[b], Martin Schlegel[c,d] and Kenneth Dumack[a]

[a]Aquatic Ecosystem Analyses, Institute for Integrated Natural Sciences, University of Koblenz, Universitätsstraße 1, 56070 Koblenz, Germany

[b]Terrestrial Ecology, Institute of Zoology, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Zülpicher Str. 47b, 50674 Köln, Germany

[c]Biodiversity and Evolution, Institute of Biology, University Leipzig, Talstraße 33, 04103 Leipzig, Germany

[d]German Centre for Integrative Biodiversity Research (iDiv) Halle Jena Leipzig, Puschstraße 4, 04103 Leipzig, Germany

Email: jule.freudenthal@uni-koblenz.de, m.bonkowski@uni-koeln.de, nengelm2@uni-koeln.de, schlegel@uni-leipzig.de, kenneth.dumack@uni-koblenz.de

Corresponding authors: Jule Freudenthal[1] and Kenneth Dumack[2]

[1] +49 261 287 - 2389

[2] +49 261 287 - 2380

Abstract

Protists are vastly diverse, forming over 20 supergroups of the eukaryotic diversity and fulfilling plentiful functions. Rhizaria is a widespread and highly abundant supergroup comprising important parasites and a huge diversity of free-living heterotrophic predators. Despite the diversity and biogeochemical importance of Rhizaria, our understanding of their physiology and metabolic capabilities remains limited, mainly due to a general lack of data and bioinformatic tools for cross-species comparisons of physiological traits. In this study, we assembled a total of 15 transcriptomes of the parasite-related bacterivorous *Rhogostoma* and their eukaryvorous relatives. By phylogenomic analyses and whole transcriptome comparison, we established an evolutionary framework to which we relate physiological traits. The morphologically highly similar *Rhogostoma* strains branch in two distinct clusters differing in orthogroups and gene expression patterns related to cell adhesion and biofilm formation. Furthermore, we reveal considerable intra-genus variation in amino acid and lipid metabolism, which might be explained by an ancient streamlining through gradual specialization to parasitism, bearing the potential for subsequent metabolic radiation. We conclude that even closely related and morphologically similar species in Rhizaria may differ distinctly in their functional repertoire. With the here established and showcased analyses, we create a basis for future characterization of the physiological traits of microeukaryotes.

Keywords

protists, functional traits, microbial ecology, multi-gene phylogeny, parasites, metabolism

1. Introduction

The Rhizaria are among the least understood but most diverse microbial eukaryotes. Rhizaria exhibits a high morphological and ecological diversity, including (1) Foraminifera, Radiolaria, and Phaeodaria, which are important oceanic predators; (2) phototrophic taxa (algae), such as *Paulinella chromatophora* and Chlorarachniophytes, which are crucial for our understanding of the evolution of photosynthesis; and (3) numerous parasitic species, such as *Plasmodiophora brassicae*, which causes about 15% of cabbage yields losses worldwide (Cavalier-Smith et al., 2018; Dumack et al., 2020; Nakamura and Suzuki, 2015; Neuhauser et al., 2010; Nowack, 2014). In addition, Rhizaria includes a variety of free-living and heterotrophic species, many of which exhibit an enormous species richness and genetic diversity (Bass et al., 2009; Flues et al., 2018; Howe et al., 2009).

Thecofilosea is one of the most abundant Rhizaria phyla across ecosystems, for example, in soils and the polar oceans, as shown by global surveys (Oliverio et al., 2020; Sommeria-Klein et al., 2021). Thecofilosea comprise a remarkable ecological and morphological diversity, including bacterivores, eukaryvores, and specialized, parasite-like predators of algae (Dumack et al., 2020). Among the Thecofilosea, free-living and heterotrophic taxa of the genus *Rhogostoma* (Rhogostomidae) are most abundant in terrestrial ecosystems (Öztoprak et al., 2020; Walden et al., 2021). *Rhogostoma* viciously preys on bacteria, whereas most of its relatives feed on eukaryotes (Seppey et al., 2017). Furthermore, *Rhogostoma* exhibits remarkable 18S marker gene variation but a highly conserved morphology, indicating ongoing cryptic speciation with a hidden diversity that remains to be fully uncovered (Öztoprak et al., 2020).

Despite the widespread occurrence and high abundance of Rhizaria, and *Rhogostoma* in particular, we still lack a well-supported phylogeny of the Rhogostomidae as single- or few-gene phylogenies, in contrast to many other microbial taxa, do not resolve most interspecific phylogenetic relationships (Cavalier-Smith et al., 2018). Furthermore, the transcriptome representation of Rhizaria is scarce (Sibbald and Archibald, 2017), and their physiology and functional diversity remain to be explored. With modern high-throughput sequencing techniques and bioinformatics tools, it is now feasible to

explore the physiological traits of individual species in an evolutionary context, which in turn will shed light on their potential ecological impact (Gerbracht et al., 2022; Ribeiro et al., 2020).

In this study, we explore the phylogenetic relationships and physiological traits of Thecofilosea. We present a compiled data set of 12 novel *Rhogostoma* transcriptomes, two Tectofilosida transcriptomes, and one Ebriida sp. transcriptome. Using a multi-gene phylogenomic approach, we resolve the phylogenetic backbone of Thecofilosea, in particular of *Rhogostoma*. Additionally, we perform whole transcriptome comparisons, along with functional annotation and enrichment analyses, to explore the inter- and intraspecific diversity in the physiological traits of the Thecofilosea in an evolutionary context.

2. Methods

2.1 RNA Extraction

We extracted total RNA of monoclonal cultures of 11 *Rhogostoma* strains previously described and cultured by Öztoprak et al. (2020), Pohl et al. (2021), and Martínez Rendón et al. (2025), along with *Katarium polorum* (Solbach et al., 2025). Additionally, one *Rhogostoma* strain was isolated and cultured from a sample originating from the Leipzig floodplain forest (51.3657 N, 12.3094 E) in Germany by isolating single cells using sterile glass micropipettes and culturing them in cell culture flasks (SARSTEDT AG & Co. KG T25; Nümbrecht, Germany) with wheat grass (WG)-medium, at temperatures ranging from 4 to 21°C.

For each RNA extraction from the cultures, cells were detached from the bottles by vigorous shaking and/or thorough scraping with a sterile cell scraper. Subsequently, the medium was filtered with a filter pore size of 3 µm (cellulose nitrate membranes, Whatman™, Buckinghamshire, United Kingdom) until the filter was completely covered in cells. The filter was transferred to a 1.5 ml tube (SARSTEDT AG & Co. KG, Nümbrecht, Germany) containing 1 ml ice-cold Sørensen buffer. The tube was vigorously shaken and centrifuged at 1500 rpm for 5 minutes at 4°C to detach the cells from the filter. The filter was carefully removed without disturbing the pellet, followed by an additional centrifugation step for

2 minutes to firm the pellet. The Sørensen buffer was replaced with 1 ml of clean Sørensen buffer and the tube was centrifuged again for 5 min. Finally, the Sørensen buffer was discarded, and 170 µl of ice-cold RLN buffer was added.

RNA extraction was carried out using the RNeasy® Plant Mini Kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's instructions starting from step 2 and using only 300 µl ethanol in step 4. RNA concentrations were quantified using the Qubit RNA High Sensitivity Assay Kit (Thermo Fisher Scientific Inc, Germany) and Qubit 4 Fluorometer (Thermo Fisher Scientific Inc, Germany). Sequencing was performed on an Illumina NovaSeq instrument (Illumina Inc., San Diego, CA, USA) at the Cologne Center for Genomics (Köln, Germany) with 2 × 100 bp paired-end reads, polyA selection, and a sequencing depth of about 50 Mio sequences (see Supplementary Table 1).

For single-cell RNA extraction of Ebriida sp., we followed the protocol of Hagemann-Jensen et al. (2020). Single cells of Ebriida sp. were isolated from samples originating from North Slope, Alaska (71.404558 N, 156.530021 W) and grown in F4 medium with nitschoid diatoms as prey. Before RNA extraction, cells were starved for 24 hours and given in the lysis buffer with a micromanipulator. Three replicates were pooled and sequenced with an Illumina NextSeq sequencer (Illumina Inc., San Diego, CA, USA) at the Cologne Center for Genomics (Köln, Germany) with 2 × 100 bp paired-end reads, polyA selection, and a sequencing depth of about 50 Mio sequences as well (see Supplementary Table 1).

2.2 Transcriptome assembly

The quality of the 15 newly sequenced transcriptomes (*Rhogostoma*, *Katarium polorum*, and Ebriida sp.) as well as of *Fisculla terrestris* (Gao et al., 2025; Solbach et al., 2021) and the two Protaspidae strains provided by Gordon Lax was assessed using FastQC v. 0.11.9 (Andrews, 2010), followed by RNA-seq error corrections with Rcorrector v 1.0.6 (Song and Florea, 2015) and quality filtering and adapter trimming with FastP v 0.23.2 (Chen et al., 2018). Potential contaminations from prokaryotes, plants, fungi, and humans were excluded using Kraken2 v 2.1.2 (Wood et al., 2019). Ribosomal RNA reads were identified using SortMeRNA v 4.3.4 (Kopylova et al., 2012) and blasted against the PR^2 database v. 4.14.0 (Guillou et al., 2013) using BLASTN v. 2.10.0 (Camacho et al., 2009) to confirm any

Rhogostoma/Thecofilosea sequences. The messenger RNA reads were assembled using Trinity v 2.14.0 (Grabherr et al., 2011), and candidate coding regions were identified using TransDecoder v 5.5.0 (Haas, 2018). Subsequently, the RNA-Seq read representation of the Trinity transcripts was validated using Bowtie2 v 2.4.1 (Langmead et al., 2019; Langmead and Salzberg, 2012).

We extended our data set by including 28 rhizarian transcriptomes/genomes (Balzano et al., 2015; Burki et al., 2013; Gerbracht et al., 2022; Gomaa et al., 2021; Grant et al., 2012; Keeling et al., 2014; Lhee et al., 2021; Rodríguez-Ezpeleta et al., 2007; Schwelm et al., 2015; Sierra et al., 2016) (see Supplementary Table 2). Candidate coding regions of nucleotide assemblies were identified using TransDecoder v 5.5.0 (Haas, 2018). The completeness of all assembled transcriptomes was evaluated with benchmarking universal single-copy orthologs (BUSCO) v 5.2.2 (Manni et al., 2021a, 2021b) and the eukaryote odb10 database.

2.3 Phylogenomic analyses

PhyloFisher v. 1.2.13 (Tice et al., 2021) was employed to identify orthologs for the muti-gene phylogeny of the Rhizaria, based on the provided database comprising 240 genes from 304 taxa across the eukaryotic tree of life. To screen for paralogs and contaminants, single gene trees of all 2490 genes were revised manually and using a customized R script. The cleaned data sets were filtered for Rhizaria transcriptomes, considering only taxa with more than 65 % amino acid coverage across all 240 genes. Exceptions were the single-cell transcriptomes of Protaspidae sp. and Ebriida sp., which had overall low gene coverage but were still included in the phylogenomic analysis. Further, the 240 genes were filtered, keeping only genes present in more than 50% of the taxa. A concatenated alignment consisting of 222 genes spanning 40 Rhizaria strains was built using PhyloFisher v. 1.2.13 (Tice et al., 2021).

The maximum likelihood (ML) tree was constructed with IQ-Tree v. 2.1.2 (Minh et al., 2020) using the site-heterogeneous mixture model LG + C60 + F + Γ and 1000 Ultra-Fast Bootstrap replicates (UFB) as well as SH-like approximate likelihood ratio test (SH-aLRT).

2.4 Transcriptome annotation and comparison

OrthoFinder v. 2.5.2 (Emms and Kelly, 2019, 2015) was employed to identify phylogenetic hierarchical orthogroups across all *Rhogostoma* strains, *Fisculla terrestris*, and *Katarium polorum*. Beforehand, the redundancy of the Trinity contigs was reduced by clustering at 95% identity over at least 90% of the shorter contig length using CD-HIT v. 4.8.1 (Fu et al., 2012). A customized R script was employed to filter the TransDecoder protein sequences by the clustered Trinity contigs and to split TransDecoder protein sequences with multiple predicted coding regions. In the case of overlapping coding regions, the longer one was retained. To verify the quality of the clustered contigs, we checked their completeness using BUSCO v 5.2.2 (Manni et al., 2021a, 2021b) and the eukaryotic odb10 database, and the RNA-Seq read mapping using Bowtie2 v 2.4.1 (Langmead et al., 2019; Langmead and Salzberg, 2012). Further, the gene expression was quantified with Salmon v 1.9.0 (Patro et al., 2017) in alignment-based mode. The protein sequences were functionally annotated based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa, 2019; Kanehisa et al., 2023; Kanehisa and Goto, 2000) using eggNOG-mapper v. 2.1.9 (Cantalapiedra et al., 2021).

The number of shared and unique orthogroups across all strains was visualized using ggupset v. 0.4.0 (Ahlmann-Eltze, 2024). Only the top 25 combinations with the highest number of orthogroups were considered. In addition, a general overview of the relative proportions of functional annotations from 26 categories was provided by bar charts visualized with ggplot2 (Wickham, 2011). The relative proportions of functional annotations were calculated for all species, the core set of orthogroups (i.e., orthogroups shared by all strains), the two *Rhogostoma* clusters, *Katarium polorum* and *Fisculla terrestris*, respectively.

A detailed overview of the metabolic pathways of the amino acid, nucleotide, carbohydrate, and lipid metabolism was created based on the KEGG module. For this, the presence or absence of each KEGG orthology (KO) term was assessed per KEGG module, the minimum KO terms required for each module were determined, and the percentage of KO terms for each reaction step (the number of preset KO terms divided by the minimum number of KO terms required for the respective reaction) was calculated. Only KEGG modules for which (for at least one strain) 50% or more of the KO terms were

present and no more than a total of three KO terms were missing were considered for the calculation of the heatmaps and the summarisation of the amino acid, central carbohydrate, and lipid metabolism modules into a graph. The graph was visualized using Cytoscape v. 3.9.0 (Shannon et al., 2003).

For the gene enrichment analysis of the two phylogenetically distinct *Rhogostoma* clusters multiple steps were conducted: First, the quantified gene expressions from salmon were summarized for each strain and orthogroup using tximport v. 1.32.0 (Soneson et al., 2016), generating matrices containing the weighted mean of the contig length, the effective contig length, the number of reads (counts) and the transcript per million (TPM) for each orthogroup. Second, low count orthogroups were excluded, keeping only orthogroups with counts per million > 1 in six or more strains. Third, principal component analysis (PCA) based on the filtered expression level of the orthogroup was calculated using the functions vst and plotPCA (DESeq2 v. 1.44.0; (Love et al., 2014)). Fourth, differential expression analysis was conducted with DESeq2 v. 1.44.0 (Love et al., 2014) comparing the two *Rhogostoma* clusters with six strains each. Orthogroups with |log2 fold change| ≥ 1 and adjusted p-value < 0.001 were considered as differentially expressed. The enrichment analysis of KO terms was carried out using GOseq v. 1.56.0 (Young et al., 2010). The enrichment analysis was performed for all 12 Rhogostoma strains, using the length and the functional assignment of the orthogroups of each strain, respectively. Only KO terms that were significantly enriched in the enrichment analysis of all strains within one Rhogostoma cluster were retained.

3. Results

3.1 Phylogenomic analysis

To shed light on the evolutionary relationships of Thecofilosea, we conducted a phylogenetic analysis based on a comprehensive multi-gene data set of 222 concatenated genes (71,479 amino acids) derived from, in total, 40 Rhizaria transcriptomes (Fig. 1, Supplementary Table 2, Supplementary Fig. 1 & 2). Among these transcriptomes are a total of 15 newly added Thecofilosea transcriptomes: 14 derived from monoclonal cultures with high gene and site coverage - 12 *Rhogostoma* (Cryomonadida)

and two Tectofilosida transcriptomes, as well as one single-cell transcriptome of Ebriida sp. with moderate gene and site coverage, as it is typical for single-cell transcriptomes (Fig. 1, Supplementary Table 1).

The phylogenetic analyses provided a highly supported backbone of the Thecofilosea, Imbricatea, Sarcomonadea and Helkesea (Fig. 1). At the base of the Thecofilosea, *Fisculla terrestris* (Fiscullidae) and *Katarium polorum* (Chlamydophryidae) form a fully supported monophyletic group, the Tectofilosida. The Ebriida sp. branch with high support basal to Matazida. Within Cryomonadida, two distinct, fully supported monophyletic subclades were identified. The first subclade included all Protaspidae strains and clustered at the base of the second subclade, including all *Rhogostoma* strains. The *Rhogostoma* strains were further divided into two fully supported clusters, each comprising six strains: Cluster 1, characterized by short branches, included *R. kyoshii*, *R. epiphylla*, *R. minus*, *R. kappa* and *R. tahiri*, and cluster 2, with longer branches, included *R. florea*, *R. pseudocylindrica* and two strains each of *R. schuessleri* and unidentified *Rhogostoma* species.

3.2 Comparative transcriptomics

To explore the physiological traits of free-living the heterotrophic Thecofilosea, we compared and functionally annotated the whole transcriptomes of Thecofilosea species based on the KEGG database. Only the culture-based and thus deeply sequenced transcriptomes of *Rhogostoma*, *Fisculla*, and *Katarium* were considered, not the single-cell transcriptome of Ebriida sp. due to the comparably lower coverage of single-cell transcriptomes (Supplementary Table 1, Supplementary Figure 3). The transcriptomes comprised an average of 67,766 clustered contigs, with a mean mapping rate of 95% when aligning the reads back to the assembled contigs and a mean completeness of 86% according to benchmarking universal single-copy orthologs (BUSCO) of the Eukaryota dataset (Fig. 2B). Further, these transcriptomes contained an average of 44,972 predicted open reading frames (ORFs, Supplementary Table 1).

Overall, 93% of the ORFs were assigned to 58,676 orthogroups (Fig. 2A). We identified a core set of 4,178 orthogroups that was shared by all investigated Thecofilosea species (Fig. 2A). Compared to the functional annotations of all orthogroups, the core orthogroups exhibited a higher relative proportion of KO terms related to genetic information processing (Fig. 2C), in particular, related to replication and repair, transcription and translation. The two phylogenetically distinct *Rhogostoma* clusters showed equally distinct sets of orthogroups. Cluster 1 shared a higher number of unique orthogroups (4,612) than cluster 2 (704, Fig. 2A). Further, cluster 1 comprised a higher proportion of KO terms related to signal transduction, sensory systems, and environmental adaptation. In contrast, Cluster 2 comprised a higher proportion of KO terms related to transport and catabolism compared to the overall functional annotation (Fig. 2C). Notably, both Tectofilosida species, *Fisculla terrestris* and *Katarium polorum* exhibited high numbers of unique orthogroups, 5,134 and 3,417, respectively (Fig. 2A). Both species showed a higher relative proportion of KO terms involved in cell motility. *Fisculla terrestris* additionally showed a relative increase of KO terms associated with cellular community processes such as focal adhesion, adherens junctions, tight junctions, and gap junctions compared to the overall functional annotation (Fig. 2C).

To gain insights into the metabolic adaptations of these free-living, heterotrophic Thecofilosea strains, we analyzed KEGG modules involved in amino acid, carbohydrate, nucleotide, and lipid metabolism (Fig. 3) and reconstructed the pathways for central carbohydrate and amino acid metabolism (Fig. 4 & 5, Supplementary Figures 4-29). The majority of genes involved in the central carbohydrate metabolism were expressed, although the complete pyruvate oxidation pathway could only be reconstructed for 6 out of 14 strains. The *Rhogostoma* cluster 2 lacked few genes for the non-oxidative pentose phosphate pathway. Furthermore, most of the genes involved in nucleotide biosynthesis were present throughout the dataset, except for genes related to the *de novo* purine biosynthesis, which were exclusively found in *Katarium polorum*. In contrast to the relatively conserved central carbohydrate and nucleotide metabolisms, greater variability was observed in amino acid and lipid metabolism. All strains expressed most genes involved in the biosynthesis of proline, valine, isoleucine,

cysteine, methionine, serine, threonine, glutamate, and glycine. However, only *Fisculla terrestris* and *Katarium polorum* appeared to synthesize arginine, tryptophan, histidine, and lysine. *Fisculla terrestris* additionally possessed most genes to synthesize leucine, phenylalanine, and tyrosine. Differences in the expression of genes involved in lipid metabolism between the strains were detected, particularly genes involved in sterol biosynthesis. For example, strains belonging to the *Rhogostoma* Cluster 2 expressed a greater number of enzymes involved in the biosynthesis of C18/19/21-steroids.

3.3 Enrichment analysis

In addition to the analyses of presence-absence data, we compared the relative expression patterns of the two *Rhogostoma* clusters based on 20,000 orthogroups. A PCA showed a clear differentiation of *Rhogostoma* clusters 1 and 2, explaining 82% of the total variation (Fig. 6A). In addition, differences in the expression patterns of the long-branched, i.e., evolutionary distant *Rhogostoma* cluster 2 (Fig. 1) explained 5% of the total variation, with strains of the same species grouping together.

Differential expression analysis revealed 7,502 out of the 20,000 orthogroups to be differentially expressed. About one-third of these orthogroups could be functionally annotated. A subsequent enrichment analysis identified a total of 15 KO terms that were consistently enriched across all six species of *Rhogostoma* cluster 1 (Fig. 6B). In contrast, no KO terms were consistently enriched across all six species of *Rhogostoma* cluster 2. ANKRD28 and ANKRD44 (Ankyrin repeat domains 28 and 44) were the most frequent KO terms of the enriched KO terms of *Rhogostoma* cluster 1, with 84 and 66 counts, respectively. ANKRD28 and ANKRD44 occurred almost exclusively (~90%) in the significantly enriched orthogroups (hit percentage, Fig. 6B). Furthermore, ata/sadA/emaA (trimeric autotransporter adhesin), ALS (agglutinin-like protein), MUC13 (mucin 13), and TRPA1/ANKTM1 (transient receptor potential cation channel subfamily A member 1) were identified among the enriched KO terms of *Rhogostoma* cluster 1. They also exhibited a high prevalence, with 26, 18, 18,

and 17 counts, respectively, and ~80% hit percentage in the significantly enriched orthogroups (Fig. 6B).

4. Discussion

4.1 The physiological capacity of Thecofilosea (Cecozoa)

We identified a core set of conserved genes and pathways that were shared across all investigated Thecofilosea species. First, we identified a core set of shared orthogroups that included an exceptionally high proportion of genes associated with genetic information processing, i.e., translation, transcription, and replication. As these genes are essential for all living cells and are known to be highly conserved (Yao and O'Donnell, 2016), it was expected that these genes would be overrepresented in the core set of orthogroups. Second, we show an overall high completeness in the central carbohydrate and nucleotide metabolism, including the presence of glycolysis, the citrate cycle, and the capacity to synthesize all nucleotides. The few missing enzymes can most likely be explained by differences in the homologs due to rapid evolutionary rates or incompleteness of the transcriptome data. Overall, the robust identification of conserved genes and pathways emphasizes the soundness of our methods and data.

We were able to reconstruct the *de novo* synthesis for at least nine amino acids for all investigated Thecofilosea species. However, we also found variations in the physiological traits of the species. Heterotrophic microorganisms usually lack pathways for synthesizing certain amino acids and consequently depend on salvaging them from their prey. For example, the predatory amoebae *Dictyostelium* and *Arcella* have lost the capability to synthesize 11 and 5 amino acids, respectively (Payne and Loomis, 2006; Ribeiro et al., 2020). Our data indicate the absence of numerous amino acid pathways, for instance, tryptophan and histidine, particularly in *Rhogostoma* – indicating their dependence on the uptake of these compounds from their prey.

4.2 Physiology in an evolutionary context

4.2.1 Multi-gene phylogeny

By incorporating the new Thecofilosea transcriptomes into existing public Rhizaria data, our phylogenetic analysis expands the latest Cercozoa multi-gene phylogeny of Irwin et al. (2019). In our phylogenetic tree, the superclass Ventrifilosa is highly supported – a hypothesized monophylum of the predominantly shell-bearing and free-living cecozoan groups Thecofilosea and Imbricatea (Cavalier-Smith and Karpov, 2012). We further confirm that the Tectofilosida are indeed monophyletic as suggested by Dumack et al. (2017). Lastly, the Sarcomonadea branch basal to Ventrifilosa, as indicated by numerous single-gene phylogenies (Howe et al., 2011; Scoble and Cavalier-Smith, 2014).

4.2.2 Orthogroup similarity reflects the phylogenetic distance

The high coverage and saturation of our culture-derived transcriptomic data allowed for whole transcriptome comparisons of physiological traits among Thecofilosea. The physiological capabilities based on nearly complete transcriptomes highly reflect evolutionary distance across all studied Thecofilosea strains. Within Tectofilosida, although only represented by two strains, the high evolutionary distance – indicated by long branches in the phylogenetic analyses – was reflected by the high number of unique orthogroups in each species and a moderate amount of shared orthogroups. Notably, *Fisculla terrestris* exhibited a higher proportion of unique orthogroups related to cell adherence, fusion, and cell-to-cell communication, likely reflecting its nature of frequent fusion (Gao et al., 2025).

Rhogostoma radiated into two clusters with notable differences in both orthogroups and gene expression patterns. The short-branched, i.e., evolutionary close, *Rhogostoma* cluster 1 exhibited a high number of shared orthogroups and clustered closer together in the PCA, compared to the long-branched, i.e., evolutionary distant, *Rhogostoma* cluster 2. Remarkably, *Rhogostoma* cluster 1 showed a higher proportion of genes related to signal transduction, sensory systems, and environmental adaptation in the orthogroups unique to this cluster. In addition, differences in gene expression patterns revealed enrichment in genes in *Rhogostoma* cluster 1 associated with first, sensory

processes, including the reception of heat, pain, or environmental irritants (TRPA1/ANKTM1, Bautista et al., 2006), second protection and lubrication of cell surfaces (Muc13, Williams et al., 2001), third, cell migration and focal adhesion formation (ANKRD28, Tachibana et al., 2009), and fourth, cell adhesion to biotic and abiotic surfaces and biofilm formation (ata/sadA/emaA and ALS, Bentancor et al., 2012; Mintz, 2004; Oh et al., 2019; Raghunathan et al., 2011).

Aside from inter-cluster variation, there is additionally a large inter-specific variation in *Rhogostoma*, showcasing that even closely related and morphologically similar species exhibit distinct physiological traits and thus, potentially distinct ecological influence. It is important to note that although transcriptomic responses exhibit a snapshot of metabolic activity (Martin and Wang, 2011; Raghavan et al., 2022), we analyzed a high density of input cells reflecting the transcriptomic response of thousands of individuals, minimizing temporal and individual variation. To further increase the comparability, we grew all *Rhogostoma* strains in the same medium, strengthening our interpretation of the results.

4.2.3 Reduced gene set of *Rhogostoma*

In addition to a core set of orthogroups, each investigated evolutionary clade expressed unique transcripts, providing evidence that protistan functions in ecosystems cannot be easily generalized, even at a low taxonomic level. We show that *Rhogostoma*, a genus containing morphologically highly similar species (Öztoprak et al., 2020) and being the most derived clade in our phylogenetic tree, exhibited an exceptionally high variability in its physiological repertoire. The question arises, which selective forces led to the evolution of different physiological traits in species with highly similar morphologies?

As our phylogenetic tree shows, the Rhogostomidae (Cryomonadida) are closely related to the Protaspidae (Cryomonadida), highly specialized parasites of algae (Drebes et al., 1996; Schnepf and Kühn, 2000). The transition from a free-living to a parasitic lifestyle usually leads to the loss of existing functions and a simplified metabolism, as parasites exploit the resources of their host (Jackson et al., 2016; Poulin and Randhawa, 2015). Thus, the transition to parasitism is often thought to be

irreversible, yet several studies, for example on diplomonads, provide evidence for a secondary free-living lifestyle (Wiśniewska et al., 2024; Xu et al., 2016). We hypothesize that the ancestors of the Cryomonadida underwent a loss of functional and genomic diversity with the specialization towards parasitism and that the ancestor of the Rhogostomidae broadened the prior narrow prey spectrum and specialized on feeding on bacteria in addition to eukaryotes.

Horizontal gene transfer (HGT) is, in general, considered to be a minor driver of eukaryote evolution (Keeling, 2024, 2019). Instead, eukaryotes are considered mainly to evolve by gene duplication and the subsequent adaption of homologs to different functions. Nonetheless, there is evidence that HGT in microbial eukaryotes may represent an escape from increasing adaptation to parasitism, which typically involves a loss of genomic and functional diversity (Wiśniewska et al., 2024; Xu et al., 2016), i.e., that HGT may contribute to the reversibility of parasitism and readaptation to a free-living lifestyle through the acquisition of new functions.

Overall, the potential back-transition of *Rhogostoma* from parasitism to free-living might have caused the acquisition of even single functional genes to lead to a new speciation event, thus contributing to the current remarkable species diversity of *Rhogostoma*.

5. Conclusion

The outstanding high coverage of the transcriptomes and the deeply sampled phylogeny of the free-living and heterotrophic Thecofilosea allowed us to draw conclusions on their physiological capabilities and compare them in an evolutionary context. The Thecofilosea possess a core set of conserved genes involved in genetic information processing, nucleotide metabolism, and central carbohydrate metabolism. However, the high variability in amino acid and lipid metabolism, as well as the differentially expressed and enriched genes, indicate potentially distinct functional roles. These findings emphasize the remarkable physiological diversity even among closely related, morphologically highly similar taxa. Considering that the Thecofilosea represent only a fraction of the tremendous

protistan diversity, the need for further research is evident. Our methodological approach paves the way for subsequent studies on a larger scale.

6. Acknowledgements

We would like to thank Lukas Isenhart for carrying out most of the RNA extractions and Gordon Lax for providing the single-cell sequences of Protaspida, which will soon be publicly available. We furthermore thank the Regional Computing Center of the University of Cologne (RRZK) for providing computing time on the DFG-funded (Funding number: INST 216/512/1FUGG) High Performance Computing (HPC) system CHEOPS as well as support.

7. Declarations

7.1 Competing interests

The authors declare no competing interests.

8. References

Ahlmann-Eltze, C., 2024. ggupset: Combination Matrix Axis for "ggplot2" to Create "UpSet" Plots. R package version 0.4.0.

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data.

Balzano, S., Corre, E., Decelle, J., Sierra, R., Wincker, P., Da Silva, C., Poulain, J., Pawlowski, J., Not, F., 2015. Transcriptome analyses to investigate symbiotic relationships between marine protists. Front Microbiol 6. https://doi.org/10.3389/fmicb.2015.00098

Bass, D., Howe, A.T., Mylnikov, A.P., Vickerman, K., Chao, E.E., Edwards Smallbone, J., Snell, J., Cabral, C., Cavalier-Smith, T., 2009. Phylogeny and classification of Cercomonadida (Protozoa, Cercozoa): *Cercomonas*, *Eocercomonas*, *Paracercomonas*, and *Cavernomonas* gen. nov. Protist 160, 483–521. https://doi.org/10.1016/j.protis.2009.01.004

Bautista, D.M., Jordt, S.-E., Nikai, T., Tsuruda, P.R., Read, A.J., Poblete, J., Yamoah, E.N., Basbaum, A.I., Julius, D., 2006. TRPA1 mediates the inflammatory actions of environmental irritants and proalgesic agents. Cell 124, 1269–1282. https://doi.org/10.1016/j.cell.2006.02.023

Bentancor, L.V., Camacho-Peiro, A., Bozkurt-Guzel, C., Pier, G.B., Maira-Litrán, T., 2012. Identification of Ata, a multifunctional trimeric autotransporter of *Acinetobacter baumannii*. J Bacteriol 194, 3950–3960. https://doi.org/10.1128/jb.06769-11

Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., Keeling, P.J., 2013. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. Curr Biol 23, 1541–1547. https://doi.org/10.1016/j.cub.2013.06.033

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinforma 10, 421. https://doi.org/10.1186/1471-2105-10-421

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38, 5825–5829. https://doi.org/10.1093/molbev/msab293

Cavalier-Smith, T., Chao, E.E., Lewis, R., 2018. Multigene phylogeny and cell evolution of chromist infrakingdom Rhizaria: contrasting cell organisation of sister phyla Cercozoa and Retaria. Protoplasma 255, 1517–1574. https://doi.org/10.1007/s00709-018-1241-1

Cavalier-Smith, T., Karpov, S.A., 2012. *Paracercomonas* Kinetid Ultrastructure, Origins of the Body Plan of Cercomonadida, and Cytoskeleton Evolution in Cercozoa. Protist 163, 47–75. https://doi.org/10.1016/j.protis.2011.06.004

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinform 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Drebes, G., Kühn, S.F., Gmelch, A., Schnepf, E., 1996. *Cryothecomonas aestivalis* sp. nov., a colourless nanoflagellate feeding on the marine centric diatom *Guinardia delicatula* (Cleve) Hasle. Helgolander Meeresunters 50, 497–515. https://doi.org/10.1007/BF02367163

Dumack, K., Fiore-Donno, A.M., Bass, D., Bonkowski, M., 2020. Making sense of environmental sequencing data: Ecologically important functional traits of the protistan groups Cercozoa and Endomyxa (Rhizaria). Mol Ecol Resour 20, 398–403. https://doi.org/10.1111/1755-0998.13112

Dumack, K., Mausbach, P., Hegmann, M., Bonkowski, M., 2017. Polyphyly in the thecate amoeba genus *Lecythium* (Chlamydophryidae, Tectofilosida, Cercozoa), redescription of its type species *L. hyalinum*, description of *L. jennyae* sp. nov. and the establishment of *Fisculla* gen. nov. and Fiscullidae fam. nov. Protist 168, 294–310. https://doi.org/10.1016/j.protis.2017.03.003

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20, 238. https://doi.org/10.1186/s13059-019-1832-y

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16, 157. https://doi.org/10.1186/s13059-015-0721-2

Flues, S., Blokker, M., Dumack, K., Bonkowski, M., 2018. Diversity of cercomonad species in the phyllosphere and rhizosphere of different plant species with a description of *Neocercomonas epiphylla* (Cercozoa, Rhizaria) a leaf-associated protist. J Eukaryot Microbiol 65, 587–599. https://doi.org/10.1111/jeu.12503

Freudenthal, J., 2025. New approaches for investigating microbial food webs across ecosystems (text.thesis.doctoral). Universität zu Köln.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinform 28, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Gao, S., Solbach, M.D., Bast, J., Dumack, K., 2025. Meiosis-associated expression patterns during starvation-induced cell fusion in the protist *Fisculla terrestris*. BMC Biol 23, 140. https://doi.org/10.1186/s12915-025-02246-3

Gerbracht, J.V., Harding, T., Simpson, A.G.B., Roger, A.J., Hess, S., 2022. Comparative transcriptomics reveals the molecular toolkit used by an algivorous protist for cell wall perforation. Curr Biol S0960982222008521. https://doi.org/10.1016/j.cub.2022.05.049

Gomaa, F., Utter, D.R., Powers, C., Beaudoin, D.J., Edgcomb, V.P., Filipsson, H.L., Hansel, C.M., Wankel, S.D., Zhang, Y., Bernhard, J.M., 2021. Multiple integrated metabolic strategies allow foraminiferan protists to thrive in anoxic marine sediments. Sci Adv 7, eabf1586. https://doi.org/10.1126/sciadv.abf1586

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29, 644–652. https://doi.org/10.1038/nbt.1883

Grant, J.R., Lahr, D.J.G., Rey, F.E., Burleigh, J.G., Gordon, J.I., Knight, R., Molestina, R.E., Katz, L.A., 2012. Gene discovery from a pilot study of the transcriptomes from three diverse microbial

eukaryotes: *Corallomyxa tenera*, *Chilodonella uncinata*, and *Subulatomonas tetraspora*, in: Protist Genomics. https://doi.org/10.2478/prge-2012-0002

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R., 2013. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41, 597–604. https://doi.org/10.1093/nar/gks1160

Haas, B., 2018. TransDecoder [WWW Document]. URL https://github.com/TransDecoder/TransDecoder/wiki (accessed 3.12.24).

Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.-J., Larsson, A.J.M., Faridani, O.R., Sandberg, R., 2020. Smart-seq3 Protocol. protocols.io. https://doi.org/dx.doi.org/10.17504/protocols.io.bcq4ivyw

Howe, A.T., Bass, D., Scoble, J.M., Lewis, R., Vickerman, K., Arndt, H., Cavalier-Smith, T., 2011. Novel cultured protists identify deep-branching Environmental DNA clades of Cercozoa: New genera *Tremula*, *Micrometopion*, *Minimassisteria*, *Nudifila*, *Peregrinia*. Protist 162, 332–372. https://doi.org/10.1016/j.protis.2010.10.002

Howe, A.T., Bass, D., Vickerman, K., Chao, E.E., Cavalier-Smith, T., 2009. Phylogeny, taxonomy, and astounding genetic diversity of Glissomonadida ord. nov., the dominant gliding zooflagellates in soil (Protozoa: Cercozoa). Protist 160, 159–189. https://doi.org/10.1016/j.protis.2008.11.007

Irwin, N.A.T., Tikhonenkov, D.V., Hehenberger, E., Mylnikov, A.P., Burki, F., Keeling, P.J., 2019. Phylogenomics supports the monophyly of the Cercozoa. Mol Phylogenet and Evol 130, 416–423. https://doi.org/10.1016/j.ympev.2018.09.004

Jackson, A.P., Otto, T.D., Aslett, M., Armstrong, S.D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J.M., Dacks, J.B., Acosta-Serrano, A., Field, M.C., Ginger, M.L., Berriman, M., 2016. Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. Curr Biol 26, 161–172. https://doi.org/10.1016/j.cub.2015.11.055

Kanehisa, M., 2019. Toward understanding the origin and evolution of cellular organisms. Protein Sci 28, 1947–1951. https://doi.org/10.1002/pro.3715

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., Ishiguro-Watanabe, M., 2023. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res 51, D587–D592. https://doi.org/10.1093/nar/gkac963

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27–30. https://doi.org/10.1093/nar/28.1.27

Keeling, P.J., 2024. Horizontal gene transfer in eukaryotes: aligning theory with data. Nat Rev Genet 25, 416–430. https://doi.org/10.1038/s41576-023-00688-5

Keeling, P.J., 2019. Combining morphology, behaviour and genomics to understand the evolution and ecology of microbial eukaryotes. Philos Trans R Soc B Biol Sci 374, 20190085. https://doi.org/10.1098/rstb.2019.0085

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L.,

Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLOS Biol 12, e1001889. https://doi.org/10.1371/journal.pbio.1001889

Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinform 28, 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Wilks, C., Antonescu, V., Charles, R., 2019. Scaling read aligners to hundreds of threads on general-purpose processors. Bioinform 35, 421–432. https://doi.org/10.1093/bioinformatics/bty648

Lhee, D., Lee, J., Ettahi, K., Cho, C.H., Ha, J.-S., Chan, Y.-F., Zelzion, U., Stephens, T.G., Price, D.C., Gabr, A., Nowack, E.C.M., Bhattacharya, D., Yoon, H.S., 2021. Amoeba genome reveals dominant host contribution to plastid endosymbiosis. Mol Biol Evol 38, 344–357. https://doi.org/10.1093/molbev/msaa206

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021a. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol 38, 4647–4654. https://doi.org/10.1093/molbev/msab199

Manni, M., Berkeley, M.R., Seppey, M., Zdobnov, E.M., 2021b. BUSCO: Assessing genomic data quality and beyond. Curr Protoc 1, e323. https://doi.org/10.1002/cpz1.323

Martin, J.A., Wang, Z., 2011. Next-generation transcriptome assembly. Nat Rev Genet 12, 671–682. https://doi.org/10.1038/nrg3068

Martínez Rendón, C., Braun, C., Kappelsberger, M., Boy, J., Casanova-Katny, A., Glaser, K., Dumack, K., 2025. Enhancing microbial predator–prey detection with network and trait-based analyses. Microbiome 13, 37. https://doi.org/10.1186/s40168-025-02035-8

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mintz, K.P., 2004. Identification of an extracellular matrix protein adhesin, EmaA, which mediates the adhesion of *Actinobacillus actinomycetemcomitans* to collagen. Microbiology 150, 2677–2688. https://doi.org/10.1099/mic.0.27110-0

Nakamura, Y., Suzuki, N., 2015. Phaeodaria: Diverse marine cercozoans of world-wide distribution, in: Ohtsuka, S., Suzaki, T., Horiguchi, T., Suzuki, N., Not, F. (Eds.), Marine Protists: Diversity and Dynamics. Springer Japan, Tokyo, pp. 223–249. https://doi.org/10.1007/978-4-431-55130-0_9

Neuhauser, S., Bulman, S., Kirchmair, M., 2010. Plasmodiophorids: The challenge to understand soil-borne, obligate biotrophs with a multiphasic life cycle, in: Gherbawy, Y., Voigt, K. (Eds.), Molecular Identification of Fungi. Springer, Berlin, Heidelberg, pp. 51–78. https://doi.org/10.1007/978-3-642-05042-8_3

Nowack, E.C.M., 2014. *Paulinella chromatophora* – rethinking the transition from endosymbiont to organelle. Acta Soc Bot Pol 83, 387–397. https://doi.org/10.5586/asbp.2014.049

Oh, S.-H., Smith, B., Miller, A.N., Staker, B., Fields, C., Hernandez, A., Hoyer, L.L., 2019. Agglutinin-like sequence (ALS) genes in the *Candida parapsilosis* species complex: Blurring the boundaries between gene families that encode cell-wall proteins. Front Microbiol 10. https://doi.org/10.3389/fmicb.2019.00781

Oliverio, A.M., Geisen, S., Delgado-Baquerizo, M., Maestre, F.T., Turner, B.L., Fierer, N., 2020. The global-scale distributions of soil protists and their contributions to belowground systems. Sci Adv 6, eaax8787. https://doi.org/10.1126/sciadv.aax8787

Öztoprak, H., Walden, S., Heger, T., Bonkowski, M., Dumack, K., 2020. What drives the diversity of the most abundant terrestrial cercozoan family (Rhogostomidae, Cercozoa, Rhizaria)? Microorganisms 8, 1123. https://doi.org/10.3390/microorganisms8081123

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417–419. https://doi.org/10.1038/nmeth.4197

Payne, S.H., Loomis, W.F., 2006. Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. Eukaryot Cell 5, 272–276. https://doi.org/10.1128/ec.5.2.272-276.2006

Pohl, N., Solbach, M.D., Dumack, K., 2021. The wastewater protist *Rhogostoma minus* (Thecofilosea, Rhizaria) is abundant, widespread, and hosts Legionellales. Water Res 203, 117566. https://doi.org/10.1016/j.watres.2021.117566

Poulin, R., Randhawa, H.S., 2015. Evolution of parasitism along convergent lines: From ecology to genomics. Parasitology 142, S6–S15. https://doi.org/10.1017/S0031182013001674

Raghavan, V., Kraft, L., Mesny, F., Rigerte, L., 2022. A simple guide to *de novo* transcriptome assembly and annotation. Brief Bioinform 23, bbab563. https://doi.org/10.1093/bib/bbab563

Raghunathan, D., Wells, T.J., Morris, F.C., Shaw, R.K., Bobat, S., Peters, S.E., Paterson, G.K., Jensen, K.T., Leyton, D.L., Blair, J.M.A., Browning, D.F., Pravin, J., Flores-Langarica, A., Hitchcock, J.R., Moraes, C.T.P., Piazza, R.M.F., Maskell, D.J., Webber, M.A., May, R.C., MacLennan, C.A., Piddock, L.J., Cunningham, A.F., Henderson, I.R., 2011. SadA, a trimeric autotransporter from *Salmonella enterica* Serovar Typhimurium, can promote biofilm formation and provides limited protection against infection. Infect Immun 79, 4342–4352. https://doi.org/10.1128/iai.05592-11

Ribeiro, G.M., Porfírio-Sousa, A.L., Maurer-Alcalá, X.X., Katz, L.A., Lahr, D.J.G., 2020. De novo sequencing, assembly, and annotation of the transcriptome for the free-Living testate amoeba *Arcella intermedia*. J Eukaryot Microbiol 67, 383–392. https://doi.org/10.1111/jeu.12788

Rodríguez-Ezpeleta, N., Brinkmann, H., Burger, G., Roger, A.J., Gray, M.W., Philippe, H., Lang, B.F., 2007. Toward Resolving the Eukaryotic Tree: The Phylogenetic Positions of Jakobids and Cercozoans. Curr Biol 17, 1420–1425. https://doi.org/10.1016/j.cub.2007.07.036

Schnepf, E., Kühn, S.F., 2000. Food uptake and fine structure of *Cryothecomonas longipes* sp. nov., a marine nanoflagellate incertae sedis feeding phagotrophically on large diatoms. Helgol Mar Res 54, 18–32. https://doi.org/10.1007/s101520050032

Schwelm, A., Fogelqvist, J., Knaust, A., Jülke, S., Lilja, T., Bonilla-Rosso, G., Karlsson, M., Shevchenko, A., Dhandapani, V., Choi, S.R., Kim, H.G., Park, J.Y., Lim, Y.P., Ludwig-Müller, J., Dixelius, C., 2015. The Plasmodiophora brassicae genome reveals insights in its life cycle and ancestry of chitin synthases. Sci Rep 5, 11153. https://doi.org/10.1038/srep11153

Scoble, J.M., Cavalier-Smith, T., 2014. Scale evolution, sequence phylogeny, and taxonomy of thaumatomonad Cercozoa: 11 new species and new genera *Scutellomonas*, *Cowlomonas*, *Thaumatospina* and *Ovaloplaca*. Eur J Protistol 50, 270–313. https://doi.org/10.1016/j.ejop.2013.12.005

Seppey, C.V.W., Singer, D., Dumack, K., Fournier, B., Belbahri, L., Mitchell, E.A.D., Lara, E., 2017. Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling. Soil Biol Biochem 112, 68–76. https://doi.org/10.1016/j.soilbio.2017.05.002

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498–2504. https://doi.org/10.1101/gr.1239303

Sibbald, S.J., Archibald, J.M., 2017. More protist genomes needed. Nat Ecol Evol 1, 0145. https://doi.org/10.1038/s41559-017-0145

Sierra, R., Cañas-Duarte, S.J., Burki, F., Schwelm, A., Fogelqvist, J., Dixelius, C., González-García, L.N., Gile, G.H., Slamovits, C.H., Klopp, C., Restrepo, S., Arzul, I., Pawlowski, J., 2016. Evolutionary

origins of rhizarian parasites. Mol Biol Evol 33, 980–983. https://doi.org/10.1093/molbev/msv340

Solbach, M.D., Bonkowski, M., Dumack, K., 2025. *Katarium polorum* n. sp., n. g., a novel thecofilosean amoeba (Cercozoa, Rhizaria) from the polar oceans. J Eukaryot Microbiol 72, e13071. https://doi.org/10.1111/jeu.13071

Solbach, M.D., Bonkowski, M., Dumack, K., 2021. Novel endosymbionts in rhizarian amoebae imply universal infection of unrelated free-living amoebae by Legionellales. Front Cell Infect Microbiol 11. https://doi.org/10.3389/fcimb.2021.642216

Sommeria-Klein, G., Watteaux, R., Ibarbalz, F.M., Pierella Karlusich, J.J., Iudicone, D., Bowler, C., Morlon, H., 2021. Global drivers of eukaryotic plankton biogeography in the sunlit ocean. Science 374, 594–599. https://doi.org/10.1126/science.abb3717

Soneson, C., Love, M., Robinson, M., 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 4. https://doi.org/10.12688/f1000research.7563.2

Song, L., Florea, L., 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience 4, 48. https://doi.org/10.1186/s13742-015-0089-y

Tachibana, M., Kiyokawa, E., Hara, S., Iemura, S., Natsume, T., Manabe, T., Matsuda, M., 2009. Ankyrin repeat domain 28 (ANKRD28), a novel binding partner of DOCK180, promotes cell migration by regulating focal adhesion formation. Exp Cell Res 315, 863–876. https://doi.org/10.1016/j.yexcr.2008.12.005

Tice, A.K., Žihala, D., Pánek, T., Jones, R.E., Salomaki, E.D., Nenarokov, S., Burki, F., Eliáš, M., Eme, L., Roger, A.J., Rokas, A., Shen, X.-X., Strassert, J.F.H., Kolísko, M., Brown, M.W., 2021. PhyloFisher: A phylogenomic package for resolving eukaryotic relationships. PLOS Biol 19, e3001365. https://doi.org/10.1371/journal.pbio.3001365

Walden, S., Jauss, R.-T., Feng, K., Fiore-Donno, A.M., Dumack, K., Schaffer, S., Wolf, R., Schlegel, M., Bonkowski, M., 2021. On the phenology of protists: recurrent patterns reveal seasonal variation of protistan (Rhizaria: Cercozoa and Endomyxa) communities in tree canopies. FEMS Microbiol Ecol 97, 2021.02.15.431229. https://doi.org/10.1093/femsec/fiab081

Wickham, H., 2011. ggplot2. Wiley Interdiscip Rev Comput Stat 3, 180–185. https://doi.org/10.1002/wics.147

Williams, S.J., Wreschner, D.H., Tran, M., Eyre, H.J., Sutherland, G.R., McGuckin, M.A., 2001. MUC13, a novel human cell surface mucin expressed by epithelial and hemopoietic cells. J Biol Chem 276, 18327–18336. https://doi.org/10.1074/jbc.M008850200

Wiśniewska, M.M., Salomaki, E.D., Silberman, J.D., Terpis, K.X., Mazancová, E., Táborský, P., Jinatham, V., Gentekaki, E., Čepička, I., Kolisko, M., 2024. Expanded gene and taxon sampling of diplomonads shows multiple switches to parasitic and free-living lifestyle. BMC Biol 22, 217. https://doi.org/10.1186/s12915-024-02013-w

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20, 257. https://doi.org/10.1186/s13059-019-1891-0

Xu, F., Jerlström-Hultqvist, J., Kolisko, M., Simpson, A.G.B., Roger, A.J., Svärd, S.G., Andersson, J.O., 2016. On the reversibility of parasitism: adaptation to a free-living lifestyle via gene acquisitions in the diplomonad *Trepomonas* sp. PC1. BMC Biol 14, 62. https://doi.org/10.1186/s12915-016-0284-z

Yao, N.Y., O'Donnell, M.E., 2016. Evolution of replication machines. Crit Rev Biochem Mol Biol 51, 135–149. https://doi.org/10.3109/10409238.2015.1125845

Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11, R14. https://doi.org/10.1186/gb-2010-11-2-r14

9. Data Accessibility

The raw sequencing data are available under NCBI BioProject PRJNA1262159 and the trinity transcriptomes will be available on Zenodo (DOI: https://doi.org/10.5281/zenodo.15091355) upon publication. For peer reviews, please refer to this private link for the trinity transcriptomes: https://zenodo.org/records/15383709?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6IjNmZDg4ZDk4LWU1OGEtNDY5Mi05NmIwLTQ4YTQ4ZTU4Nzc0MiIsImRhdGEiOnt9LCJyYW5kb20iOiIzNzFiOGFjMDRjNDcwOGJlMGRkYjRkOTM1MWVlZjBhMiJ9.QNSpUjiBnubMQVzVPRWoBUiYx744I_AH-0GKuWIy8rwJBACzi5pH0rrWbaas7Amtk1n1kLAQRDI50IP-FTzvgA. Please note that this manuscript has been published as part of Jule Freudenthal PhD thesis in accordance with institutional requirements of the University of Cologne (Freudenthal, 2025) (https://kups.ub.uni-koeln.de/78069/).

10.Authors' contributions

JF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing

MB: Conceptualization, Funding acquisition, Project administration, Supervision,Writing – review & editing

NJME: Investigation, Resources

MS: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing

KD: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing

11.Funding

This study was supported by grants from the German Research Foundation (DFG) in the framework of the priority program SPP 1991: TAXON-OMICS (Project IDs 447013012 and 221301018) and with the grant number 555596351.

12. Declarations of Competing interests

The authors declare no competing interests.

13. Figures

Figure 1: Multi-gene phylogeny of the Rhizaria. Maximum likelihood tree (LG+C60+F+G model) based on an alignment comprising 222 concatenated genes (71,479 amino acids) derived from 40 Rhizaria transcriptomes. Support values were obtained from Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) and 1000 ultrafast bootstraps (UFB). The percentage of gene coverage and the percentage of the site coverage of the present genes for each taxon are shown on the left. Newly added transcriptomes are highlighted in red.

Figure 2: Shared orthogroups and functional annotations of the Thecofilosea. **(A)** The upper bar chart shows the number of shared orthogroups that are unique to the respective combination of Thecofilosida strains displayed in the matrix below. Only the top 25 strain combinations, based on the number of shared orthogroups, are shown. **(B)** BUSCOs assessment of the assembled Thecofilosida transcriptome, based on the Eukaryota database. **(C)** A selection of the functional annotations (KEGG orthology) of the Thecofilosida transcriptomes The grey bar charts in the background display the relative proportions of functional annotations of all transcriptomes for the respective category. In addition, the relative proportions of functional annotations of the core set of orthogroups (purple), the *Rhogostoma* clusters 1 (blue) and 2 (dark green), *Katarium polorum* (light green) and *Fisculla terrestris* (yellow) are shown for the respective category.

Figure 3: Overview of metabolic pathways for the Thecofilosea transcriptomes based on KEGG modules. The heat maps show the presence and absence of the KEGG modules for amino acid metabolism (A), nucleotide metabolism (B), carbohydrate metabolism (C) and lipid metabolism (D) for

each culture-based Thecofilosea transcriptome. The colour indicates the percentage of KO terms present, normalised to the minimum number of KO terms required for each module. White boxes indicate that either more than 50% or more than three of the required KO terms were missing. Modules that were absent in all Thecofilosea transcriptomes are not shown. Abbreviations: BS Biosynthesis, PW Pathway, MB Metabolism, DG Degradation.

Figure 4: Overview of the central carbohydrate and amino acid metabolism for *Rhogostoma kyoshii* (WM). The graph illustrates a reconstruction of the central carbohydrate (green boxes) and amino acid metabolism for *Rhogostoma kyoshii* (WM) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. Amino acids are highlighted in orange, central compounds of carbohydrate metabolism in green. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

Figure 5: Overview of the lipid metabolism for *Rhogostoma kyoshii* (WM). The graph illustrates a reconstruction of the fatty acid (purple boxes), sterol (blue boxes), and lipid (red boxes) metabolism for *Rhogostoma kyoshii* (WM) based on KEGG ontologies and KEGG modules. Nodes represent components and edges represent enzymatic reactions. The edge color indicates the percentage of KO terms present, normalized to the minimum number of KO terms required for the respective reaction. Solid edges indicate that all KO terms were present for the respective reaction.

Figure 6: Comparative transcriptomics of *Rhogostoma*. (A) Principal component analysis of all *Rhogostoma* strains based on the expression patterns of 20,000 orthogroups. The color and shape of the points encode for *Rhogostoma* clusters 1 (blue shades, round) and 2 (green shades, triangular). **(B)** Significantly enriched KO terms in all six *Rhogostoma* strains of cluster 1. The enrichment analysis was based on 7,502 out of 20,000 differentially expressed orthogroups (adjusted p-value < 0.001, |log2

fold change| ≥ 1). The size of the points indicates the frequency of the respective KO terms in the higher expressed orthogroups of *Rhogostoma* cluster 1. The hit percentage describes the ratio of KO terms in higher expressed orthogroups of *Rhogostoma* cluster 1 compared to all orthogroups.

Figure 1

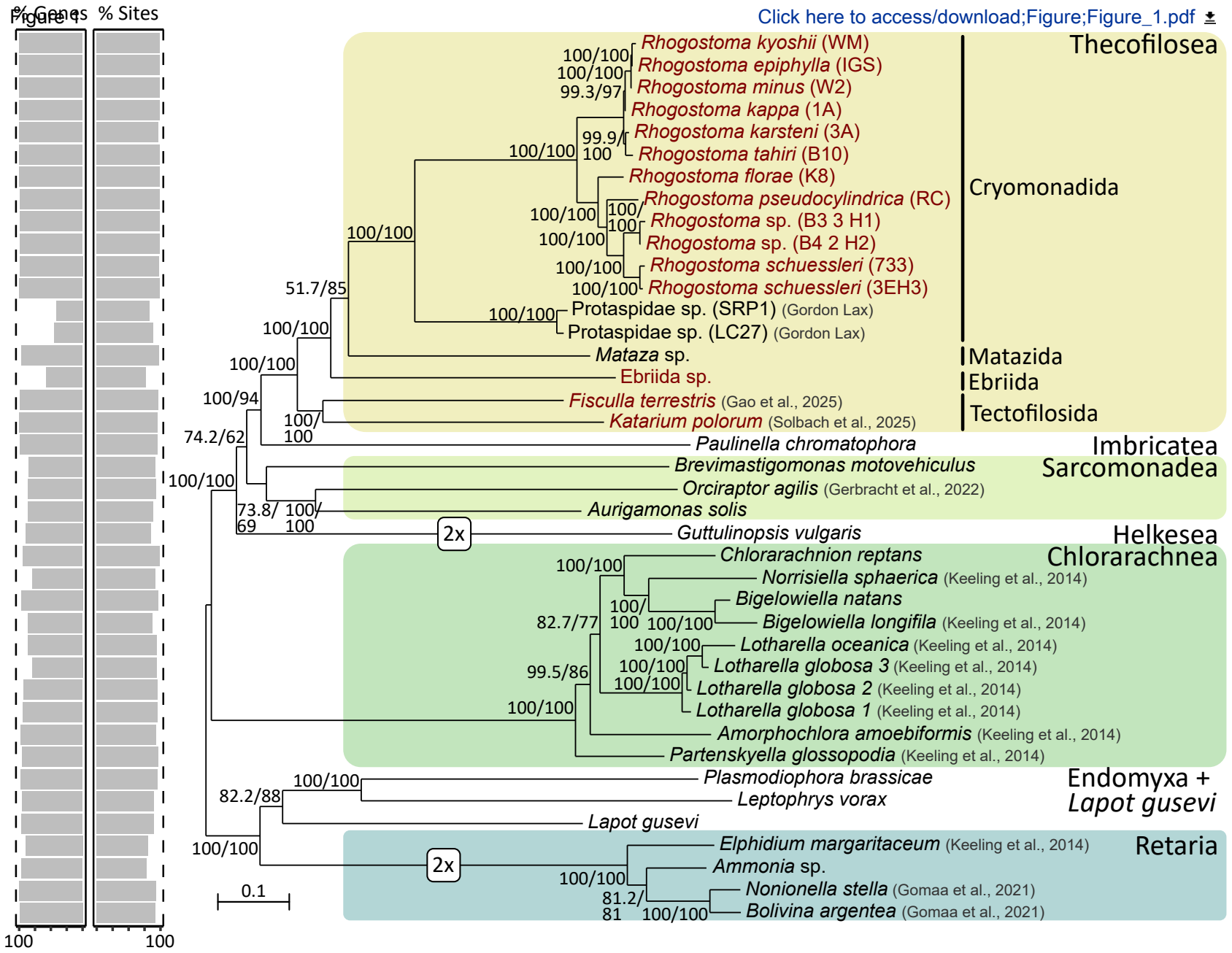

% Genes
% Sites
100
100
Click here to access/download;Figure;Figure_1.pdf
Rhogostoma kyoshii (WM)
Rhogostoma epiphylla (IGS)
Rhogostoma minus (W2)
Rhogostoma kappa (1A)
Rhogostoma karsteni (3A)
Rhogostoma tahiri (B10)
Rhogostoma florae (K8)
Rhogostoma pseudocylindrica (RC)
Rhogostoma sp. (B3 3 H1)
Rhogostoma sp. (B4 2 H2)
Rhogostoma schuessleri (733)
Rhogostoma schuessleri (3EH3)
Protaspidae sp. (SRP1) (Gordon Lax)
Protaspidae sp. (LC27) (Gordon Lax)
Mataza sp.
Ebriida sp.
Fisculla terrestris (Gao et al., 2025)
Katarium polorum (Solbach et al., 2025)
Paulinella chromatophora
Brevimastigomonas motovehiculus
Orciraptor agilis (Gerbracht et al., 2022)
Aurigamonas solis
Guttulinopsis vulgaris
Chlorarachnion reptans
Norrisiella sphaerica (Keeling et al., 2014)
Bigelowiella natans
Bigelowiella longifila (Keeling et al., 2014)
Lotharella oceanica (Keeling et al., 2014)
Lotharella globosa 3 (Keeling et al., 2014)
Lotharella globosa 2 (Keeling et al., 2014)
Lotharella globosa 1 (Keeling et al., 2014)
Amorphochlora amoebiformis (Keeling et al., 2014)
Partenskyella glossopodia (Keeling et al., 2014)
Plasmodiophora brassicae
Leptophrys vorax
Lapot gusevi
Elphidium margaritaceum (Keeling et al., 2014)
Ammonia sp.
Nonionella stella (Gomaa et al., 2021)
Bolivina argentea (Gomaa et al., 2021)
100/100
100/100
99.3/97
99.9/
100
100/100
100/100
100/
100
100/100
100/100
100/100
100/100
100/100
51.7/85
100/100
100/100
100/100
100/94
100/
100
74.2/62
100/100
73.8/
69
100/
100
2x
100/100
82.7/77
100/
100
100/100
100/100
100/100
99.5/86
100/
100
100/100
100/100
82.2/88
100/100
100/100
2x
81.2/
81
100/100
0.1
Thecofilosea
Cryomonadida
Matazida
Ebriida
Tectofilosida
Imbricatea
Sarcomonadea
Helkesea
Chlorarachnea
Endomyxa +
Lapot gusevi
Retaria

Figure 2

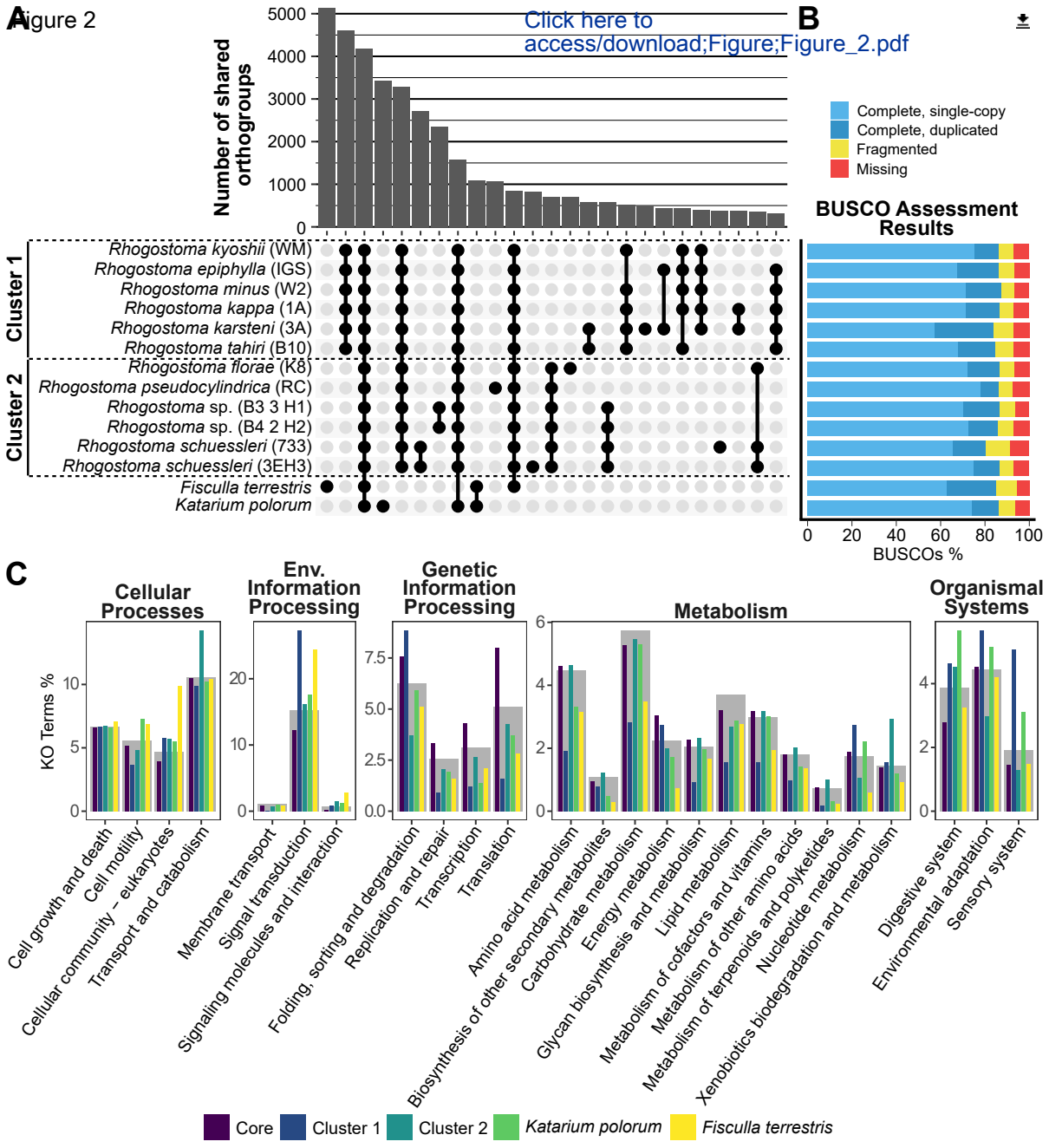


Figure 3

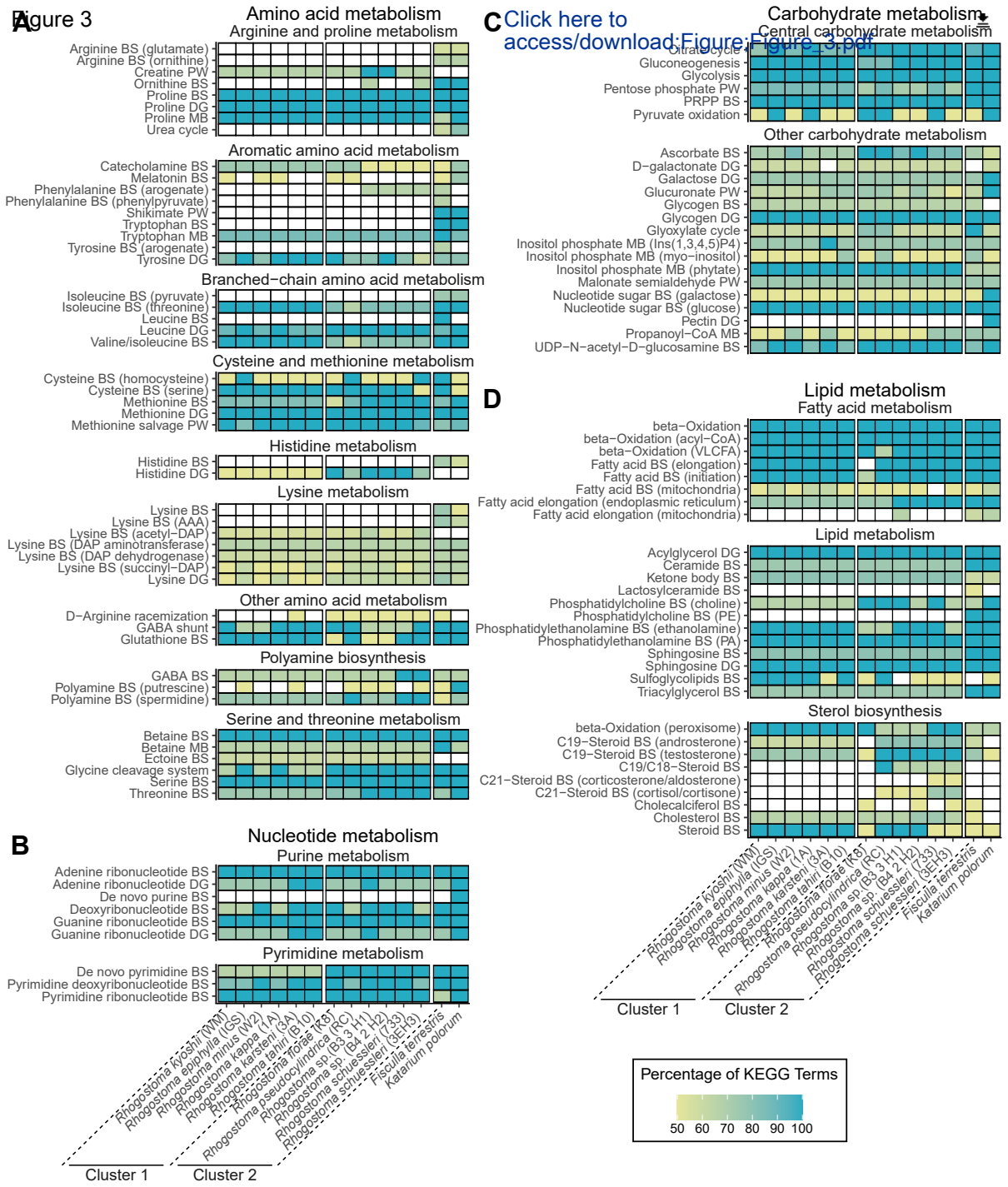

A
Amino acid metabolism
Arginine and proline metabolism
Arginine BS (glutamate)
Arginine BS (ornithine)
Creatine PW
Ornithine BS
Proline BS
Proline DG
Proline MB
Urea cycle
Aromatic amino acid metabolism
Catecholamine BS
Melatonin BS
Phenylalanine BS (arogenate)
Phenylalanine BS (phenylpyruvate)
Shikimate PW
Tryptophan BS
Tryptophan MB
Tyrosine BS (arogenate)
Tyrosine DG
Branched−chain amino acid metabolism
Isoleucine BS (pyruvate)
Isoleucine BS (threonine)
Leucine BS
Leucine DG
Valine/isoleucine BS
Cysteine and methionine metabolism
Cysteine BS (homocysteine)
Cysteine BS (serine)
Methionine BS
Methionine DG
Methionine salvage PW
Histidine metabolism
Histidine BS
Histidine DG
Lysine metabolism
Lysine BS
Lysine BS (AAA)
Lysine BS (acetyl−DAP)
Lysine BS (DAP aminotransferase)
Lysine BS (DAP dehydrogenase)
Lysine BS (succinyl−DAP)
Lysine DG
Other amino acid metabolism
D−Arginine racemization
GABA shunt
Glutathione BS
Polyamine biosynthesis
GABA BS
Polyamine BS (putrescine)
Polyamine BS (spermidine)
Serine and threonine metabolism
Betaine BS
Betaine MB
Ectoine BS
Glycine cleavage system
Serine BS
Threonine BS
B
Nucleotide metabolism
Purine metabolism
Adenine ribonucleotide BS
Adenine ribonucleotide DG
De novo purine BS
Deoxyribonucleotide BS
Guanine ribonucleotide BS
Guanine ribonucleotide DG
Pyrimidine metabolism
De novo pyrimidine BS
Pyrimidine deoxyribonucleotide BS
Pyrimidine ribonucleotide BS
Rhogostoma kyoshii (WMT
Rhogostoma epiphylla (ICS)
Rhogostoma minus (W2)
Rhogostoma kappa (1A)
Rhogostoma karsteni (3A)
Rhogostoma tahiri (B10)
Rhogostoma floreae (KBT
Rhogostoma pseudocylindrica (RC)
Rhogostoma sp. (B3 3 H1)
Rhogostoma sp. (B4 2 H2)
Rhogostoma schuessleri (733)
Rhogostoma schuessleri (3EH9)
Fisculla terrestris
Katarium polorum
Cluster 1
Cluster 2
C
Click here to access/download;Figure;Figure_3.pdf
Carbohydrate metabolism
Central carbohydrate metabolism
Citrate cycle
Gluconeogenesis
Glycolysis
Pentose phosphate PW
PRPP BS
Pyruvate oxidation
Other carbohydrate metabolism
Ascorbate BS
D−galactonate DG
Galactose DG
Glucuronate PW
Glycogen BS
Glycogen DG
Glyoxylate cycle
Inositol phosphate MB (Ins(1,3,4,5)P4)
Inositol phosphate MB (myo−inositol)
Inositol phosphate MB (phytate)
Malonate semialdehyde PW
Nucleotide sugar BS (galactose)
Nucleotide sugar BS (glucose)
Pectin DG
Propanoyl−CoA MB
UDP−N−acetyl−D−glucosamine BS
D
Lipid metabolism
Fatty acid metabolism
beta−Oxidation
beta−Oxidation (acyl−CoA)
beta−Oxidation (VLCFA)
Fatty acid BS (elongation)
Fatty acid BS (initiation)
Fatty acid BS (mitochondria)
Fatty acid elongation (endoplasmic reticulum)
Fatty acid elongation (mitochondria)
Lipid metabolism
Acylglycerol DG
Ceramide BS
Ketone body BS
Lactosylceramide BS
Phosphatidylcholine BS (choline)
Phosphatidylcholine BS (PE)
Phosphatidylethanolamine BS (ethanolamine)
Phosphatidylethanolamine BS (PA)
Sphingosine BS
Sphingosine DG
Sulfoglycolipids BS
Triacylglycerol BS
Sterol biosynthesis
beta-Oxidation (peroxisome)
C19−Steroid BS (androsterone)
C19−Steroid BS (testosterone)
C19/C18−Steroid BS
C21−Steroid BS (corticosterone/aldosterone)
C21−Steroid BS (cortisol/cortisone)
Cholecalciferol BS
Cholesterol BS
Steroid BS
Rhogostoma kyoshii (WMT
Rhogostoma epiphylla (ICS)
Rhogostoma minus (W2)
Rhogostoma kappa (1A)
Rhogostoma karsteni (3A)
Rhogostoma tahiri (B10)
Rhogostoma floreae (KBT
Rhogostoma pseudocylindrica (RC)
Rhogostoma sp. (B3 3 H1)
Rhogostoma sp. (B4 2 H2)
Rhogostoma schuessleri (733)
Rhogostoma schuessleri (3EH9)
Fisculla terrestris
Katarium polorum
Cluster 1
Cluster 2
Percentage of KEGG Terms
50 60 70 80 90 100

Figure 4

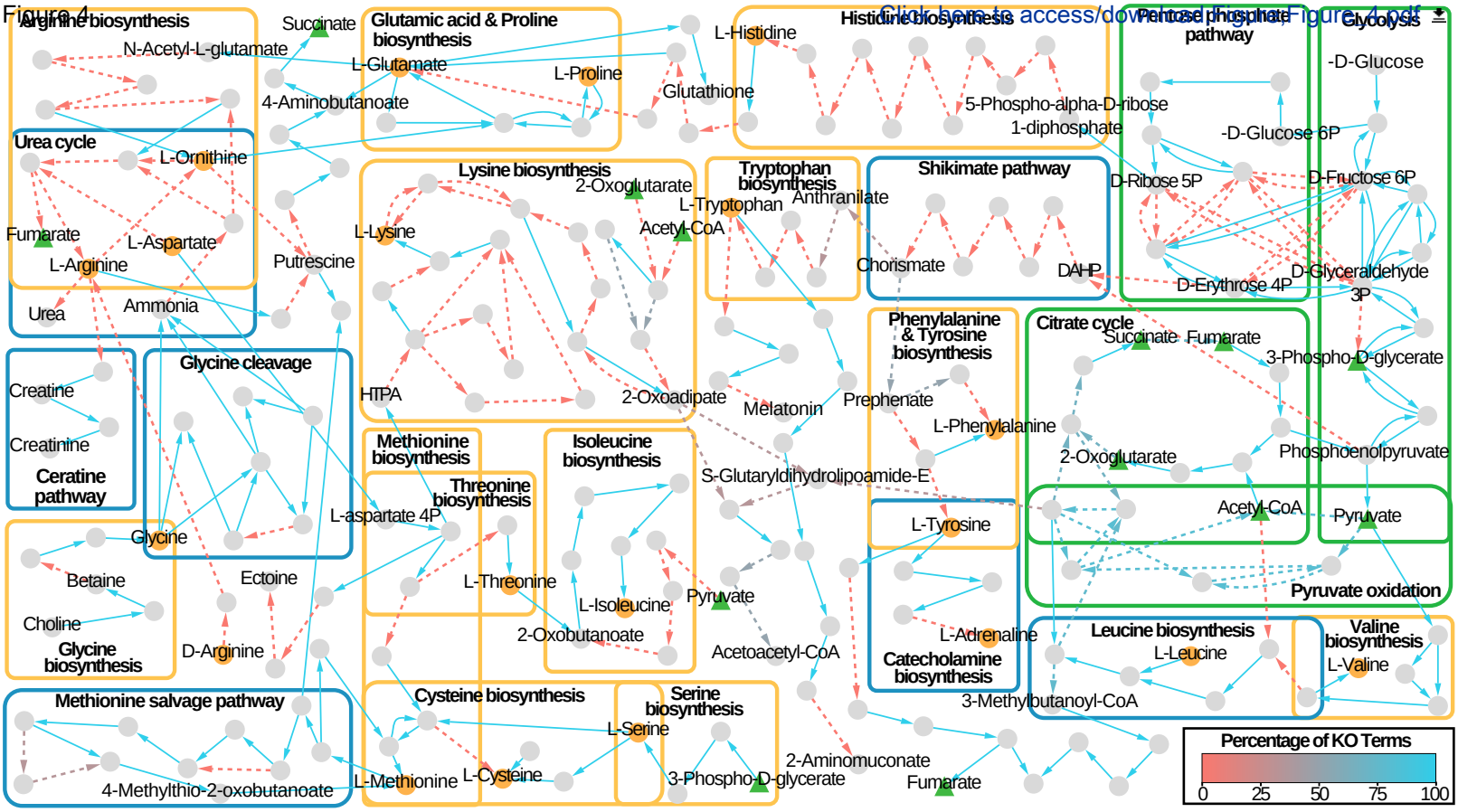

Click here to access/download Figure 4.pdf
Arginine biosynthesis
N-Acetyl-L-glutamate
Succinate
Urea cycle
L-Ornithine
Fumarate
L-Aspartate
L-Arginine
Urea
Ammonia
Putrescine
Glutamic acid & Proline biosynthesis
L-Glutamate
L-Proline
4-Aminobutanoate
Glutathione
L-Histidine
Histidine biosynthesis
5-Phospho-alpha-D-ribose
1-diphosphate
Pentose phosphate pathway
Glycolysis
-D-Glucose
-D-Glucose 6P
D-Ribose 5P
D-Fructose 6P
D-Erythrose 4P
D-Glyceraldehyde 3P
Lysine biosynthesis
2-Oxoglutarate
L-Lysine
Acetyl-CoA
HTPA
2-Oxoadipate
Tryptophan biosynthesis
L-Tryptophan
Anthranilate
Shikimate pathway
Chorismate
DAHP
Phenylalanine & Tyrosine biosynthesis
Prephenate
L-Phenylalanine
L-Tyrosine
Catecholamine biosynthesis
L-Adrenaline
Citrate cycle
Succinate
Fumarate
3-Phospho-D-glycerate
2-Oxoglutarate
Phosphoenolpyruvate
Acetyl-CoA
Pyruvate
Pyruvate oxidation
Creatine
Creatinine
Ceratine pathway
Glycine cleavage
Glycine
Betaine
Choline
Glycine biosynthesis
Ectoine
D-Arginine
Methionine biosynthesis
Threonine biosynthesis
L-aspartate 4P
L-Threonine
Isoleucine biosynthesis
L-Isoleucine
2-Oxobutanoate
Pyruvate
Acetoacetyl-CoA
S-Glutaryldihydrolipoamide-E
Melatonin
Leucine biosynthesis
L-Leucine
3-Methylbutanoyl-CoA
Valine biosynthesis
L-Valine
Methionine salvage pathway
4-Methylthio-2-oxobutanoate
L-Methionine
Cysteine biosynthesis
L-Cysteine
Serine biosynthesis
L-Serine
3-Phospho-D-glycerate
2-Aminomuconate
Fumarate
Percentage of KO Terms
0
25
50
75
100

Figure 5