Taxon-rich transcriptomics supports higher-level phylogeny and major evolutionary trends in Foraminifera

Sierra Miranda, Roberto Mario; Mauffrey, Florian; Cruz, Joana; Holzmann, Maria; Gooday, Andrew J.; Maurer-Alcalá, Xyrus; Thakur, Rabindra; Greco, Mattia; Weiner, Agnes K.M.; Katz, Laura A.; Pawlowski, Jan Wojciech

How to cite

SIERRA MIRANDA, Roberto Mario et al. Taxon-rich transcriptomics supports higher-level phylogeny and major evolutionary trends in Foraminifera. In: Molecular phylogenetics and evolution, 2022, vol. 174, p. 107546. doi: 10.1016/j.ympev.2022.107546

This publication URL: https://archive-ouverte.unige.ch/unige:175221

Publication DOI: 10.1016/j.ympev.2022.107546

Molecular Phylogenetics and Evolution 174 (2022) 107546

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Taxon-rich transcriptomics supports higher-level phylogeny and major evolutionary trends in Foraminifera

Roberto Sierra [a,*], Florian Mauffrey [a], Joana Cruz [a], Maria Holzmann [a], Andrew J. Gooday [b,c], Xyrus Maurer-Alcalá [d], Rabindra Thakur [e,f], Mattia Greco [e,g], Agnes K.M. Weiner [e,h], Laura A. Katz [e,f], Jan Pawlowski [a,g,i]

[a] Department of Genetics and Evolution, University of Geneva, 1211 Geneva, Switzerland
[b] National Oceanography Centre, European Way, Southampton SO14 3ZH, UK
[c] Life Sciences Department, Natural History Museum, Cromwell Road, London SW7 5BD, UK
[d] Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA
[e] Department of Biological Sciences, Smith College, Northampton, MA 01063, USA
[f] University of Massachusetts Amherst, Program in Organismic and Evolutionary Biology, Amherst, MA, USA
[g] Institute of Oceanology, Polish Academy of Sciences, 81-712 Sopot, Poland
[h] NORCE Climate and Environment, NORCE Norwegian Research Centre AS, Jahnebakken 5, 5007 Bergen, Norway
[i] ID-Gene Ecodiagnostics, Chemin du Pont-du-Centenaire 109, CH-1228 Plan-les-Ouates, Switzerland

ARTICLE INFO

Keywords:
Foraminifera
Monothalamids
Globothalamea
Tubothalamea
Xenophyophores
Phylogenomics

ABSTRACT

Foraminifera, classified in the supergroup Rhizaria, are a common and highly diverse group of mainly marine protists. Despite their evolutionary and ecological importance, only limited genomic data (one partial genome and nine transcriptomic datasets) have been published for this group. Foraminiferal molecular phylogeny is largely based on 18S rRNA gene sequence analysis. However, due to highly variable evolutionary rates of substitution in ribosomal genes plus the existence of intragenomic variation at this locus, the relationships between and within foraminiferal classes remain uncertain. We analyze transcriptomic data from 28 species, adding 19 new species to the previously published dataset, including members of the strongly under-represented class Monothalamea. A phylogenomic reconstruction of Rhizaria, rooted with alveolates and stramenopiles, based on 199 genes and 68 species supports the monophyly of Foraminifera and their sister relationship to Polycystinea. The phylogenomic tree of Foraminifera is very similar to the 18S rRNA tree, with the paraphyletic single-chambered monothalamids giving rise to the multi-chambered Tubothalamea and Globothalamea. Within the Monothalamea, our analyses confirm the monophyly of the giant, deep-sea xenophyophores that branch within clade C and indicate the basal position of monothalamous clades D and E. The multi-chambered Globothalamea are monophyletic and comprise the paraphyletic Textulariida and monophyletic Rotaliida. Our phylogenomic analyses support major evolutionary trends of Foraminifera revealed by ribosomal phylogenies and reinforce their current higher-level classification.

1. Introduction

Foraminifera are classified in the supergroup Rhizaria that is sister to Alveolata and Stramenopila, and altogether are referred to as the SAR group. Foraminifera is a common group of protists widespread in all marine environments from estuaries to deep-sea trenches, with few lineages also living in freshwater settings (Pawlowski et al., 1999; Holzmann et al., 2003; Murray 2006; Siemensma et al., 2017; Holzmann et al., 2021). The group evolved in the Precambrian, possibly from radiolarian-like ancestors with which they form the well-supported group Retaria (Pawlowski and Burki 2009; Burki et al., 2010). Foraminiferal origins date to about 650–900 mya, according to a Bayesian relaxed clock (Groussin et al., 2011). No Precambrian fossils can be assigned unambiguously to the Foraminifera, but the group was presumably represented at that time by naked or single-chambered monothalamous species that may have had proteinaceous tests (i.e., shells), consistent with the appearance of a large paraphyletic assemblage of such lineages as early-diverging branches in ribosomal DNA trees

* Corresponding author.
E-mail address: Roberto.Sierra@me.com (R. Sierra).

https://doi.org/10.1016/j.ympev.2022.107546
Received 10 November 2021; Received in revised form 10 May 2022; Accepted 18 May 2022
Available online 8 June 2022

(Pawlowski 2012). Foraminifera possibly played a crucial role in late Precambrian ecosystems, as suggested by study of recent microbialites (Bernhard et al., 2013) and fossil lipid biomarkers (Nettersheim et al., 2019).

Until recently, the molecular phylogeny of Foraminifera has been based nearly exclusively on ribosomal RNA (rRNA) genes (Pawlowski et al., 2003; Pawlowski 2012; Holzmann and Pawlowski 2017). The current database includes over 4000 sequences from about 600 species representing all extant orders except for the order Lagenida. Phylogenetic studies based on 18S rRNA genes divide Foraminifera into three classes: Monothalamea, Tubothalamea and Globothalamea (Pawlowski et al., 2013). Each of these classes is characterized by specific morphogenetic features. Monothalamea are a paraphyletic group that comprises a tremendous variety of single-chambered forms with organic or agglutinated walls, as well as a few naked (non-testate) forms. The Tubothalamea and Globothalamea are characterized by multi-chambered agglutinated or calcareous tests that differ in terms of chamber number, shape, and arrangement. While the relationships within the Globothalamea and Tubothalamea have been relatively well-established based on 18S rRNA gene sequences, the phylogeny of the Monothalamea remains poorly known. The classification proposed by Pawlowski et al. (2002) subdivides this group into 13 clades, named from A to M. However, this classification has been established based on very limited number of sequences representing 50 species only and it is currently under revision (Pawlowski et al., in prep). The presence of multiple indels and the large variation of evolutionary rates in the ribosomal genes of Foraminifera and their close radiolarian relatives impede any reliable inferences about the branching order of monothalamid clades at the base of the foraminiferal tree.

Among the protein-coding genes, single-gene trees for Foraminifera have been inferred from actin (Flakowski et al., 2005, 2006; Krabberod et al., 2017), RNA Polymerase II (Longet and Pawlowski 2007), and alpha- and beta-tubulin (Hou et al., 2013; Krabberod et al., 2017). Both actin and tubulin gene families, however, are not the best phylogenetic markers due to the large number of paralogs. The only foraminiferal genomic and transcriptomic data published until now include the freshwater species *Reticulomyxa filosa* (Glockner et al., 2014), seven rotaliid and one miliolid species (Sierra et al., 2013). These analyses confirmed the monophyly of Rotaliida but contributed little to resolving the relationships of other groups, especially within Monothalamea.

The limited amount of genomic data available for Foraminifera reflects various factors: the difficulty of culturing most species and the lack of access to well-established cultures of the few cultivable species. With the notable exception of the freshwater *R. filosa* (Nauss 1949; Glockner et al., 2014), all other foraminiferal genomic and transcriptomic data have been obtained using multiple specimens that have been sorted from environmental samples, cleaned, and characterized. Only recently have new single-cell RNA extraction methods helped to override this limitation, providing high-quality material for exploring the transcriptomics of uncultured protists (Yan et al., 2019).

Here, we report 19 additional transcriptomes of Foraminifera, including a wide range of species from all major taxonomic groups. We analyzed these transcriptomes in order to revise the molecular phylogeny of Foraminifera, focusing on relationships among Monothalamea and between them and the two multi-chambered classes: Globothalamea and Tubothalamea.

2. Material and methods

2.1. Sampling

Foraminifera were collected in different ecosystems and in different geographic areas. The three xenophyophores (*Aschemonella monilis, Psammina* aff. *limbata* sensu Gooday et al., 2018, *Shinkaiya lindsayi*) were collected from the seafloor at abyssal depths (>4000 m) in the Clarion-Clipperton Zone (eastern equatorial Pacific) (Gooday et al., 2017; Gooday et al., 2018) and the Japan Trench (NW Pacific) (Lecroq et al., 2009). The other 16 species were obtained in much shallower water (<300 m depth), with nine originating from Patagonian fjords, three from Greenland fjords, three from the Faroe Islands and Svalbard, and one from the Gulf of Eilat (Table S1). Except for the large-sized xenophyophores, which were picked directly from freshly-collected sediment core surfaces, all specimens were sorted from sieved sediment samples, washed in prefiltered seawater and stored in RNA-later until processing, either in the refrigerator or frozen at −80 °C. Fragments of freshly-collected xenophyophores were broken off and preserved in RNA-later (Qiagen, Germany), and parts of the branching cellular system was later dissected from these fragments.

2.2. RNA extraction and cDNA synthesis

Total RNA was extracted by using the commercial kit RNeasy® Micro Kit (RNeasy®, Qiagen®) following the manufacturer's instructions except for the lysis step for which we heated the samples at 65 °C for 3 min. RNA quantification was performed with a Qubit fluorometer (Life Technologies®) and RNA integrity was assessed with a Bioanalyzer (Agilent Technologies®). From total RNA, cDNA was synthesized according to the iScript cDNA Synthesis Kit standard protocol (Bio-Rad).

2.3. 18S rRNA gene sequencing

To confirm the morphological identification of the analyzed species, all of them were sequenced for the foraminiferal barcoding fragment of the 18S rRNA gene. The PCR amplifications were performed on cDNA either using primer pairs s14F3 (ACGCAMGTGTGAAACTTG) − 20r (GACGGGCGGTGTGTACAA) for monothalamids, or primer pairs s14F3 - sB (TGATCCTTCTGCAGGTTCACCTAC) for globothalamids. The amplified PCR products were purified using High Pure PCR Purification Kit (Roche Diagnostics) and sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems®) on a 3130XL Genetic Analyzer (Applied Biosystems®). Direct sequencing was not possible for some species (*Globobulimina* sp., *Eggerelloides scaber*, *Reophax* sp., *Spiroplectammina* sp., *Hippocrepina (?) alba* and saccamminids). For these species, PCR products were cloned with the TOPO TA Cloning Kit (Invitrogen®) following the manufacturer's instructions and transformed into competent *E. coli* cells.

2.4. 18S phylogeny

The phylogeny of Tubothalamea and monothalamids (62 taxa and 2236 sites, of which 1531 sites were used for phylogenetic analysis) and Globothalamea (69 taxa and 1770 sites of which 1177 sites were used for phylogenetic analysis) were derived from 18S sequences (Pawlowski and Holzmann 2014). The obtained sequences were added to an existing database using MUSCLE automatic alignment option, as implemented in SeaView vs. 4.3.3 (Gouy et al., 2010). The maximum likelihood phylogenetic trees were constructed using PhyML 3.0 (Guindon et al., 2010). An automatic model selection by SMS (Lefort et al., 2017) based on Akaike Information Criterion (AIC) was used, resulting in a GTR + G + I substitution model being selected for the analysis. The initial trees were based on BioNJ. Statistical branch support was given by bootstrap values based on 100 replicates.

2.5. RNA-seq

For the bulk of our analyses, the TruSeq mRNA stranded kit from Illumina was used for the library preparation with 100 ng of total RNA as input. Library quantifications, molarity and quality were assessed with a Qubit fluorometer (Life Technologies®) and a Tapestation using a DNA High sensitivity chip (Agilent Technologies®). Eight to ten libraries were pooled at 2 nM each and loaded for clustering on a lane of a single-read Illumina Flow Cell. Reads of 100 bases were generated using the

TruSeq SBS chemistry on an Illumina HiSeq 4000 sequencer yielding approximately 400 million reads per lane.

2.6. Reads filtering and assembly

For each transcriptome, fastq files were analyzed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads with bad quality were filtered out using the FASTX-Toolkit, keeping reads with at least 90 % of bases having a minimum quality score of 20 (https://hannonlab.cshl.edu/fastx_toolkit). Filtered reads were assembled into transcripts using rnaSpades with default parameters (Bushmanova et al., 2019). Assembly metrics were assessed using QUAST (Gurevich et al., 2013). Putative protein sequences were derived from contig sequences using the Transeq tool from EMBOSS with the "-clean", "-trim" and "-frame 6" arguments (Rice et al., 2000). Finally, proteins sequences were used as databases in the search for genes of interest.

The newly sequenced transcriptome of *Ammonia* sp. was obtained from the same sample as previously sequenced and thus merged with the existing sequences for the genus, and either the complete or the longest gene sequence were kept.

2.7. Phylogenetic analyses

An existing database of 229 protein coding genes used as eukaryotic phylogenic markers was utilized for this study (Sierra et al., 2016). Based on these sequences, we used blastp from the blast + tools suite (Camacho et al., 2009) to extract homologous sequences from the newly sequenced Foraminifera transcriptomes. Only hits with a minimum e-value of 1E-05 were considered. For all species, the three homologous contigs with the lowest e-values were chosen for each of the 229 genes tested. Reciprocal best blastp hit to the existing ortholog set was used to exclude paralogous sequences. Corresponding sequences were extracted and aligned with their corresponding reference sequences using MAFFT (Katoh and Standley 2013) option L-INS-i. Based on these alignments, sequences were manually curated, and those that were poorly aligned were removed. Twelve genes were removed from the analysis for lack of good quality sequences, leaving 217 genes from the 229 protein-coding genes tested. These alignments were trimmed with BMGE using a BLOSUM75 matrix, a minimum block size of 5 and a gap rate cut-off of 0.3 (Criscuolo and Gribaldo 2010). Independent trees were generated from the 217 alignments using RAxML v. 8.2 using PROTGAMMALG model and 100 bootstrap replicates (Stamatakis 2014). Three trees where the foraminiferal origin of the genes was not clear were considered of poor quality and corresponding alignments were discarded. Fifteen trees highlighting the putative presence of paralogous sequences were also removed. Finally, a total of 199 alignments were concatenated using SCaFoS (Roure et al., 2007) and new trees were generated from this super matrix using Bayesian inference and maximum likelihood (ML) analyses. The ML analysis were performed with RAxML v.8.2 using PROTGAMMALG and 1000 bootstrap replicates, and IQ-Tree v.1.6 (Nguyen et al., 2015) using LG + G and 1000 ultrafast (Hoang et al., 2018) and standard bootstrap replicates. Bayesian Inference was conducted using ExaBayes v. 1.5 (Aberer et al., 2014) with GTR + G model, running four independent chains in parallel until topological convergence was achieved (average standard deviation of split frequencies (ASDSF) $<$ 5%, at least 10,000 generations) sampling every 500th generation starting from a parsimony tree. After convergence, the extended majority consensus trees with a burnin proportion of 0.5 were calculated using consense (implemented in the ExaBayes package).

3. Results

3.1. RNA sequencing, assembly, and matrix construction

Nineteen new foraminiferal transcriptome libraries were sequenced, obtaining between ~30 and 50 million raw reads per sample (Table S1). The reads were assembled with rnaSpades and the resulting contigs were queried for eukaryotic phylogenetic markers. In the matrix for phylogenomic analyses we included 47 Rhizaria, 11 stramenopiles and 10 alveolates, with accession numbers for data from GenBank in Table S1. The matrix comprised 199 gene families and represented 49,227 amino acids. Compared to previous studies, the rate of missing data for the newly sequenced foraminiferal species was very low (on average 15.95%), especially for textulariids, xenophyophores and monothalamids (Fig. 1, % of positions). The overall rate of missing data in the matrix is 39.35% (based on the number of sites).

3.2. 18S phylogeny

The 18S rRNA trees were reconstructed separately for monothalamous/tubothalamous and globothalamous Foraminifera. The trees included representatives of main families and clades, as well as the sequences of species for which transcriptomic data were obtained (Figs. S1–S2). The tree for the Monothalamea and Tubothalamea (Fig. S1) confirms the position of the three morphologically identified xenophyophore species branching within clade C, together with *Hippocrepina indivisa* and *Toxisarcon taimyr*. It also confirms the position of *Psammophaga fuegia* branching within clade E, and *Hippocrepina hirudinea* within clade D. Regarding the two morphologically unidentified monothalamids, the 18S tree indicated that the undetermined allogromiid (undet monothalamid) branches in clade B as sister group to *Psammosphaera*, while the undetermined saccamminid branches within clade F, as sister to *Hemisphaerammina*. The 18S rRNA analysis also confirms the position of previously sequenced miliolids (*Sorites orbiculus* and *Quinqueloculina seminula*) as well as that of *R. filosa*.

The 18S rRNA tree of Globothalamea (Fig. S2) confirms the morphological classification of all species except *Amphistegina* sp. This calcareous rotaliid is characterized by very rapidly evolving 18S genes and branches outside the globothalamid group in all analyses (data not shown). Other globothalamid species branch as expected. The two rotaliids, *Globobulimina turgida* and *Elphidium macellum,* branch within the Globobulimindae and Elphidiidae, respectively. The position of six textulariids analyzed here is also congruent with the 18S-based classification, *Reophax* sp. branches with other Reophacidae, while *Spiroplectammina* sp., *Trochammina* sp., *Eggerelloides scaber, Textularia gramen* and *Cribrostomoides crassimargo* branch with corresponding genera.

3.3. Foraminiferal phylogenomics

All trees showed the same relationships between the different species and the three groups comprising the SAR group (Stramenopiles + Alveolates + Rhizaria). The four independent chains from the Bayesian inference converged (ASDSF $<$ 5%) after 295,000 generations and the obtained consensus trees were topological identical. The final tree was rooted using the stramenopiles and alveolates (Fig. S3). The global topology of the tree is well supported with posterior probabilities (PP) of 1 and bootstrap values (BS) of 100 for most nodes. Nodes that are not fully supported are noted otherwise (Figs. 1 and S3).

The tree in Fig. 1 depicts the retarian clade taken from the complete SAR tree in Fig. S3. Globally, all major nodes grouping Foraminifera, Polycystinea and Acantharea are strongly supported with full PP and BS values. The position of *Sticholonche* as sister group to the Acantharea is not supported. There is also no statistical support for the sister relationship between Polycystinea and Foraminifera. Within the Foraminifera, all the major clades present in the 18S phylogeny were recovered (Fig. 1). The Rotaliida (including *Amphistegina*), Xenophyophorea and Miliolida formed well defined monophyletic clades with high support according to PP and BS. In particular, the Rotaliida includes the genus *Amphistegina* that branches outside in the 18S phylogenies, despite its morphological affinity to other rotaliids. Textulariida species formed a series of fully supported paraphyletic branches at the base of the

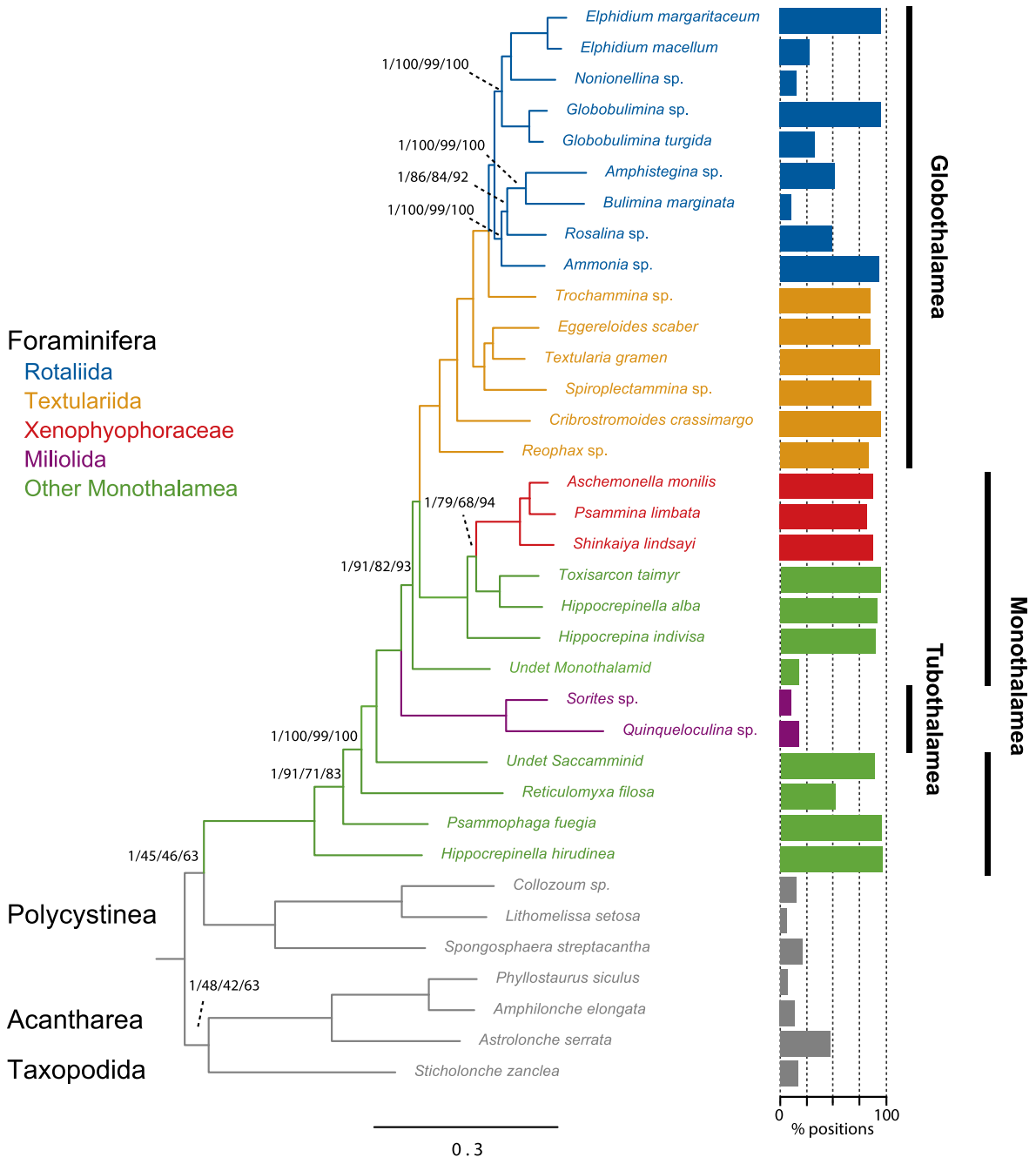

Fig. 1. Retarian phylogeny showing all major groups of Foraminifera and radiolarians. Adapted from the recovered SAR (Stramenopiles + Alveolates + Rhizaria) phylogeny in Fig. S3. The tree corresponds to the best ML tree and Bayesian inference from a concatenated analysis of 199 proteins. Phylogenetic support at nodes was obtained from ML analysis and Bayesian posterior probabilities, red circles denote 100 bootstrap value and posterior probability of 1, otherwise support is shown at the node, dashes (-) represent an alternative branching in the analysis. The support at nodes correspond to posterior probabilities, RAxML bootstrap, IQtree standard bootstrap and IQtree UFBoot values, respectively. The bars next to each species name denote the percentage of sites present in the matrix per data set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Rotaliida. Finally, the Monothalamea species formed a paraphyletic group at the base of Foraminifera. Among the species or genera for which new transcriptomic data were obtained, *Globobulimina* sp. and *E. macellum* sequences clustered with previously sequenced *G. turgida* and *E. margaritaceum*, respectively.

4. Discussion

Our study reports the first comprehensive phylogenomic analysis of Foraminifera, including all major taxonomic groups of this phylum, for which transcriptomic data is available. Until now, Foraminifera were represented in eukaryotic phylogenomic trees by only a few species, most of them belonging to the order Rotaliida (Sierra et al., 2013; Sierra et al., 2016; Burki et al., 2020). This situation was also typical of other Rhizaria, which are generally under-represented in phylogenomic studies (Burki et al., 2010; Burki et al., 2013; Burki and Keeling 2014). Our study has changed this situation, by adding 19 new foraminiferal transcriptomes to the nine already existing, meaning that this group is now represented in phylogenomic trees by more species than all other rhizarian taxa combined. Moreover, the new foraminiferal transcriptomes contain a very low proportion of missing data compared to the previous studies, especially regarding the Polycystinea and

Acantharea (Burki et al., 2010). These high-quality transcriptomes confirm the efficiency of new RNA extraction methods applied to uncultured rhizarian taxa. A recent study by Weiner et al., (2020) also reported on the successful sequencing of a wide range of protist single-cell transcriptomes, including rhizarian lineages, isolated from environmental samples. Together, these data highlight the potential of this method for future phylogenomic and functional diversity studies.

Overall, the foraminiferal phylogenomic tree reported here (Fig. 1) is in congruence with previous single-gene phylogenies. Its general structure is very similar to the trees published based on 18S rRNA (Pawlowski et al., 2013), actin (Flakowski et al., 2005, 2006), RNA polymerase (Longet and Pawlowski 2007), and beta-tubulin (Hou et al., 2013), as well as small-scale multigene analyses (Groussin et al., 2011; Krabberod et al., 2017). Our phylogenomic analyses confirmed with high support the distinction of the two major classes of Foraminifera: the multi-chambered Globothalamea and Tubothalamea, and the paraphyly of single-chambered monothalamids (Pawlowski et al., 2013; Adl et al., 2019). The fourth large group, the calcareous Lagenida, is still not represented by any reliable genomic data, which makes it impossible to establish its phylogenetic position. As our study did not add any new data for Tubothalamea, we will focus our discussion on Globothalamea and Monothalamea only.

Focusing on the ecologically important Globothalamea, our main contribution consists in adding six new transcriptomes of species belonging to the order Textulariida, a group not represented previously in any phylogenomic tree. The Textulariida are common multi-chambered Foraminifera characterized by an agglutinated test wall. Until now, all available molecular data for this order have been 18S rRNA and actin genes (Pawlowski et al., 2003; Groussin et al., 2011). Our analyses confirm the paraphyletic nature of the Textulariida and its position at the base of the calcareous Rotaliida (Groussin et al., 2011; Holzmann and Pawlowski 2017). They also indicate the sister position of *Reophax* relative to other Globothalamea, which is congruent with the relatively simple linear test morphology of this genus and micropaleontological studies that reveal it to be one of the early evolving multi-chambered foraminiferal lineages (Pawlowski et al., 2013). The relationships between other textulariid genera inferred from our data are also in broad agreement with morphotaxonomic classifications, although it is impossible to make any general conclusions about the phylogeny of this order based on our inclusion of only six species.

Regarding the phylogeny of the Rotaliida, another important order of Globothalamea now represented in phylogenomic trees by nine transcriptomes, the relations inferred from transcriptomic data substantially differ from the 18S phylogenies that comprise many more taxa (Holzmann and Pawlowski 2017). For example, the genera *Ammonia* and *Elphidium* are classified together within the superfamily Rotalioidea according to 18S phylogenies but branch far from each other in the multigene tree. The strong support for the cluster of *Elphidium*, *Nonionellina* and *Globobulimina* is also in disagreement with the 18S-based classification of the Rotaliida, which place each of these genera in different superfamilies (Holzmann and Pawlowski 2017). This is also true in the case of *Amphistegina* and *Bulimina*, which branch together despite very different morphological features. Interestingly, the genus *Amphistegina* had presented particular problems, as its phylogenetic position could not be established based on the 18S gene; its unusually fast-evolving 18S sequences place *Amphistegina* somewhere among the monothalamid clades, despite its hyaline calcareous test, which is characteristic of the order Rotaliida (Pawlowski et al., 2013). Transcriptomic data solves this problem by placing *Amphistegina* within Rotaliida, although its position within this order remains to be confirmed.

The most important contribution of our study concerns the phylogeny of the class Monothalamea. In previous phylogenomic studies, this class was represented by a unique, naked, freshwater species, *R. filosa* (Glockner et al., 2014). Here, we added transcriptomic data for 10 additional species representing five different clades. The analyses of this expanded dataset reveal some very interesting relationships. First, we confirm the monophyly of the Xenophyophorea, megafauna-sized agglutinated Foraminifera that are common in many parts of the deep sea (Gooday et al., 2017). The three species of this very unusual taxon form a strongly supported monophyletic group, which clusters together with the three other monothalamous species. All of them are classified in clade C according to 18S phylogeny (Pawlowski et al., 2002). The monothalamous clade C is very heterogenous and contains various organic-walled and agglutinated taxa. Placing the xenophyophores within this clade (Pawlowski et al., 2003; Gooday et al., 2017) was quite unexpected, given their large size, often bizarre morphologies, and unusual cellular organization. However, the transcriptomic data apparently confirm the 18S phylogeny, providing new evidence that xenophyophores are not some kind of early evolved group of protists, as suggested by some palaeontologists (e.g., Seilacher et al., (2003)), but a highly derived group of monothalamous Foraminifera (Antcliffe et al., 2011).

Another interesting finding concerns the root of the foraminiferal tree. In previous phylogenomic studies the root was placed either with the tubothalamid genus *Quinqueloculina* (Sierra et al., 2013) or *R. filosa* (Burki et al., 2013; Sierra et al., 2016; Krabberod et al., 2017). Earlier studies based on single protein families placed the root on *Astrammina rara* in the beta-tubulin 2 tree (Hou et al., 2013) or *Allogromia* sp. in the actin 1 and actin 2 trees (Flakowski et al., 2005) and RPB1 tree (Longet and Pawlowski 2007). Our transcriptomic data, however, show *Hippocrepinella hirudinea* and *Psammophaga fuegia* at the base of the foraminiferal tree. The former species belongs to monothalamid clade D and is characterized by a relatively large tubular test with an agglutinated wall. The second species represents clade E and is characterized by an organic or agglutinated test and the very particular feature of ingesting mineral grains (Gschwend et al., 2016). Both species may represent some early foraminiferal lineages, but it is probably too soon to consider them as close to the foraminiferal ancestor, especially given the limited number of analyzed monothalamous lineages.

To conclude, the phylogenomic data presented here confirm major trends of evolution in Foraminifera revealed by previous studies, with single-chambered, organic-walled, or agglutinated Monothalamea evolving at least twice into more complex multi-chambered Globothalamea and Tubothalamea characterized by agglutinated or calcareous tests. Within Monothalamea, our data confirm that megafauna-sized abyssal Xenophyophorea are a highly evolved lineage of single-chambered Foraminifera. Within Globothalamea, our study further supports the paraphyletic character of agglutinated Textulariida and the monophyly of calcareous Rotaliida.

CRediT authorship contribution statement

Roberto Sierra: Conceptualization, Data curation, Formal analysis, Writing – review & editing. **Florian Mauffrey:** Data curation, Formal analysis. **Joana Cruz:** Investigation. **Maria Holzmann:** Formal analysis. **Andrew J. Gooday:** Funding acquisition, Writing – review & editing. **Xyrus Maurer-Alcalá:** Investigation. **Rabindra Thakur:** Investigation. **Mattia Greco:** Investigation. **Agnes K.M. Weiner:** Investigation. **Laura A. Katz:** Funding acquisition, Writing – review & editing. **Jan Pawlowski:** Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability.

Data are available under accession numbers SAMN15195961 -

SAMN15195979 and alignments at https://doi.org/10.17632/pmscvgxf23.1.

Acknowledgments

We thank the sequencing platform of the University of Geneva. The Cushman Foundation for Foraminiferal Research and the Johanna M. Resig Foraminiferal Research Fellowship (RS) for supporting this study. This study was supported by the Swiss National Science Foundation grants 31003A_179125 and 316030_150817 (JP). Xenophyophores from the Clarion-Clipperton Zone were collected as part of the ABYSSLINE project, funded through a commercial arrangement with UK Seabed Resources Development Ltd.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107546.

References

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31, 2553–2556.
Adl, S.M., Bass, D., Lane, C.E., Lukes, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., et al., 2019. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. J. Eukaryot. Microbiol. 66, 4–119.
Antcliffe, J.B., Gooday, A.J., Brasier, M.D., 2011. Testing the Protozoan Hypothesis for Ediacaran Fossils: A Developmental Analysis of Palaeopascichnus. Palaeontology 54, 1157–1175.
Bernhard, J.M., Edgcomb, V.P., Visscher, P.T., McIntyre-Wressnig, A., Summons, R.E., Bouxsein, M.L., Louis, L., Jeglinski, M., 2013. Insights into foraminiferal influences on microfabrics of microbialites at Highborne Cay, Bahamas. Proc. Natl. Acad. Sci. U. S. A. 110, 9830–9834.
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.
Burki, F., Keeling, P.J., 2014. Rhizaria. Curr. Biol. 24, R103–R107.
Burki, F., Kudryavtsev, A., Matz, M.V., Aglyamova, G.V., Bulman, S., Fiers, M., Keeling, P.J., Pawlowski, J., 2010. Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. BMC Evol. Biol. 10, 377.
Burki, F., Roger, A.J., Brown, M.W., Simpson, A.G.B., 2020. The New Tree of Eukaryotes. Trends Ecol. Evol. 35, 43–55.
Bushmanova, E., Antipov, D., Lapidus, A., Prjibelski, A.D., 2019. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience 8.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421.
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.
Flakowski, J., Bolivar, I., Fahrni, J., Pawlowski, J., 2005. Actin phylogeny of foraminifera. J. Foramin. Res. 35, 93–102.
Flakowski, J., Bolivar, I., Fahrni, J., Pawlowski, J., 2006. Tempo and mode of spliceosomal intron evolution in actin of foraminifera. J. Mol. Evol. 63, 30–41.
Glockner, G., Hulsmann, N., Schleicher, M., Noegel, A.A., Eichinger, L., Gallinger, C., Pawlowski, J., Sierra, R., Euteneuer, U., Pillet, L., et al., 2014. The genome of the foraminiferan Reticulomyxa filosa. Curr. Biol. 24, 11–18.
Gooday, A.J., Holzmann, M., Caulle, C., Goineau, A., Jones, D.O.B., Kamenskaya, O., Simon-Lledó, E., Weber, A.A.T., Pawlowski, J., 2018. New species of the xenophyophore genus Aschemonella (Rhizaria: Foraminifera) from areas of the abyssal eastern Pacific licensed for polymetallic nodule exploration. Zool. J. Linn. Soc. 182, 479–499.
Gooday, A.J., Holzmann, M., Caulle, C., Goineau, A., Kamenskaya, O., Weber, A.A.T., Pawlowski, J., 2017. Giant protists (xenophyophores, Foraminifera) are exceptionally diverse in parts of the abyssal eastern Pacific licensed for polymetallic nodule exploration. Biol. Conserv. 207, 106–116.
Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.
Groussin, M., Pawlowski, J., Yang, Z., 2011. Bayesian relaxed clock estimation of divergence times in foraminifera. Mol. Phylogenet. Evol. 61, 157–166.
Gschwend, F., Majda, A., Majewski, W., Pawlowski, J., 2016. Psammophaga fuegia sp. nov., a New Monothalamid Foraminifera from the Beagle Channel. South America. Acta Protozool 55, 101–110.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.
Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075.
Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35, 518–522.
Holzmann, M., Gooday, A.J., Siemensma, F., Pawlowski, J., 2021. Review: Freshwater and Soil Foraminifera – A Story of Long-Forgotten Relatives. J. Foramin. Res. 51, 318–331.
Holzmann, M., Habura, A., Giles, H., Bowser, S.S., Pawlowski, J., 2003. Freshwater foraminiferans revealed by analysis of environmental DNA samples. J. Eukaryot. Microbiol. 50, 135–139.
Holzmann, M., Pawlowski, J., 2017. An updated classification of rotaliid foraminifera based on ribosomal DNA phylogeny. Mar. Micropaleontol. 132, 18–34.
Hou, Y., Sierra, R., Bassen, D., Banavali, N.K., Habura, A., Pawlowski, J., Bowser, S.S., 2013. Molecular evidence for beta-tubulin neofunctionalization in Retaria (Foraminifera and radiolarians). Mol. Biol. Evol. 30, 2487–2493.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Krabberod, A.K., Orr, R.J.S., Brate, J., Kristensen, T., Bjorklund, K.R., Shalchian-Tabrizi, K., 2017. Single Cell Transcriptomics, Mega-Phylogeny, and the Genetic Basis of Morphological Innovations in Rhizaria. Mol. Biol. Evol. 34, 1557–1573.
Lecroq, B., Gooday, A.J., Tsuchiya, M., Pawlowski, J.A.N., 2009. A new genus of xenophyophores (Foraminifera) from Japan Trench: morphological description, molecular phylogeny and elemental analysis. Zool. J. Linn. Soc. 156, 455–464.
Lefort, V., Longueville, J.E., Gascuel, O., 2017. SMS: Smart Model Selection in PhyML. Mol. Biol. Evol. 34, 2422–2424.
Longet, D., Pawlowski, J., 2007. Higher-level phylogeny of Foraminifera inferred from the RNA polymerase II (RPB1) gene. Eur J Protistol 43, 171–177.
Murray, J.W., 2006. Ecology and Applications of Benthic Foraminifera. Cambridge University Press, Cambridge.
Nauss, R.N., 1949. Reticulomyxa filosa gen. et sp. nov., a new primitive plasmodium. Bull. Torrey Bot. Club 76, 161–173.
Nettersheim, B.J., Brocks, J.J., Schwelm, A., Hope, J.M., Not, F., Lomas, M., Schmidt, C., Schiebel, R., Nowack, E.C.M., De Deckker, P., et al., 2019. Putative sponge biomarkers in unicellular Rhizaria question an early rise of animals. Nat. Ecol. Evol. 3, 577–581.
Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Pawlowski, J., 2012. Foraminifera. In: Schaechter, M. (Ed.), Eukaryotic Microbes. Elsevier, San Diego, USA, pp. 291–309.
Pawlowski, J., Bolivar, I., Fahrni, J.F., De Vargas, C., Bowser, S.S., 1999. Molecular evidence that Reticulomyxa filosa is a freshwater naked foraminifer. J. Eukaryot. Microbiol. 46, 612–617.
Pawlowski, J., Burki, F., 2009. Untangling the phylogeny of amoeboid protists. J. Eukaryot. Microbiol. 56, 16–25.
Pawlowski, J., Holzmann, M., 2014. A plea for DNA barcoding of Foraminifera. J. Foramin. Res. 44, 62–67.
Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Cedhagen, T., Bowser, S.S., 2002. Phylogeny of allogromiid foraminifera inferred from SSU rRNA gene sequences. J. Foramin. Res. 32, 334–343.
Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Gooday, A.J., Cedhagen, T., Habura, A., Bowser, S.S., 2003. The evolution of early Foraminifera. Proc. Natl. Acad. Sci. U. S. A. 100, 11494–11498.
Pawlowski, J., Holzmann, M., Tyszka, J., 2013. New supraordinal classification of Foraminifera: Molecules meet morphology. Mar. Micropaleontol. 100, 1–10.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.
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.
Seilacher, A., Grazhdankin, D., Legouta, A., 2003. Ediacaran biota: The dawn of animal life in the shadow of giant protists. Paleontolog. Res. 7, 43–54.
Siemensma, F., Apotheloz-Perret-Gentil, L., Holzmann, M., Clauss, S., Volcker, E., Pawlowski, J., 2017. Taxonomic revision of freshwater foraminifera with the description of two new agglutinated species and genera. Eur. J. Protistol. 60, 28–44.
Sierra, R., Canas-Duarte, S.J., Burki, F., Schwelm, A., Fogelqvist, J., Dixelius, C., Gonzalez-Garcia, L.N., Gile, G.H., Slamovits, C.H., Klopp, C., et al., 2016. Evolutionary Origins of Rhizarian Parasites. Mol. Biol. Evol. 33, 980–983.
Sierra, R., Matz, M.V., Aglyamova, G., Pillet, L., Decelle, J., Not, F., de Vargas, C., Pawlowski, J., 2013. Deep relationships of Rhizaria revealed by phylogenomics: a farewell to Haeckel's Radiolaria. Mol. Phylogenet. Evol. 67, 53–59.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Weiner, A.K.M., Ceron-Romero, M.A., Yan, Y., Katz, L.A., 2020. Phylogenomics of the Epigenetic Toolkit Reveals Punctate Retention of Genes across Eukaryotes. Genome Biol. Evol. 12, 2196–2210.
Yan, Y., Maurer-Alcala, X.X., Knight, R., Kosakovsky Pond, S.L., Katz, L.A., 2019. Single-Cell Transcriptomics Reveal a Correlation between Genome Architecture and Gene Family Evolution in Ciliates. MBio 10.