AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Single-Cell Genomics Reveals the Divergent Mitochondrial Genomes of Retaria (Foraminifera and Radiolaria)

Jan-Niklas Macher,[a] Nicole L. Coots,[b,c] Yu-Ping Poh,[b] Elsa B. Girard,[a,e] Anouk Langerak,[a] Sergio A. Muñoz-Gómez,[d] Savar D. Sinha,[b] Dagmar Jirsová,[b] Rutger Vos,[a] Richard Wissels,[a] Gillian H. Gile,[c] Willem Renema,[a,e] Jeremy G. Wideman[b,c]

[a]Naturalis Biodiversity Center, Marine Biodiversity Group, Leiden, The Netherlands
[b]Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, Arizona, USA
[c]School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[d]Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA
[e]University of Amsterdam, Department of Ecosystem & Landscape Dynamics, Institute for Biodiversity & Ecosystem Dynamics, Amsterdam, The Netherlands

Jan-Niklas Macher and Nicole L. Coots are co-first authors. The author order was determined based on initial concept ideas, the materials they provided, their seniority, and the extent to which they financially supported the analyses.

ABSTRACT Mitochondria originated from an ancient bacterial endosymbiont that underwent reductive evolution by gene loss and endosymbiont gene transfer to the nuclear genome. The diversity of mitochondrial genomes published to date has revealed that gene loss and transfer processes are ongoing in many lineages. Most well-studied eukaryotic lineages are represented in mitochondrial genome databases, except for the superphylum Retaria—the lineage comprising Foraminifera and Radiolaria. Using single-cell approaches, we determined two complete mitochondrial genomes of Foraminifera and two nearly complete mitochondrial genomes of radiolarians. We report the complete coding content of an additional 14 foram species. We show that foraminiferan and radiolarian mitochondrial genomes contain a nearly fully overlapping but reduced mitochondrial gene complement compared to other sequenced rhizarians. In contrast to animals and fungi, many protists encode a diverse set of proteins on their mitochondrial genomes, including several ribosomal genes; however, some aerobic eukaryotic lineages (euglenids, myzozoans, and chlamydomonas-like algae) have reduced mitochondrial gene content and lack all ribosomal genes. Similar to these reduced outliers, we show that retarian mitochondrial genomes lack ribosomal protein and tRNA genes, contain truncated and divergent small and large rRNA genes, and contain only 14 or 15 protein-coding genes, including *nad1*, *-3*, *-4*, *-4L*, *-5*, and *-7*, *cob*, *cox1*, *-2*, and *-3*, and *atp1*, *-6*, and *-9*, with forams and radiolarians additionally carrying *nad2* and *nad6*, respectively. In radiolarian mitogenomes, a noncanonical genetic code was identified in which all three stop codons encode amino acids. Collectively, these results add to our understanding of mitochondrial genome evolution and fill in one of the last major gaps in mitochondrial sequence databases.

IMPORTANCE We present the reduced mitochondrial genomes of Retaria, the rhizarian lineage comprising the phyla Foraminifera and Radiolaria. By applying single-cell genomic approaches, we found that foraminiferan and radiolarian mitochondrial genomes contain an overlapping but reduced mitochondrial gene complement compared to other sequenced rhizarians. An alternative genetic code was identified in radiolarian mitogenomes in which all three stop codons encode amino acids. Collectively, these results shed light on the divergent nature of the mitochondrial genomes from an ecologically important group, warranting further questions into the biological underpinnings of gene content variability and genetic code variation between mitochondrial genomes.

KEYWORDS Foraminifera, mitochondrial evolution, mitochondrial genome, Radiolaria, Retaria, Rhizaria

Editor Jianping Xu, McMaster University

Address correspondence to Jan-Niklas Macher, jan.macher@naturalis.nl, or Jeremy G. Wideman, Jeremy.Wideman@asu.edu.

The authors declare no conflict of interest.

Received 7 February 2023
Accepted 14 February 2023
Published 20 March 2023

Endosymbiosis, the functional integration of one cell into another, has shaped the evolution of eukaryotes (1, 2). The oldest extant endosymbioses, those of mitochondria and chloroplasts, both originate from ancient bacterial endosymbionts (3, 4). From their origin to the present, mitochondrial and chloroplast genomes have undergone reductive evolution by gene loss or endosymbiont gene transfer (EGT) to the nuclear genome (3, 5–7). For mitochondrial genomes, most of this reduction occurred prior to the emergence of the last eukaryotic common ancestor (LECA). What was once a bacterial genome encoding thousands of proteins became a mitochondrial genome encoding fewer than a hundred proteins in the LECA (3). Plants and many protist lineages still retain a diverse set of between 30 and 70 protein-coding genes on their mitochondrial genomes (3, 8–11). In addition to electron transport chain (ETC) components from complexes I to V, mitochondrial genomes can encode upward of 40 other proteins involved in transcription, translation, complex maturation, and transport (9, 12). In extreme cases of reduction, the highly reduced mitochondrion-related organelles (MROs) have entirely lost their organellar genomes (3, 13–15). Although these extreme cases of reduction are associated with anaerobic lifestyles, several aerobic lineages have also undergone major reductions in their mitochondrial genome content, mostly via EGT to the nucleus (16–19). Why the organelle retains certain mitochondrial genes is hotly debated (20–25), and it remains unclear what functional consequences varying gene complements might entail.

Mitochondria are best known for their role in synthesizing ATP using a proton gradient across their inner membrane (26). In fact, mitochondrial genomes largely code for proteins directly or indirectly related to the function of the ETC and ATP synthase (27). Furthermore, when the need for the ETC is lost in anaerobic lineages, the mitochondrial genome is also lost (14, 28). In aerobic lineages, a few components of the ETC (e.g., parts of complexes I, III, and IV) and the mitochondrial rRNAs are always encoded in mitochondrial genomes (29). Apart from these few components, all other genes can be lost, replaced (30), or transferred to the nuclear genome. In the lineages leading to animals and fungi, all genes encoding ribosomal proteins (except *rps3* in most fungi [31]) were independently transferred to the nuclear genomes (19). Thus, most animal and fungal mitochondrial genomes encode only rRNAs, tRNAs, and 13 or 14 proteins (fewer if complex I is lost, as in *Saccharomyces cerevisiae*), all of which are involved in electron transport or ATP synthesis (32).

In addition to the mitochondrial genome reductions seen in animals and fungi, some aerobic protist lineages also exhibit ancient reductions of their mitochondrial coding repertoire and contain many fewer proteins, few or no tRNAs, and highly divergent or fragmented rRNAs. For example, myzozoans, which include apicomplexans and dinoflagellates, encode only one to four ETC proteins in addition to highly fragmented and extremely divergent rRNAs (33–37); euglenids like *Euglena gracilis* encode 8 ETC proteins and very short divergent rRNAs (38, 39); and chlorophycean algae like *Chlamydomonas reinhardtii* encode 7 ETC proteins and extremely fragmented rRNAs (40, 41). In addition to these major aerobic lineages, one aerobic genus, the red alga *Galdieria* (42), has also lost its mitoribosomal proteins from its mitochondrial genomes via EGT. Thus, although mitochondrial genomes often encode many proteins, certain evolutionary circumstances result in massive gene loss or EGT, resulting in reduced mitochondrial genome coding content.

While several orphan taxa still lack representation in mitochondrial genome databases, only one major eukaryotic lineage is completely absent: the Retaria, the rhizarian lineage, comprising the phyla Foraminifera and Radiolaria (43). As a whole, rhizarians are important members of marine communities (44–47), contributing significantly to marine biogeochemical cycling (48–50). Retarians are aerobes (although some forams thrive under anoxic conditions [51]) and are abundant in many environments, especially in the global ocean. With ~9,000 recognized mostly marine extant species, Foraminifera are estimated to account for ~25% of present-day carbonate production (52, 53). Silicified Radiolaria, with their 600 to 800 named species, are estimated to account for between 2

and 19% of total biogenic silica production (54). Despite their importance, the paucity of retarian genomes and transcriptomes in sequence databases has made a deeper understanding of these lineages impossible (55–60).

In order to obtain mitochondrial genome sequences from Retaria, we chose to use single-cell approaches. Single-cell genomics can effectively recover mitochondrial genomes from diverse protists (10, 61, 62). Even though most species of Foraminifera and Radiolaria are not in culture, contain a multitude of symbionts (63, 64), and show high levels of intragenomic polymorphisms (65, 66), we show that single-cell approaches can effectively recover mitochondrial genomes from these complex assemblages. Our data demonstrate that foraminiferan and radiolarian mitochondrial genomes have an overlapping but reduced gene complement compared to other sequenced rhizarians, similar to other reduced mitochondrial genomes from other lineages. Retarian mitochondrial genomes do not encode ribosomal proteins or tRNAs. However, they do contain truncated and divergent small and large rRNAs and contain only 14 or 15 protein-coding genes, including *nad1*, *-3*, *-4*, *-4L*, *-5*, and *-7*, *cob*, *cox1*, *-2*, and *-3*, and *atp1*, *-6*, and *-9*, with forams and radiolarians additionally carrying *nad2* and *nad6*, respectively. An alternative genetic code was identified in radiolarian mitogenomes in which all three stop codons encode amino acids (TGA = W, TAG = Y, and TAA = Y/stop). These results further add to our understanding of mitochondrial genome evolution across the eukaryotic tree of life.

RESULTS AND DISCUSSION

Retarian mitochondrial, but not nuclear, genomes can be readily recovered using single-cell methods. We isolated individual cells and Illumina sequenced and assembled mini-metagenomes of 31 Foraminifera from 15 species (which are impossible to separate from their symbionts) and single-cell amplified genomes (SAGs) of 5 Radiolaria from 2 species (see Table S1 in the supplemental material for a complete list). Foram mini-metagenomes are referred to here as SAGs. One additional foraminiferan metagenome (*Globobulimina* sp.) was downloaded and reassembled from the NCBI sequence read archive (SRA accession number SRX3312059 [67]). Assemblies from *Calcarina*, *Neorotalia*, *Lithomelissa*, and *Acanthometra* SAGs are available for BLAST at SAGdb (https://evocellbio.com/SAGdb/macher_et_al/).

Both forams and radiolarians associate with many eukaryotic and bacterial endosymbionts (68–70), making it difficult to obtain *bona fide* sequence data from either lineage. To assess the contamination in foram and radiolarian SAGs, we collected all 18S and 16S sequences from all assemblies using *Cafeteria roenbergensis* 18S and *Escherichia coli* 16S sequences as BLAST queries. We found foraminifera 18S genes or gene fragments in 23 of 31 SAGs from 13 of 15 species (Table S1). We also identified specific symbiont 28S sequences (from a dinoflagellate symbiont) or *rbcL* (from diatom symbionts) in all foram SAGs and species except *Calcarina mayori* and the reassembled *Globobulimina* (which is expected, since *Globobulimina* does not contain photosymbionts) (Table S2). The inability to identify the foram 18S genes in all specimens is likely due to their extreme within-cell variability (65, 71), which prevented proper assembly. 16S BLAST searches recovered diatom symbiont chloroplast and mitochondrial genes. In addition, 16S sequences from two common bacterial genera were also recovered (*Burkholderia* and *Cutibacterium*). Blob plots from foram assemblies confirm 18S BLAST findings as large contigs of symbiont organelles (Fig. S1A and B). From these data, we concluded that our foram SAG assemblies predominantly contain symbiont contigs, with only some representation from the host nuclear genome.

In radiolarians, we obtained high-coverage contigs with complete 18S sequences only from radiolarians (Table S2). In *Acanthometra* and *Amphibelone* (*nc69*, *-78*, *-87*, and *-96*) SAGs, a few contaminating 18S sequences (e.g., diatom, cryptophyte, and ciliate) were detected, but these contigs were fragmented with low coverage, indicating relatively few eukaryotic contaminants. Similarly, only a few fragmented low-coverage 16S contigs were recovered, again indicating very little prokaryotic contamination. These results are corroborated by blob plots showing relatively little contamination in acantharian SAGs (Fig. S1C). In the *Lithomelissa* SAG (r2m), only radiolarian 18S sequences

were recovered. However, many high-coverage bacterial contigs containing 16S sequences were identified, indicating that eukaryotic contamination in this SAG was low but bacterial contamination was very high. These 18S and 16S results are corroborated by blob plots showing a degree of bacterial and eukaryotic contamination but a large proportion of "unknown" reads with no similar hits in the NCBI nonredundant database (Fig. S1D). To assess the contamination of our nuclear data, we used a phylogenetic placement approach to assess SAG contamination (Fig. S2). Forams were excluded because they lacked sufficient identifiable nuclear contigs. Briefly, we extracted BUSCO proteins from SAG assemblies and added them to existing alignments from EukProt (72). Even with low radiolarian BUSCO scores (nc69, 10.6%; nc78, 5.1%; and nc96, 6.3% for *Acanthometra* SAGs; nc87, 3.9% for *Amphibelone* SAGs; and 11.4% for *Lithomelissa* SAGs), *Acanthometra* SAGs were correctly placed with full support alongside the only radiolarian (*Astrolonche serrata*) in the EukProt data set (Fig. S2). Conversely, the *Lithomelissa* SAG was placed within alveolates with full support (Fig. S2), suggesting unseen eukaryotic contamination, even though no contaminating 18S could be detected. Collectively, these data indicate that our radiolarian SAGs contain a substantial amount of radiolarian nuclear contigs, though the *Lithomelissa* SAG contains a large degree of bacterial and possibly eukaryote contamination.

Since mitochondrial genomes are often overrepresented in genome assemblies (10), we sought to identify foram and radiolarian mitochondrial genomes in our SAGs. Using protein sequences encoded by the *Andalucia godoyi* mitochondrial genome, one of the most gene-rich mitogenomes known (12), we identified several putative retarian mitochondrial contigs in foram and most radiolarian SAGs. *Amphibelone* SAG nc87 (96% identical 18S sequences to other nc SAGs) lacked any obvious mitochondrial contigs and was not investigated further. Since many assemblies exhibited both eukaryotic and prokaryotic contamination, great care was taken to inspect the validity of each contig manually. In forams, the putative mitochondrial contigs had orders-of-magnitude-higher read coverage and much lower GC content than symbiont or putative foram nuclear contigs (seen clearly in blob plots [Fig. S1A and B]). Contigs representing nearly complete or complete symbiont organelle genomes were also found in many foram SAGs, though these contigs had much lower coverage than the foram mitochondrial genomes (Fig. S1A and B). In radiolarians, the results were less clear-cut. While the coverage (~40 to 60×) of the putative mitochondrial contigs was much higher than the median for the SAG (~3 to 5× for the *Acanthometra* SAGs and ~11 for *Lithomelissa*, likely due to some very high-coverage contigs [Fig. S3]), the GC content was similar to that of the putative nuclear contigs (Fig. S1). Thus, though these contigs had relatively high coverage, they were not clearly separated from the majority of contigs in blob plots. The coverage of mitochondrial genome contigs and contigs containing radiolarian 18S sequences had similar coverage (~40 to 60×). Since both mitochondrial genomes and 18S sequences are generally found in multiple copies in a cell, we reasoned that we likely sequenced *bona fide* mitochondrial genomes and not nuclear mitochondrial genomes (NuMts), which would likely have much lower read coverage.

Retarian mitochondrial genomes carry a reduced gene complement. From each assembly, we extracted mitochondrial contigs collectively representing the putatively complete mitochondrial gene complement from 16 foraminiferan and two radiolarian species (Fig. 1). We obtained complete circular-mapping mitochondrial genomes of the forams *Calcarina hispida* and *Neorotalia gaimardi*. The mitogenomes were 46 kb (*Calcarina hispida*) and 50 kb (*Neorotalia gaimardi*) long, and each had the same set of 14 protein-coding genes (Fig. 2). We recovered contigs with mitochondrial genes from two species of radiolarians (*Lithomelissa* sp. and *Acanthometra* sp.) that almost completely overlap the foraminiferan complement (Fig. 1A). We concluded from these data that we likely extracted the complete, or nearly complete, coding complement of these radiolarian mitochondrial genomes. However, we were unable to recover complete circular mitochondrial genomes from either radiolarian, likely due to repetitive intergenic regions that prevented proper assembly. To complete these genomes, we attempted

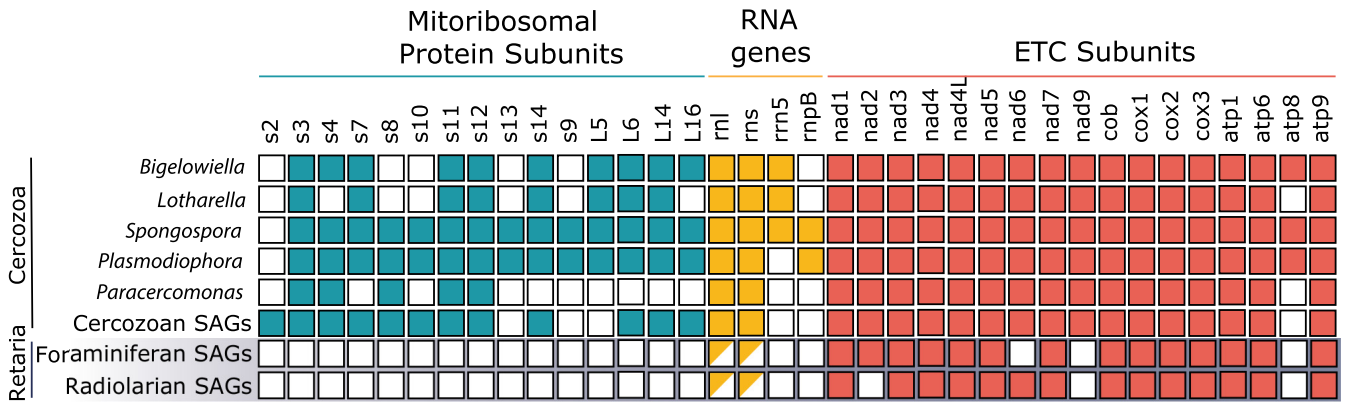

FIG 1 Mitochondrial genomes of forams and radiolarians overlap in gene content. Gene content of rhizarian mitochondrial genomes. Half-filled boxes indicate the presence of fragmented and shortened genes. Mitochondrial genes present in other mitochondrial genomes but absent from all sequenced rhizarian mitochondrial genomes are not listed. "Cercozoan SAGs" refers to the singly amplified genomes published previously (10).

to link contigs using primers designed to PCR amplify missing regions between contigs but were unsuccessful, likely due to complex repetitive regions.

Between the two phyla, retarians displayed a near-identical gene complement, including cytochrome *c* oxidase subunits (*cox1*, *cox2*, and *cox3*), cytochrome *b* (*cob*), and ATP synthase subunits (*atp1*, *atp6*, and *atp9*) and NADH dehydrogenase subunits (*nad1*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad7*; *nad2* is missing from radiolarians and *nad6* is missing from foraminifera), The lack of *nad2* in both radiolarian mitochondrial genomes and the lack of *nad6* in one radiolarian and both foram mitochondrial genomes is not without precedent, as both are either lost or extremely diverged and transferred to the nuclear genome in euglenids (73). Fragments of large-subunit (LSU)-rRNA (*rrnL*) and small-subunit (SSU)-rRNA (*rrnS*) genes were identified in the mitogenomes of the foraminiferans *Calcarina hispida* and *Neorotalia gaimardi* and the radiolarian *Lithomelissa* sp., while only *rrnL* was identified in the mitogenome of the radiolarian *Acanthometra* sp.; however, neither full-length ribosomal protein-coding genes nor tRNAs were detected (Fig. 1). The *nad9* gene was not found in our retarian mitochondrial genomes, even though all other sequenced rhizarian mitochondrial DNAs contain this gene (10, 74–77) (except for *Brevimastigomonas*, which has lost complex I altogether). Since most core complex I subunit genes appear to be retained in rhizarian mitogenomes (including those of retarians), missing complex I genes could be carried by the nuclear genome; however, these genes have not been identified in the nuclear genomes of euglenids (73). However, we were unable to find any complex I components in the nuclear assemblies, likely indicative of their incompleteness (BUSCO scores $<$ 10%). Another conspicuous absence from retarian mitochondrial genomes is *atp8*, which encodes subunit 8 of ATP synthase. Subunit 8 is likely an essential component of ATP synthase in most organisms (77) but appears to be absent in *Caenorhabditis elegans* (78) and cannot be identified in many rhizarian mitochondrial genomes (77, 79). To further confirm that we collected *bona fide* mitochondrial contigs, we reconstructed the phylogeny of forams with radiolarians as an outgroup using concatenated mitochondrial proteins predicted from the contigs (Fig. S4). The resulting phylogeny at the family level recapitulates the topology seen in 18S rDNA trees of Foraminifera (80–82), except for Peneroplidae clustering within the Soritidae.

We also obtained 25 fragmented mitochondrial genomes from 14 additional foraminiferan species (see Table S1 for a list of samples) that could not be linked in a single contig but had the same set of 14 protein-coding genes (all for ETC subunits) present in the circular-mapping mitochondrial genomes of *Calcarina hispida* and *Neorotalia gaimardi*. We also downloaded the available (meta)genomes of the foraminiferans *Reticulomyxa filosa* (55) and *Astrammina rara* (56) but could not identify mitochondrial genes.

Retarian mitochondrial genomes have large intergenic regions. Since we found large intergenic regions in both foraminiferan and radiolarian mitochondrial genomes, we conducted searches for genes or gene fragments within these intergenic regions

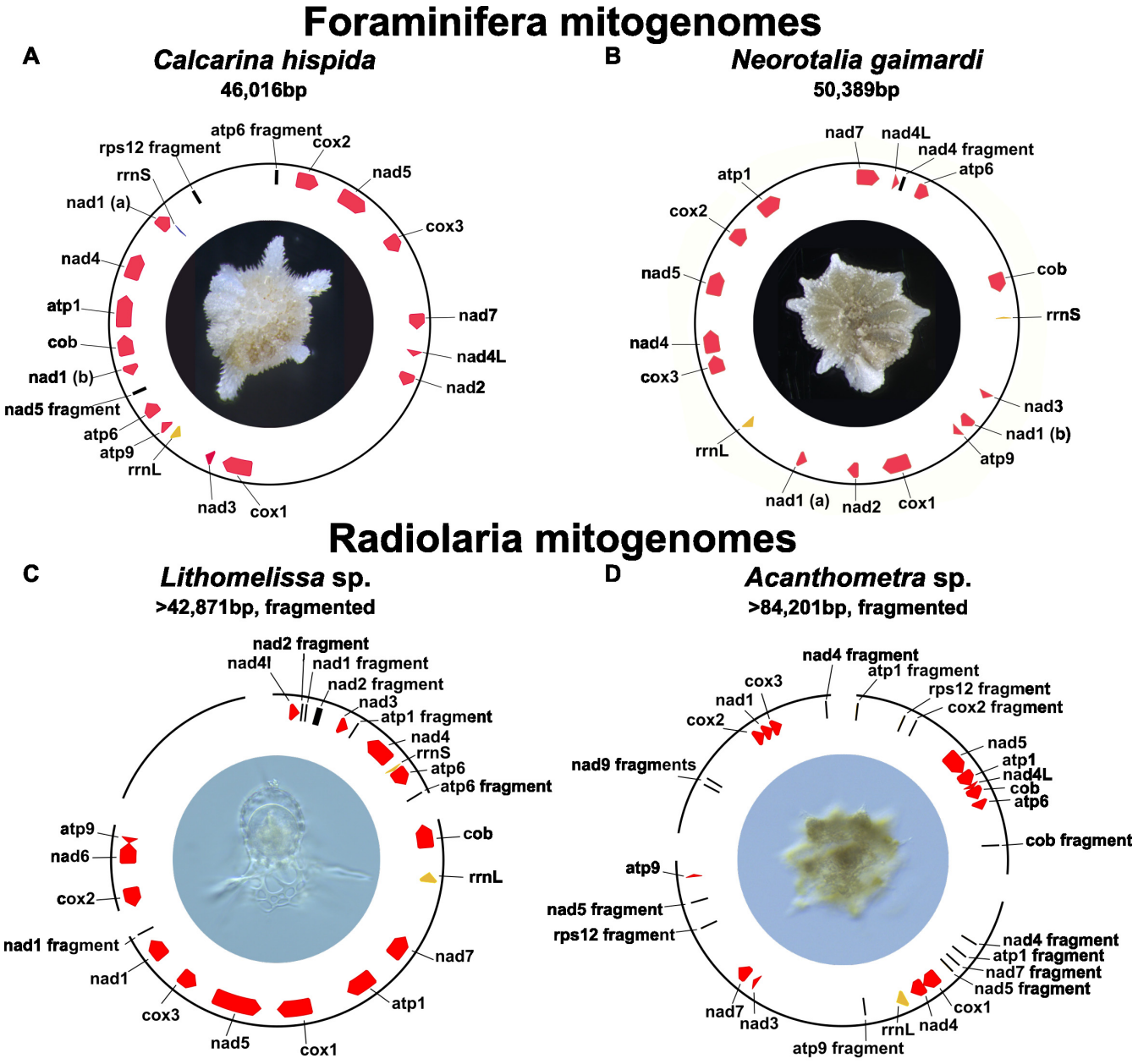

FIG 2 Retarian mitochondrial genomes contain large noncoding intergenic regions. Complete circular-mapping of mitochondrial genomes of the Foraminifera *Calcarina hispida* (A) and *Neorotalia gaimardi* (B) and the inferred nearly complete mitogenomes of the Radiolaria *Lithomelissa* sp. (C) and *Acanthometra* sp. (D). Protein-coding genes are highlighted in red; rRNAs (rrnL and rrnS) are highlighted in yellow. Gene fragments are shown in black. Gaps in the radiolarian mitochondrial genomes show the start and end of assembled mitochondrial contigs. Photos were taken of each individual organism before lysis.

using blastx (v.2.11.0) (83), mfannot (https://github.com/BFL-lab/Mfannot), and hmmer (v3.3.2) (84). Twenty-four regions were identified as putatively homologous to genes typically encoded by rhizarian mitochondrial genomes (Fig. 2, black lines). Eighteen of these are very similar to fragments of genes present elsewhere within the retarian mitochondrial genomes (*atp1*, *atp6*, *cob*, *nad4*, *nad5*, *nad7*, and *cox2*). The remaining six fragments are homologous to genes normally present in rhizarians, including *nad2* and *nad9* in radiolarians and *rps12* in a foram and a radiolarian. These fragments could represent pseudogenes or horizontally transferred DNA sequences (85, 86) or could reflect past genomic recombinations and rearrangements. In all mitochondrial contigs, small (~50-bp) stretches were nearly identical in many places, differing by one or two nucleotides. In

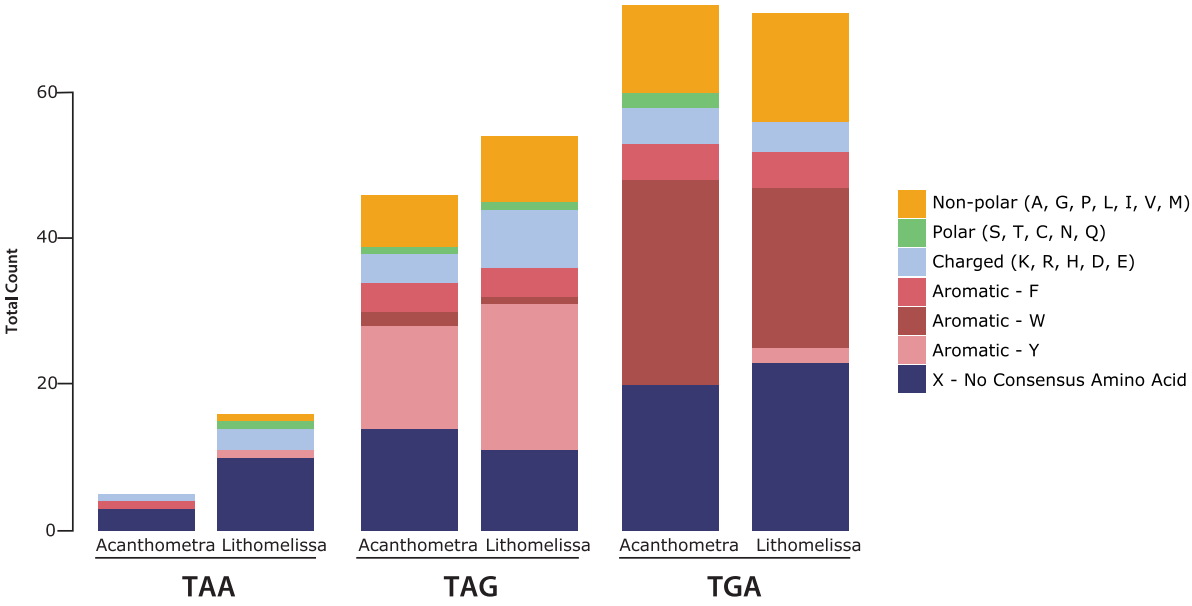

FIG 3 All three stop codons have been recoded to sense codons in radiolarian mitochondrial genomes. Proteins encoded in diverse mitochondrial genomes were aligned with their orthologues from radiolarians. Consensus (>50%) amino acids at sites containing internal radiolarian stop codons (TGA, TAG, and TAA) were noted. Amino acids were grouped based on their biochemical properties (nonpolar, polar, charged, or aromatic). When a large proportion of sites are occupied by a particular amino acid, this suggests that the in-frame stop codon encodes that amino acid.

the *Lithomelissa* sp. SAG, a large contig with similar read coverage was detected that contained these ~50-bp pseudorepeats but no mitochondrial genes or fragments (Fig. 2C). Perhaps the missing mitochondrial rRNAs have diverged beyond recognition.

All three standard stop codons are likely recoded in radiolarian mitochondrial genomes. Deviations from the ancestral standard genetic code have evolved in numerous lineages (87, 88). In particular, lineages with extremely low GC content and limited opportunities for recombination (i.e., organellar genomes) exhibit genetic code changes more frequently (89, 90). One common trend of genetic code variability, and the easiest to detect, is when stop codons are reassigned as sense codons. The most common version of stop codon reassignment by far is the TGA stop codon being recoded to tryptophan (normally encoded only by TGG) (91). This change has occurred several times across mitochondrial genomes and in other bacterial lineages. The TAA and TAG stop codons can also be recoded. For example, in the mitochondrial genome of certain thraustochytrid stramenopiles, a new stop codon (TTA) (GenBank accession no. AF288091.2) evolved and, in some species, both TAA and TAG were recoded to tyrosine (normally encoded by TAT and TAC) (10). All three stop codons have been recoded in the nuclear genomes of the ciliate *Condylostoma magnum* (92) and the kinetoplastid genus *Blastocrithidia* (93, 94). In both cases, TGA encodes tryptophan, and TAA and TAG encode glutamine (normally encoded only by CAA and CAG). For *Blastocrithidia*, authors showed that highly expressed genes have fewer TAA and TAG codons and speculate that changes in tRNA usage enable ribosomes to read through TAA and TAG sense codons in the middle of genes with moderate and low levels of expression, while TAA is still used as a termination codon at the end of transcripts (93, 94). Here, we identified a similar example in radiolarian mitochondria, where all three stop codons are likely recoded to sense codons.

To determine the genetic code of radiolarian mitochondrial genomes, we aligned predicted proteins with mitochondrion-encoded proteins from diverse protists. These alignments revealed that in-frame TGA and TAG codons occur at sites often occupied by tryptophan and tyrosine residues, respectively (Fig. 3). Conversely, relatively few in-frame TAA codons are present in conserved domains. The majority of in-frame TAA codons occurred at locations for which there was no consensus amino acid in the alignment (Fig. 3). In Acanthometra, only two genes contained in-frame TAAs (*nad5* [8 TAAs] and

cox2 [2 TAAs]). All eight *nad5* TAA codons were in the 3′ region, which appears to have diverged compared to the same region of other *nad5* genes. Similarly, the two *cox2* TAA codons were also in regions of the gene that are not highly conserved. In *Lithomelissa*, eight genes contained in-frame TAA codons. Like *Acanthometra*, *Lithomelissa* contained a few TAAs that aligned with conserved residues (one glutamine, one arginine, and the other tyrosine) in the middle of protein alignments (*nad7* and *cox3*).

When assessing pairwise alignments of the radiolarian proteins (e.g., pairwise alignment of *Lithomelissa* and *Acanthometra cox1*), of the 29 in-frame TAA codons, nearly half (14 TAAs) aligned with a tyrosine, phenylalanine, or tryptophan, and the majority (19 TAAs) aligned with a hydrophobic residue. In addition, 25 of 27 radiolarian mitochondrial protein-coding genes had a TAA codon near where the end of the protein is predicted. Two *Acanthometra* genes (*nad4L* and *cox2*) lacked stop codons and were contiguous with the open reading frames of *cob* and *nad1*, respectively. These data all suggest that a mechanism similar to the one proposed for the *Blastocrithidia* nuclear genome may be in place in radiolarian mitochondrial genomes. While TGA and TAG encode tryptophan and tyrosine, respectively, TAA appears to have a dual role, likely encoding tyrosine in some proteins at a few locations but acting primarily as a stop codon. Curiously, several proteins lack any in-frame TAA codon. Perhaps, similar to the case in *Blastocrithidia*, the most highly expressed mitochondrial proteins lack in-frame stop codons. The *atp9* gene is among the most highly expressed and has no TAA or TAG present in either radiolarian. These data indicate that the mitochondrial genetic code in radiolarians has diverged from the ancestral code and has recoded all three stop codons to code for amino acids. While TGA and TAG are recoded to tryptophan and tyrosine, TAA codons sometimes encode tyrosine but are the primary, and likely only, stop codon.

Retarian mitochondrial genomes contain fragmented rRNA genes, divergent *atp6* genes, and split *nad* genes. In most eukaryotic lineages, mitochondrial genomes encode a combination of proteins involved in electron transport and ATP synthesis, ribosomal proteins, and a few auxiliary proteins involved in protein maturation or translocation (3). However, five major lineages (euglenids, retarians, chlorophycean algae, myzozoans, and animals [Fig. 4]) have completely transferred all genes for mitoribosomal proteins to the nucleus, and two others are close behind (fungi contain only *rps3*, and glycomonads [Euglenozoa] contain at most *rps3* and *rps12*) (34, 77, 95–97). In all these lineages except animals and fungi, the EGT of mitoribosomal proteins has coincided with an extreme reduction or fragmentation of the mitochondrial rRNAs (Fig. 4, dark blue circles) (98). Animal and fungal mitochondrial rRNAs are truncated, but not to the extent of other mitochondrial rRNAs that are extremely divergent and nearly undetectable.

In addition to highly divergent rRNA genes, euglenozoans, retarians, chlorophycean algae, and myzozoans possess extremely divergent *atp6* genes (Fig. 4, yellow circles). Since a few TAA codons appear in the 5′ region of the *Lithomelissa atp6* gene, we decided to model the Atp6 protein using Alphafold2 (99) to determine if the N-terminal extension is part of the protein or represents a noncoding upstream sequence. Alphafold2 modeled full-length subunit a into a structure that better resembles the classic subunit a (Fig. S5). This suggests that the TAA codons are in part of the coding sequence of *Lithomelissa atp6*. Divergence of ATP synthase structure can have consequences for mitochondrial crista morphologies (e.g., chlorophycean algae, euglenids, kinetoplastids, and apicomplexans all have unique crista morphologies) (100–103). Since mitochondrial crista architecture that departs from classic lamellar and tubular morphologies present in most other eukaryotes has also been reported for Foraminifera (104) and Radiolaria (105, 106), retarian ATP synthase structures represent excellent candidates for future investigation into the structural and functional diversity of this amazing protein complex (107).

Curiously, *nad1* is split into two parts in the Foraminifera *Calcarina hispida* and *Neorotalia gaimardi*. This suggests that some trans-splicing might be present in Retaria, similar to what has been reported for mitochondrial genes in other eukaryotes (108, 109). It is also possible that two peptides are separately expressed and merged into a functional protein, as has been found in *Chromera* plastids (110). Furthermore, we identified a

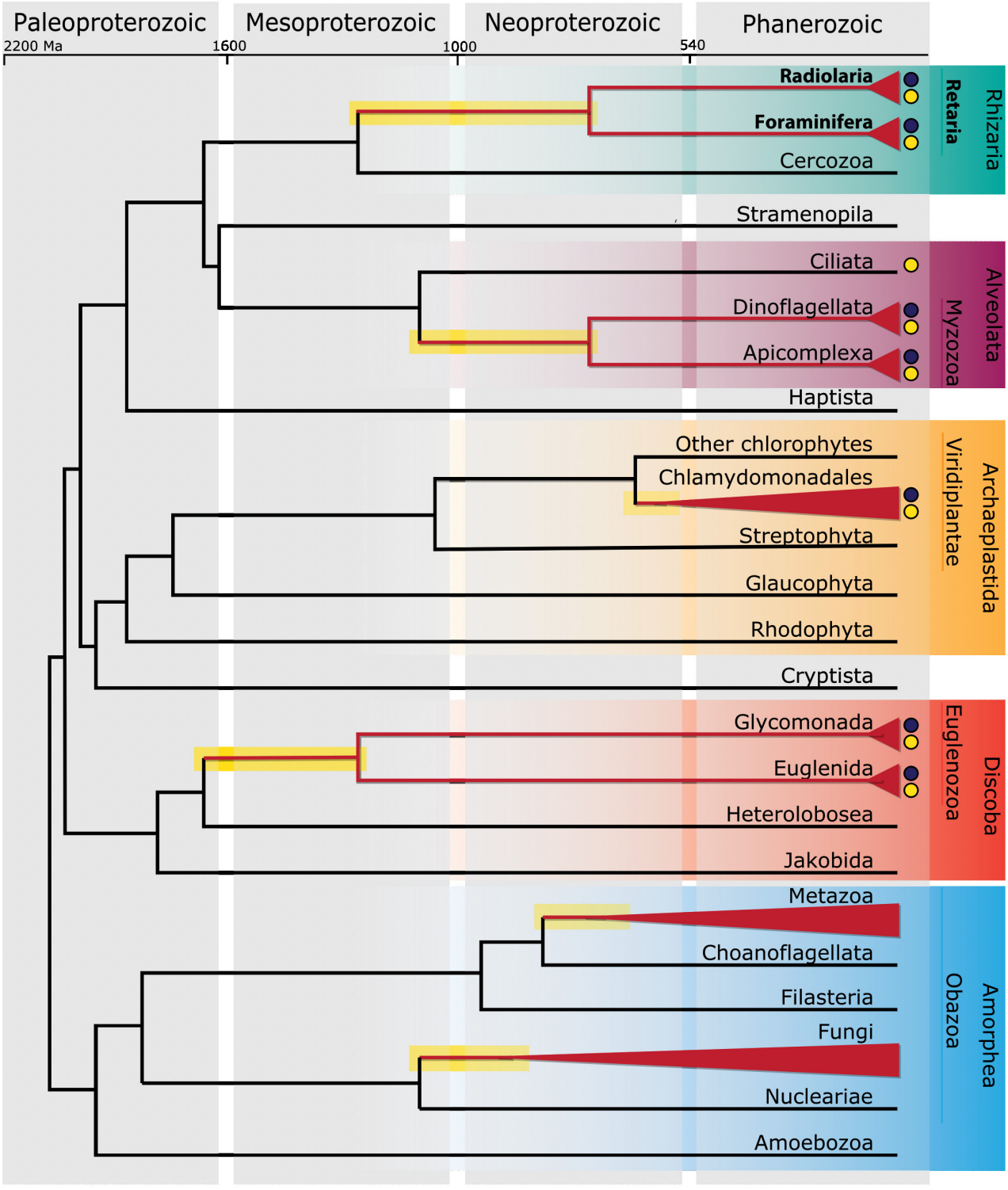

FIG 4 Divergent mitochondrial traits have persisted for hundreds of millions of years. Schematic phylogeny of extant eukaryotes with divergence times approximated based on reference 118. Clades highlighted in red have lost all mitoribosomal protein-coding genes from the mitochondrial genome. Dark blue circles indicate groups with short or fragmented mitochondrial rRNA genes. Yellow circles indicate groups with highly divergent mitochondrial *atp6* genes. Lineages possessing these traits diverged in the mid-Neoproterozoic or earlier (emphasized with a branch highlighted in yellow).

conserved frameshift in *cox1* of all four analyzed species of the foraminiferan order Miliolida, which suggests that a mechanism for stop codon read-through or posttranscriptional mRNA modification of this codon exists in this lineage. Manual insertion of a single N into the miliolid sequences resulted in a continuous open reading frame (ORF), which, when translated, spans the entire length of the *cox1* protein sequence (111). Posttranscriptional insertion modifications have evolved in several protist lineages, including euglenids and diplonemids (95, 112). As the same pattern was found in all analyzed miliolid Foraminifera but not in any rotaliid species, we conclude that this is a unique feature of Miliolida mitochondria, which might be of interest in the future characterization of this group.

Conclusions. Why do mitochondrial genomes vary so drastically across eukaryotes? Specifically, what triggers the wholesale transfer of mitochondrial ribosomal genes to the nucleus in so many lineages? There are several possible benefits to mitochondrion-to-nucleus gene transfer (113), and given enough time, mitochondrion-to-nucleus EGT is considered mathematically inevitable (114). Perhaps the diversity of mitochondrial genomes is simply a result of these evolutionary forces playing out over billions of years, with no functional cell biological consequences. However, this seems a somewhat unsatisfying answer given the diversity of mitoribosomal structures that have recently been solved (115–117).

In any case, retarian mitochondrial genomes represent a newly discovered ancient independent reduction in organellar gene content. The reduced gene complement of retarians displays more similarities to the mitochondrial genomes of euglenozoans, myzozoans, chlorophycean algae, animals, and fungi than it does to that of other rhizarians (Cercozoa). While several of these lineages may seem obscure and disparate, it is important to note that each lineage diverged in the mid-Neoproterozoic or earlier (118) (Fig. 4). These lineages therefore possess histories as deep and rich as those of animals and fungi, which are each traditionally considered independent "kingdoms." Given the ancient divergence of forams and radiolarians, the strikingly reduced mitochondrial genomes of Retaria have persisted without substantial change for over 500 million years. The persistence of mitochondrial gene content over large time spans suggests that mitochondrion-to-nucleus gene transfer does not occur consistently but rather occurs in relatively short macroevolutionary bursts. Further investigations into more deeply branching taxa at nodes of apparent sudden mass EGT will clarify this notion. In sum, the retarian mitochondrial genomes presented here bridge a major gap in our understanding and provide the first glimpse into the mitochondria of this diverse group of ancient protists.

MATERIALS AND METHODS

Sample collection. (i) Foraminifera samples. We analyzed 31 benthic Foraminifera cells (15 species) from the Spermonde Archipelago in Indonesia and from Coral Bay in Australia (see Table S1 for samples and locations). All specimens were stored in >90% ethanol after sampling and transferred to the Naturalis Biodiversity Centre laboratory for morphological species identification and molecular analyses. Specimens were sorted into morphotypes and identified and photographed using a ZeissDiscovery v12 stereomicroscope (Zeiss, Oberkochen, Germany).

(ii) Radiolaria samples. Marine surface water plankton samples were collected from the Pacific Ocean near the California coast (33.454219, −117.705215) by towing an 87-μm-mesh plankton net from the back of a kayak on 7 February 2021 at 10:00 a.m. Bulk environmental plankton samples were immediately aliquoted into 15-mL Falcon tubes and preserved with RNAlater. Plankton samples were stored on ice during transportation to the lab. Radiolarian cells were identified by morphology and imaged prior to single-cell isolation under an inverted microscope using a micropipette. Individual cells were washed four times in DNase- and RNase-free water to remove extracellular material from each radiolarian. This process was repeated twice with new water each time before each cell was transferred to 4 μL of RNAlater and then stored at −20°C before further processing.

DNA extraction and sequencing. (i) Foraminifera. Single Foraminifera specimens were dried in sterile 1.5-mL Eppendorf tubes and ground to a fine powder using a porcelain mortar and pestle. Total genomic DNA extraction was carried out using the QIAamp DNA Micro kit (Qiagen; Hilden, Germany) as described in reference 111. After extraction, DNA quantification was conducted using the FragmentAnalyzer system (Agilent Technologies, Santa Clara, CA, USA). Since extracted DNA was already fragmented to an average length of less than 500 bp, no further fragmentation using ultrasonication or enzymes was conducted.

Shotgun metagenomic libraries were prepared using the NEBNext Ultra II DNA library preparation kit (New England Biolabs, Ipswich, MA, USA) with the corresponding NEBNext multiplex oligonucleotides

for Illumina, following the manufacturer's protocol but reducing volumes by 50%. Final concentration and fragment size were checked on the Tapestation system (Agilent Technologies, Santa Clara, CA, USA). All samples were pooled in equimolar amounts before being sent for sequencing on the Illumina NovaSeq 6000 platform (2 × 150-bp read length) at Baseclear (Leiden, The Netherlands), targeting 5 million reads per sample.

(ii) Radiolaria. Single-cell DNA extractions were performed using the MasterPure DNA and RNA purification kit (Epicentre Biotechnologies) following the protocol as written, with the addition of a 30-min incubation with a solution of lysis buffer and proteinase K at 65°C and 1,000 rpm. Purified total genomic DNA was eluted into 4 μL Tris-EDTA (TE) buffer and quantified using a Qubit HS double-stranded-DNA (dsDNA) kit.

Genomic DNA from each cell was amplified using the Repli-G Advanced DNA single-cell kit and protocol (Qiagen) for amplifying purified genomic DNA. Final concentration and fragment size were checked using the Tapestation and Qubit systems. An aliquot of each singly amplified genome containing a total of 500 ng DNA was provided to the ASU Genomics Facility for library preparation using KAPA Biosystem's LTP library preparation kit before the samples were sequenced on the Illumina NovaSeq 6000 platform targeting 10 million 2 × 150-bp reads per sample.

Bioinformatic analysis. (i) Foraminifera. MultiQC (119) was used for the quality assessment of raw reads. Megahit (120) was used for the initial assembly of reads into contigs, which were loaded into Geneious Prime (v.2020) together with raw reads. Contigs were mapped against the mitochondrial genome of the rhizarian *Lotharella oceanica* deposited in GenBank (accession number NC_029731.1 [77]) with up to 50% mismatch, a word length of 5, and up to 10% gaps (gap size, 10) allowed. Since none of the assembled contigs could be mapped, raw reads and contigs were mapped against the *L. oceanica* reference with the settings mentioned above and against Foraminifera mitochondrial cytochrome oxidase subunit I (COI) barcode sequences published in reference 121. Regions with high coverage of mapped reads or with mapped contigs were manually inspected. When mapped contigs did not represent a full mitochondrial genome (which was the case only for *Neorotalia gaimardi*), mapped reads were used as a reference for repeated mapping with the Geneious Prime mapper, with a minimum of 100 bp overlap, a maximum of 1% mismatch, and no gaps allowed. Mapping was repeated until no further reads could be mapped. The resulting contigs were checked for ORFs with mitochondrial translation table 4, which was reported previously for protist mitochondrial genomes (10).

Contigs were submitted to the mfannot mitochondrial annotation web server of the University of Montréal (https://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl). ORFs identified as the cytochrome oxidase subunit 1 gene (*cox1*) were searched against the NCBI GenBank reference database (122) and the Foraminifera *cox1* database (121) using BLASTn to identify the *cox1* sequence stemming from putative symbionts and the putative foraminiferal *cox1*. Annotations were manually curated in Geneious Prime. ORFs that were not annotated by mfannot were translated to proteins, subjected to transmembrane prediction with TMHMM (123), and searched against Pfam (124), UniProt (125), SwissProt (126), and Ensembl (127) databases using the hmmer web server (84) to check for potential matches with known mitochondrial genes. When a complete mitochondrial genome could not be obtained, the putative foraminiferal mitochondrial genes were identified by mapping reads against the newly assembled *Calcarina hispida* and *Neorotalia gaimardi* mitochondrial genomes as described above.

To verify that foraminiferal mitochondrial genes could also be obtained from previously published data sets, we downloaded the *Globobulimina* (order Rotaliida) metagenome from the NCBI Sequence Read Archive (accession number SRX3312059 [67]) and assembled the foraminiferal mitochondrial genes as described above. Furthermore, we downloaded the genomic contigs of the foraminiferans *Reticulomyxa filosa* (55) and *Astrammina rara* (56) and searched for mitochondrial genes as described above, though none could be found.

(ii) Radiolaria. MultiQC (119) was used to trim and filter raw fastq reads, which were then normalized with BBNorm (an addition to BBMap v.38.12). SAGs were assembled using SPAdes (v.3.15.2) (128). Normalized reads were mapped back to contigs with BBMap, and genome completeness was assessed with BUSCO (v.5.1.2) (129). BlobTools (v.1.0) (130) was used to visualize contigs with similar read coverage and GC content. Mitochondrial contigs were identified using *Andalucia godoyi* mitochondrion-encoded proteins as queries in tblastn searches against radiolarian SAG assemblies. The mitochondrial contigs identified were manually stitched together by identifying regions with overlaps of >50 bp between contigs with similar read coverages.

Putative mitochondrial contigs were submitted to the mfannot mitochondrial annotation (https://megasun.bch.umontreal.ca/apps/mfannot/) web server. Because mfannot did not identify full-length rRNA genes within our mitochondrial genomes, nhmmer (131) was used to search each genome for rRNA genes using manually curated rRNA databases. Fragments of mitochondrial genes were also identified by searching intergenic regions and open reading frames that were not annotated by mfannot against a manually curated database of mitochondrial protein sequences with representatives from all protist genera with a sequenced mitochondrial genome in NCBI GenBank using blastx. Intergenic regions and ORFs with at least four hits from the same gene were considered significant enough for annotation on the mitochondrial genome maps. Annotations were added manually using Geneious Prime.

Stop codon analysis. Amino acid multiple sequence alignments were used to assess the locations within a mitochondrial gene at which radiolarians have a stop codon. Alignments were generated with MUSCLE (132) using radiolarian genes identified by mfannot and genes from every available protistan genus in GenBank. If more than one mitochondrial genome existed for a genus in GenBank, then the most recent two mitochondrial genomes from different species were chosen as representatives of that genus. The total number of stop codons present within each mitochondrial gene from the two radiolarian

mitochondrial genomes was visually counted. The 50% consensus amino acid identities at locations for which a radiolarian mitochondrial gene had an in-frame stop codon were also tallied to assess which amino acids radiolarian stop codons are potentially coding for instead. Radiolarian stop codons that occurred at locations where the consensus alignment contained a gap or where the majority of genes within the alignment were not present (the very beginnings and ends of the alignment) were not counted. Pairwise alignments of radiolarian mitochondrion-encoded proteins were performed using MUSCLE and inspected manually.

Phylogenetic analysis of Foraminifera and Radiolaria. Twelve mitochondrial protein-coding genes (*cox1*, *cox2*, *cox3*, *cob*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad7*, *atp1*, *atp6*, and *atp9*) were aligned with MAFFT (v7.450) (133). The split *nad1* gene was excluded from phylogenetic analyses. Aligned protein sequences per gene were manually trimmed to the same length, and stop codons were removed. All analyzed genes were manually concatenated. Gaps in the alignment were manually removed, resulting in an alignment of 2,137 amino acids. A phylogenetic tree was calculated using the IQ-TREE web server (134) with the JTT+F+G4 model and 1,000 iterations of Ultrafast Bootstrap (134). We visualized the resulting tree using FigTree (v1.4.4) (https://github.com/rambaut/figtree/).

Data availability. Raw reads are available in the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA743004. The full mitochondrial genomes of *Calcarina hispida* and *Neorotalia gaimardi* have been deposited in NCBI GenBank (accession numbers OP965949 and OP965950). The radiolarian genome assemblies, multiple sequence alignments, and predicted mitochondrial gene sequences have been deposited at Figshare (10.6084/m9.figshare.16734961). Assemblies can be searched using BLAST on a SequenceServer (135) at https://evocellbio.com/SAGdb/macher_et_al/.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 0.5 MB.

FIG S2, TIF file, 0.4 MB.

FIG S3, TIF file, 0.4 MB.

FIG S4, TIF file, 1.7 MB.

FIG S5, PDF file, 0.9 MB.

TABLE S1, XLSX file, 0.1 MB.

TABLE S2, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grant DBI-2119963 (J.G.W.).

We are grateful to T. Coots for assistance with computational analyses.

REFERENCES

1. Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. Philos Trans R Soc Lond B Biol Sci 370:20140330. https://doi.org/10.1098/rstb.2014.0330.
2. Archibald JM. 2015. Endosymbiosis and eukaryotic cell evolution. Curr Biol 25:R911–R921. https://doi.org/10.1016/j.cub.2015.07.055.
3. Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. Curr Biol 27:R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015.
4. Kishino H, Miyata T, Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol 31:151–160. https://doi.org/10.1007/BF02109483.
5. Hjort K, Goldberg AV, Tsaousis AD, Hirt RP, Embley TM. 2010. Diversity and reductive evolution of mitochondria among microbial eukaryotes. Philos Trans R Soc Lond B Biol Sci 365:713–727. https://doi.org/10.1098/rstb.2009.0224.
6. Khachane AN, Timmis KN, Martins dos Santos VAP. 2007. Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. Mol Biol Evol 24:449–456. https://doi.org/10.1093/molbev/msl174.
7. Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol 118:9–17. https://doi.org/10.1104/pp.118.1.9.
8. Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: anything goes. Trends Genet 19:709–716. https://doi.org/10.1016/j.tig.2003.10.012.
9. Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol Evol 5:418–438. https://doi.org/10.1093/gbe/evt008.
10. Wideman JG, Monier A, Rodríguez-Martínez R, Leonard G, Cook E, Poirier C, Maguire F, Milner DS, Irwin NAT, Moore K, Santoro AE, Keeling PJ, Worden AZ, Richards TA. 2020. Unexpected mitochondrial genome diversity revealed by targeted single-cell genomics of heterotrophic flagellated protists. Nat Microbiol 5:154–165. https://doi.org/10.1038/s41564-019-0605-4.
11. Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. Annu Rev Genet 38:477–524. https://doi.org/10.1146/annurev.genet.37.110801.142526.
12. Gray MW, Burger G, Derelle R, Klimeš V, Leger MM, Sarrasin M, Vlček Č, Roger AJ, Eliáš M, Lang BF. 2020. The draft nuclear genome sequence and predicted mitochondrial proteome of Andalucia godoyi, a protist with the most gene-rich and bacteria-like mitochondrial genome. BMC Biol 18:22. https://doi.org/10.1186/s12915-020-0741-6.
13. Stairs CW, Leger MM, Roger AJ. 2015. Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Philos Trans R Soc Lond B Biol Sci 370:20140326. https://doi.org/10.1098/rstb.2014.0326.
14. Yahalomi D, Atkinson SD, Neuhof M, Sally Chang E, Philippe H, Cartwright P, Bartholomew JL, Huchon D. 2020. A cnidarian parasite of salmon (Myxozoa: Henneguya) lacks a mitochondrial genome. Proc Natl Acad Sci U S A 117:5358–5363. https://doi.org/10.1073/pnas.1909907117.
15. Maguire F, Richards TA. 2014. Organelle evolution: a mosaic of "mitochondrial" functions. Curr Biol 24:R518–R520. https://doi.org/10.1016/j.cub.2014.03.075.
16. Andersson JO, Sjögren AM, Davis LAM, Embley TM, Roger AJ. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Curr Biol 13:94–104. https://doi.org/10.1016/s0960-9822(03)00003-4.

17. Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123–135. https://doi.org/10.1038/nrg1271.
18. Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. Nature 524:427–432. https://doi.org/10.1038/nature14963.
19. Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol 29:380–395. https://doi.org/10.1016/s1055-7903(03)00194-5.
20. von Heijne G. 1986. Why mitochondria need a genome. FEBS Lett 198:1–4. https://doi.org/10.1016/0014-5793(86)81172-3.
21. Björkholm P, Ernst AM, Hagström E, Andersson SGE. 2017. Why mitochondria need a genome revisited. FEBS Lett 591:65–75. https://doi.org/10.1002/1873-3468.12510.
22. Allen JF. 2017. The CoRR hypothesis for genes in organelles. J Theor Biol 434:50–57. https://doi.org/10.1016/j.jtbi.2017.04.008.
23. Lane N, Martin W. 2010. The energetics of genome complexity. Nature 467:929–934. https://doi.org/10.1038/nature09486.
24. Giannakis K, Arrowsmith SJ, Richards L, Gasparini S, Chustecki JM, Røyrvik EC, Johnston IG. 2022. Evolutionary inference across eukaryotes identifies universal features shaping organelle gene retention. Cell Syst 13:874–884.e5. https://doi.org/10.1016/j.cels.2022.08.007.
25. Johnston IG, Williams BP. 2016. Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. Cell Syst 2:101–111. https://doi.org/10.1016/j.cels.2016.01.013.
26. Song J, Pfanner N, Becker T. 2018. Assembling the mitochondrial ATP synthase. Proc Natl Acad Sci U S A 115:2850–2852. https://doi.org/10.1073/pnas.1801697115.
27. Gray MW. 2015. Mosaic nature of the mitochondrial proteome: implications for the origin and evolution of mitochondria. Proc Natl Acad Sci U S A 112:10133–10138. https://doi.org/10.1073/pnas.1421379112.
28. Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu R-Y, van der Giezen M, Tielens AGM, Martin WF. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol Mol Biol Rev 76:444–495. https://doi.org/10.1128/MMBR.05024-11.
29. Boore JL. 1999. Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780. https://doi.org/10.1093/nar/27.8.1767.
30. Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Rohwer FL, Mylnikov AP, Keeling PJ. 2017. A new lineage of eukaryotes illuminates early mitochondrial genome reduction. Curr Biol 27:3717–3724.E5. https://doi.org/10.1016/j.cub.2017.10.051.
31. Paquin B, Laforest M-J, Forget L, Roewer I, Wang Z, Longcore J, Lang BF. 1997. The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. Curr Genet 31:380–395. https://doi.org/10.1007/s002940050220.
32. Wang X, Lavrov DV. 2007. Mitochondrial genome of the homoscleromorph Oscarella carmela (Porifera, Demospongiae) reveals unexpected complexity in the common ancestor of sponges and other animals. Mol Biol Evol 24:363–373. https://doi.org/10.1093/molbev/msl167.
33. Waller RF, Jackson CJ. 2009. Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. Bioessays 31:237–245. https://doi.org/10.1002/bies.200800164.
34. Takeuchi F, Sekizuka T, Ogasawara Y, Yokoyama H, Kamikawa R, Inagaki Y, Nozaki T, Sugita-Konishi Y, Ohnishi T, Kuroda M. 2015. The mitochondrial genomes of a myxozoan genus Kudoa are extremely divergent in metazoa. PLoS One 10:e0132030. https://doi.org/10.1371/journal.pone.0132030.
35. Mathur V, Wakeman KC, Keeling PJ. 2021. Parallel functional reduction in the mitochondria of apicomplexan parasites. Curr Biol 31:2920–2928.E4. https://doi.org/10.1016/j.cub.2021.04.028.
36. Kayal E, Smith DR. 2021. Is the dinoflagellate Amoebophrya really missing an mtDNA? Mol Biol Evol 38:2493–2496. https://doi.org/10.1093/molbev/msab041.
37. John U, Lu Y, Wohlrab S, Groth M, Janouškovec J, Kohli GS, Mark FC, Bickmeyer U, Farhat S, Felder M, Frickenhaus S, Guillou L, Keeling PJ, Moustafa A, Porcel BM, Valentin K, Glöckner G. 2019. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. Sci Adv 5:eaav1110. https://doi.org/10.1126/sciadv.aav1110.
38. Roy J, Faktorová D, Lukes J, Burger G. 2007. Unusual mitochondrial genome structures throughout the Euglenozoa. Protist 158:385–396. https://doi.org/10.1016/j.protis.2007.03.002.
39. Dobáková E, Flegontov P, Skalický T, Lukeš J. 2015. Unexpectedly streamlined mitochondrial genome of the euglenozoan Euglena gracilis. Genome Biol Evol 7:3358–3367. https://doi.org/10.1093/gbe/evv229.
40. Popescu CE, Lee RW. 2007. Mitochondrial genome sequence evolution in Chlamydomonas. Genetics 175:819–826. https://doi.org/10.1534/genetics.106.063156.
41. Denovan-Wright EM, Nedelcu AM, Lee RW. 1998. Complete sequence of the mitochondrial DNA of Chlamydomonas eugametos. Plant Mol Biol 36:285–295. https://doi.org/10.1023/a:1005995718091.
42. Jain K, Krause K, Grewe F, Nelson GF, Weber APM, Christensen AC, Mower JP. 2014. Extreme features of the Galdieria sulphuraria organellar genomes: a consequence of polyextremophily? Genome Biol Evol 7:367–380. https://doi.org/10.1093/gbe/evu290.
43. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, Čepička I, Chistyakova L, Del Campo J, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karnkowska A, Karpov S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Le Gall L, Lynn DH, Mann DG, Massana R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella G, Youssef N, Zlatogursky V, Zhang Q. 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol 66:4–119. https://doi.org/10.1111/jeu.12691.
44. Xu Z, Wang M, Wu W, Li Y, Liu Q, Han Y, Jiang Y, Shao H, McMinn A, Liu H. 2018. Vertical distribution of microbial eukaryotes from surface to the hadal zone of the Mariana Trench. Front Microbiol 9:2023. https://doi.org/10.3389/fmicb.2018.02023.
45. Biard T, Krause JW, Stukel MR, Ohman MD. 2018. The significance of giant phaeodarians (rhizaria) to biogenic silica export in the California current ecosystem. Global Biogeochem Cycles 32:987–1004. https://doi.org/10.1029/2018GB005877.
46. Biard T, Ohman MD. 2020. Vertical niche definition of test-bearing protists (Rhizaria) into the twilight zone revealed by in situ imaging. Limnol Oceanogr 65:2583–2602. https://doi.org/10.1002/lno.11472.
47. Hemleben C, Spindler M, Roger AO. 2012. Modern planktonic Foraminifera. Springer Science & Business Media, Berlin, Germany.
48. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum J, Coelho LP, Espinoza JCI, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, de Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G, Tara Oceans coordinators. 2016. Plankton networks driving carbon export in the oligotrophic ocean. Nature 532:465–470. https://doi.org/10.1038/nature16942.
49. Gutierrez-Rodriguez A, Stukel MR, Lopes Dos Santos A, Biard T, Scharek R, Vaulot D, Landry MR, Not F. 2019. High contribution of Rhizaria (Radiolaria) to vertical export in the California Current ecosystem revealed by DNA metabarcoding. ISME J 13:964–976. https://doi.org/10.1038/s41396-018-0322-7.
50. Stukel MR, Biard T, Krause J, Ohman MD. 2018. Large Phaeodaria in the twilight zone: their role in the carbon cycle. Limnol Oceanogr 63:2579–2594. https://doi.org/10.1002/lno.10961.
51. Orsi WD, Morard R, Vuillemin A, Eitel M, Wörheide G, Milucka J, Kucera M. 2020. Anaerobic metabolism of Foraminifera thriving below the seafloor. ISME J 14:2580–2594. https://doi.org/10.1038/s41396-020-0708-1.
52. Langer MR. 2008. Assessing the contribution of foraminiferan protists to global ocean carbonate production. J Eukaryot Microbiol 55:163–169. https://doi.org/10.1111/j.1550-7408.2008.00321.x.
53. Moodley L, Boschker HTS, Middelburg JJ, Pel R, Herman PMJ, de Deckere E, Heip CHR. 2000. Ecological significance of benthic foraminifera: 13C labelling experiments. Mar Ecol Prog Ser 202:289–295. https://doi.org/10.3354/meps202289.
54. Monferrer NL, Boltovskoy D, Tréguer P, Sandin MM, Not F, Leynaert A. 2020. Estimating biogenic silica production of Rhizaria in the global ocean. Global Biogeochem Cycles 34:e2019GB006286. https://doi.org/10.1029/2019GB006286.
55. Glöckner G, Hülsmann N, Schleicher M, Noegel AA, Eichinger L, Gallinger C, Pawlowski J, Sierra R, Euteneuer U, Pillet L, Moustafa A, Platzer M, Groth M, Szafranski K, Schliwa M. 2014. The genome of the foraminiferan Reticulomyxa filosa. Curr Biol 24:11–18. https://doi.org/10.1016/j.cub.2013.11.027.
56. Habura A, Hou Y, Reilly AA, Bowser SS. 2011. High-throughput sequencing of Astrammina rara: sampling the giant genome of a giant foraminiferan protist. BMC Genomics 12:169. https://doi.org/10.1186/1471-2164-12-169.
57. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A,

Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, et al. 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.
58. Sierra R, Matz MV, Aglyamova G, Pillet L, Decelle J, Not F, de Vargas C, Pawlowski J. 2013. Deep relationships of Rhizaria revealed by phylogenomics: a farewell to Haeckel's Radiolaria. Mol Phylogenet Evol 67:53–59. https://doi.org/10.1016/j.ympev.2012.12.011.
59. Krabberød AK, Orr RJS, Bråte J, Kristensen T, Bjørklund KR, Shalchian-Tabrizi K. 2017. Single cell transcriptomics, mega-phylogeny, and the genetic basis of morphological innovations in rhizaria. Mol Biol Evol 34:1557–1573. https://doi.org/10.1093/molbev/msx075.
60. Liu Z, Mesrop LY, Hu SK, Caron DA. 2019. Transcriptome of Thalassicolla nucleata holobiont reveals details of a radiolarian symbiotic relationship. Front Mar Sci 6. https://doi.org/10.3389/fmars.2019.00284.
61. Wideman JG, Lax G, Leonard G, Milner DS, Rodríguez-Martínez R, Simpson AGB, Richards TA. 2019. A single-cell genome reveals diplonemid-like ancestry of kinetoplastid mitochondrial gene structure. Philos Trans R Soc Lond B Biol Sci 374:20190100. https://doi.org/10.1098/rstb.2019.0100.
62. Záhonová K, Lax G, Sinha SD, Leonard G, Richards TA, Lukeš J, Wideman JG. 2021. Single-cell genomics unveils a canonical origin of the diverse mitochondrial genomes of euglenozoans. BMC Biol 19:103. https://doi.org/10.1186/s12915-021-01035-y.
63. Sen Gupta BK. 2003. Modern Foraminifera. Springer, Berlin, Germany.
64. Suzuki N, Not F. 2015. Biology and ecology of Radiolaria, p 179–222. *In* Ohtsuka S, Toshinobu S, Horiguchi T, Suzuki N, Not F (ed), Marine protists. Springer, Berlin, Germany.
65. Weber AA-T, Pawlowski J. 2014. Wide occurrence of SSU rDNA intragenomic polymorphism in foraminifera and its implications for molecular species identification. Protist 165:645–661. https://doi.org/10.1016/j.protis.2014.07.006.
66. Decelle J, Romac S, Sasaki E, Not F, Mahé F. 2014. Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing. PLoS One 9:e104297. https://doi.org/10.1371/journal.pone.0104297.
67. Woehle C, Roy A-S, Glock N, Wein T, Weissenbach J, Rosenstiel P, Hiebenthal C, Michels J, Schönfeld J, Dagan T. 2018. A novel eukaryotic denitrification pathway in Foraminifera. Curr Biol 28:2536–2543.E5. https://doi.org/10.1016/j.cub.2018.06.027.
68. Gast RJ, Caron DA. 1996. Molecular phylogeny of symbiotic dinoflagellates from planktonic foraminifera and radiolaria. Mol Biol Evol 13:1192–1197. https://doi.org/10.1093/oxfordjournals.molbev.a025684.
69. Gast RJ, Sanders RW, Caron DA. 2009. Ecological strategies of protists and their symbiotic relationships with prokaryotic microbes. Trends Microbiol 17:563–569. https://doi.org/10.1016/j.tim.2009.09.001.
70. Prazeres M, Renema W. 2019. Evolutionary significance of the microbial assemblages of large benthic Foraminifera. Biol Rev Camb Philos Soc 94:828–848. https://doi.org/10.1111/brv.12482.
71. Girard EB, Langerak A, Jompa J, Wangensteen OS, Macher J-N, Renema W. 2022. Mitochondrial cytochrome oxidase subunit 1: a promising molecular marker for species identification in Foraminifera. Front Mar Sci 9. https://doi.org/10.3389/fmars.2022.809659.
72. Richter DJ, Berney C, Strassert JFH, Poh Y-P, Herman EK, Muñoz-Gómez SA, Wideman JG, Burki F, de Vargas C. 2022. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotes. Peer Community J 2:e56. https://doi.org/10.24072/pcjournal.173.
73. Hammond M, Zoltner M, Garrigan J, Butterfield E, Varga V, Lukeš J, Field MC. 2021. The distinctive flagellar proteome of Euglena gracilis illuminates the complexities of protistan flagella adaptation. New Phytol 232:1323–1336. https://doi.org/10.1111/nph.17638.
74. Gutiérrez P, Bulman S, Alzate J, Ortíz MC, Marín M. 2016. Mitochondrial genome sequence of the potato powdery scab pathogen Spongospora subterranea. Mitochondrial DNA A DNA Mapp Seq Anal 27:58–59. https://doi.org/10.3109/19401736.2013.873898.
75. Gawryluk RMR, Kamikawa R, Stairs CW, Silberman JD, Brown MW, Roger AJ. 2016. The earliest stages of mitochondrial adaptation to low oxygen revealed in a novel rhizarian. Curr Biol 26:2729–2738. https://doi.org/10.1016/j.cub.2016.08.025.
76. Daval S, Belcour A, Gazengel K, Legrand L, Gouzy J, Cottret L, Lebreton L, Aigu Y, Mougel C, Manzanares-Dauleux MJ. 2019. Computational analysis of the Plasmodiophora brassicae genome: mitochondrial sequence description and metabolic pathway database design. Genomics 111:1629–1640. https://doi.org/10.1016/j.ygeno.2018.11.013.
77. Tanifuji G, Archibald JM, Hashimoto T. 2016. Comparative genomics of mitochondria in chlorarachniophyte algae: endosymbiotic gene transfer and organellar genome dynamics. Sci Rep 6:21016. https://doi.org/10.1038/srep21016.
78. Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR. 1992. The mitochondrial genomes of two nematodes, Caenorhabditis elegans and Ascaris suum. Genetics 130:471–498. https://doi.org/10.1093/genetics/130.3.471.
79. Stjelja S, Fogelqvist J, Tellgren-Roth C, Dixelius C. 2019. The architecture of the Plasmodiophora brassicae nuclear and mitochondrial genomes. Sci Rep 9:15753. https://doi.org/10.1038/s41598-019-52274-7.
80. Holzmann M, Pawlowski J. 2017. An updated classification of rotaliid foraminifera based on ribosomal DNA phylogeny. Mar Micropaleontol 132:18–34. https://doi.org/10.1016/j.marmicro.2017.04.002.
81. Holzmann M, Hohenegger J, Hallock P, Piller WE, Pawlowski J. 2001. Molecular phylogeny of large miliolid foraminifera (Soritacea Ehrenberg 1839). Mar Micropaleontol 43:57–74. https://doi.org/10.1016/S0377-8398(01)00021-4.
82. Pawlowski J, Holzmann M, Tyszka J. 2013. New supraordinal classification of Foraminifera: molecules meet morphology. Mar Micropaleontol 100:1–10. https://doi.org/10.1016/j.marmicro.2013.04.002.
83. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
84. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res 46:W200–W204. https://doi.org/10.1093/nar/gky448.
85. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW, Knox EB, Palmer JD. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science 342:1468–1473. https://doi.org/10.1126/science.1246275.
86. Zou H, Jakovlić I, Chen R, Zhang D, Zhang J, Li W-X, Wang G-T. 2017. The complete mitochondrial genome of parasitic nematode Camallanus cotti: extreme discontinuity in the rate of mitogenomic architecture evolution within the Chromadorea class. BMC Genomics 18:840. https://doi.org/10.1186/s12864-017-4237-x.
87. Ling J, O'Donoghue P, Söll D. 2015. Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. Nat Rev Microbiol 13:707–721. https://doi.org/10.1038/nrmicro3568.
88. Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2:49–58. https://doi.org/10.1038/35047500.
89. McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10:13–26. https://doi.org/10.1038/nrmicro2670.
90. Knight RD, Landweber LF, Yarus M. 2001. How mitochondria redefine the code. J Mol Evol 53:299–313. https://doi.org/10.1007/s002390010220.
91. Keeling PJ. 2016. Genomics: evolution of the genetic code. Curr Biol 26:R851–R853. https://doi.org/10.1016/j.cub.2016.08.005.
92. Heaphy SM, Mariotti M, Gladyshev VN, Atkins JF, Baranov PV. 2016. Novel ciliate genetic code variants including the reassignment of all three stop codons to sense codons in Condylostoma magnum. Mol Biol Evol 33:2885–2889. https://doi.org/10.1093/molbev/msw166.
93. Záhonová K, Kostygov AY, Ševčíková T, Yurchenko V, Eliáš M. 2016. An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. Curr Biol 26:2364–2369. https://doi.org/10.1016/j.cub.2016.06.064.
94. Kachale A, Pavlíková Z, Nenarokova A, Roithová A, Durante IM, Miletínová P, Záhonová K, Nenarokov S, Votýpka J, Horáková E, Ross RL, Yurchenko V, Beznosková P, Paris Z, Valášek LS, Lukeš J. 2023. Short tRNA anticodon stem and mutant eRF1 allow stop codon reassignment. Nature 613:751–758. https://doi.org/10.1038/s41586-022-05584-2.
95. Hammond MJ, Nenarokova A, Butenko A, Zoltner M, Dobáková EL, Field MC, Lukeš J. 2020. A uniquely complex mitochondrial proteome from Euglena gracilis. Mol Biol Evol 37:2173–2191. https://doi.org/10.1093/molbev/msaa061.
96. Hosoya K-I, Amagai A, Chida J, Maeda Y. 2003. Unique behavior and function of the mitochondrial ribosomal protein S4 (RPS4) in early Dictyostelium development. Zoolog Sci 20:1455–1465. https://doi.org/10.2108/zsj.20.1455.
97. Korovesi AG, Ntertilis M, Kouvelis VN. 2018. Mt-rps3 is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. Mol Phylogenet Evol 127:74–86. https://doi.org/10.1016/j.ympev.2018.04.037.

98. Feagin JE, Harrell MI, Lee JC, Coe KJ, Sands BH, Cannone JJ, Tami G, Schnare MN, Gutell RR. 2012. The fragmented mitochondrial ribosomal RNAs of Plasmodium falciparum. PLoS One 7:e38320. https://doi.org/10.1371/journal.pone.0038320.
99. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2.
100. Strauss M, Hofhaus G, Schröder RR, Kühlbrandt W. 2008. Dimer ribbons of ATP synthase shape the inner mitochondrial membrane. EMBO J 27:1154–1160. https://doi.org/10.1038/emboj.2008.35.
101. Mühleip A, Kock Flygaard R, Ovciarikova J, Lacombe A, Fernandes P, Sheiner L, Amunts A. 2021. ATP synthase hexamer assemblies shape cristae of Toxoplasma mitochondria. Nat Commun 12:120. https://doi.org/10.1038/s41467-020-20381-z.
102. Davies KM, Anselmi C, Wittig I, Faraldo-Gómez JD, Kühlbrandt W. 2012. Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae. Proc Natl Acad Sci U S A 109:13602–13607. https://doi.org/10.1073/pnas.1204593109.
103. Paumard P, Vaillier J, Coulary B, Schaeffer J, Soubannier V, Mueller DM, Brèthes D, di Rago J-P, Velours J. 2002. The ATP synthase is involved in generating mitochondrial cristae morphology. EMBO J 21:221–230. https://doi.org/10.1093/emboj/21.3.221.
104. LeKieffre C, Bernhard JM, Mabilleau G, Filipsson HL, Meibom A, Geslin E. 2018. An overview of cellular ultrastructure in benthic foraminifera: new observations of rotalid species in the context of existing literature. Mar Micropaleontol 138:12–32. https://doi.org/10.1016/j.marmicro.2017.10.005.
105. Yuasa T, Takahashi O. 2016. Light and electron microscopic observations of the reproductive swarmer cells of nassellarian and spumellarian polycystines (Radiolaria). Eur J Protistol 54:19–32. https://doi.org/10.1016/j.ejop.2016.02.007.
106. Anderson OR. 1978. Fine structure of a symbiont-bearing colonial radiolarian, Collosphaera globularis, and ^{14}C isotopic evidence for assimilation of organic substances from its zooxanthellae. J Ultrastruct Res 62:181–189. https://doi.org/10.1016/s0022-5320(78)90031-x.
107. Sinha SD, Wideman JG. 2022. The persistent homology of mitochondrial ATP synthases. bioRxiv. https://doi.org/10.1101/2022.09.13.506888.
108. Burger G, Yan Y, Javadi P, Lang BF. 2009. Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. Trends Genet 25:381–386. https://doi.org/10.1016/j.tig.2009.07.003.
109. Jackson CJ, Waller RF. 2013. A widespread and unusual RNA trans-splicing type in dinoflagellate mitochondria. PLoS One 8:e56777. https://doi.org/10.1371/journal.pone.0056777.
110. Janouskovec J, Sobotka R, Lai D-H, Flegontov P, Koník P, Komenda J, Ali S, Prásil O, Pain A, Oborník M, Lukes J, Keeling PJ. 2013. Split photosystem protein, linear-mapping topology, and growth of structural complexity in the plastid genome of Chromera velia. Mol Biol Evol 30:2447–2462. https://doi.org/10.1093/molbev/mst144.
111. Macher J-N, Wideman JG, Girard EB, Langerak A, Duijm E, Jompa J, Sadekov A, Vos R, Wissels R, Renema W. 2021. First report of mitochondrial COI in foraminifera and implications for DNA barcoding. Sci Rep 11:22165. https://doi.org/10.1038/s41598-021-01589-5.
112. Kaur B, Záhonová K, Valach M, Faktorová D, Prokopchuk G, Burger G, Lukeš J. 2020. Gene fragmentation and RNA editing without borders: eccentric mitochondrial genomes of diplonemids. Nucleic Acids Res 48:2694–2708. https://doi.org/10.1093/nar/gkz1215.
113. Kelly S. 2021. The economics of organellar gene loss and endosymbiotic gene transfer. Genome Biol 22:345. https://doi.org/10.1186/s13059-021-02567-w.
114. Berg OG, Kurland CG. 2000. Why mitochondrial genes are most often found in nuclei. Mol Biol Evol 17:951–961. https://doi.org/10.1093/oxfordjournals.molbev.a026376.
115. Tobiasson V, Berzina I, Amunts A. 2022. Structure of a mitochondrial ribosome with fragmented rRNA in complex with membrane-targeting elements. Nat Commun 13:6132. https://doi.org/10.1038/s41467-022-33582-5.
116. Waltz F, Soufari H, Bochler A, Giegé P, Hashem Y. 2020. Cryo-EM structure of the RNA-rich plant mitochondrial ribosome. Nat Plants 6:377–383. https://doi.org/10.1038/s41477-020-0631-5.
117. Soufari H, Waltz F, Parrot C, Durrieu-Gaillard S, Bochler A, Kuhn L, Sissler M, Hashem Y. 2020. Structure of the mature kinetoplastids mitoribosome and insights into its large subunit biogenesis. Proc Natl Acad Sci U S A 117:29851–29861. https://doi.org/10.1073/pnas.2011301117.
118. Strassert JFH, Irisarri I, Williams TA, Burki F. 2021. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nat Commun 12:1879. https://doi.org/10.1038/s41467-021-22044-z.
119. Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354.
120. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.
121. Macher J-N, Bloska DM, Holzmann M, Girard EB, Pawlowski J, Renema W. 2022. Mitochondrial cytochrome c oxidase subunit I (COI) metabarcoding of Foraminifera communities using taxon-specific primers. PeerJ 10:e13952. https://doi.org/10.7717/peerj.13952.
122. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Res 44:D67–D72. https://doi.org/10.1093/nar/gkv1276.
123. Sonnhammer EL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6:175–182.
124. Sonnhammer EL, Eddy SR, Durbin R. 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins 28:405–420. https://doi.org/10.1002/(SICI)1097-0134(199707)28:3%3C405::AID-PROT10%3E3.0.CO;2-L.
125. UniProt Consortium. 2015. UniProt: a hub for protein information. Nucleic Acids Res 43:D204–D212. https://doi.org/10.1093/nar/gku989.
126. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31:365–370. https://doi.org/10.1093/nar/gkg095.
127. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, et al. 2018. Ensembl 2018. Nucleic Acids Res 46:D754–D761. https://doi.org/10.1093/nar/gkx1098.
128. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
129. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol 1962:227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.
130. Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Res 6:1287. https://doi.org/10.12688/f1000research.12232.1.
131. Wheeler TJ, Eddy SR. 2013. nhmmer: DNA homology search with profile HMMs. Bioinformatics 29:2487–2489. https://doi.org/10.1093/bioinformatics/btt403.
132. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113.
133. Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436.
134. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44:W232–W235. https://doi.org/10.1093/nar/gkw256.
135. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, Chowdhary H, Pieniak I, Maynard LJ, Gibbins MA, Moon H, Davis-Richardson A, Uludag M, Watson-Haigh NS, Challis R, Nakamura H, Favreau E, Gómez EA, Pluskal T, Leonard G, Rumpf W, Wurm Y. 2019. Sequenceserver: a modern graphical user interface for custom BLAST databases. Mol Biol Evol 36:2922–2924. https://doi.org/10.1093/molbev/msz185.