Journal of Eukaryotic Microbiology

ORIGINAL ARTICLE OPEN ACCESS

Complete Mitochondrial Genomes of Ancyromonads Provide Clues for the Gene Content and Genome Structures of Ancestral Mitochondria

Ryo Harada[1] | Takashi Shiratori[2] | Akinori Yabuki[3,4] | Yuji Inagaki[5] | Andrew J. Roger[1] | Ryoma Kamikawa[6]

[1]Department of Biochemistry and Molecular Biology, Institute for Comparative Genomics, Dalhousie University, Halifax, Nova Scotia, Canada | [2]Institute of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan | [3]Research Institute for Global Change (RIGC), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka, Japan | [4]Advanced Institute for Marine Ecosystem Change (WPI-AIMEC), Yokohama, Japan | [5]Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan | [6]Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Correspondence: Ryo Harada (rharada@dal.ca) | Ryoma Kamikawa (kamikawa.ryoma.7v@kyoto-u.ac.jp)

Received: 2 April 2025 | **Revised:** 1 May 2025 | **Accepted:** 5 May 2025

Funding: This work was in part supported by the Japan Society for Promotion of Sciences (JSPS) projects, 18J02091 (awarded to T.S.), 22KJ0401 (awarded to R.H.), 24K21929 (awarded to R.K.), 23K27226, and BPI06050 (awarded to Y.I.), a Discovery grant RGPIN-2022-05430 awarded to A.J.R. by the Natural Sciences and Engineering Research Council of Canada. R.H. was also supported by JSPS Overseas Research Fellowships. R.K. was also supported by a grant from Institute for Fermentation, Osaka (G-2024-1-011). A.Y. was also supported by the World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan.

Keywords: AlphaFold | *Ancyromonas* | Diphoda | *Fabomonas* | Foldseek | inverted repeats | LECA | *Nutomonas* | Opimoda | *rps5*

ABSTRACT

Mitochondria of eukaryotic cells are direct descendants of an endosymbiotic bacterium related to Alphaproteobacteria. These organelles retain their own genomes, which are highly reduced and divergent when compared to those of their bacterial relatives. To better understand the trajectory of mitochondrial genome evolution from the last eukaryotic common ancestor (LECA) to extant species, mitochondrial genome sequences from phylogenetically diverse lineages of eukaryotes—particularly protists—are essential. For this reason, we focused on the mitochondrial genomes of Ancyromonadida, an independent and understudied protist lineage in the eukaryote tree of life. Here we report the mitochondrial genomes from three Ancyromonadida: *Ancyromonas sigmoides*, *Nutomonas longa*, and *Fabomonas tropica*. Our analyses reveal that these mitochondrial genomes are circularly mapping molecules with inverted repeats that carry genes. This inverted repeat structure has been observed in other mitochondrial genomes but is patchily distributed over the tree of eukaryotes. Ancyromonad mitochondrial genomes possess several protein-coding genes, which have not been detected from any other mitochondrial genomes of eukaryotes sequenced to date, thereby extending the known mitochondrial gene repertoire of ancestral eukaryotes, including LECA. These findings significantly expand our understanding of mitochondrial genome diversity across eukaryotes, shedding light on the early phases of mitochondrial genome evolution.

1 | Introduction

Mitochondria are double-membrane bounded organelles that are best known as the ATP-producing "powerhouses" of extant eukaryotic cells. They are the direct descendants of endosymbiotic bacteria related to Alphaproteobacteria that were integrated into host cells of archaeal origin during eukaryogenesis (Richards et al. 2024; Roger et al. 2017; Vosseberg et al. 2024). Except for several anaerobic groups (Abrahamsen et al. 2004; Leger et al. 2017; Williams et al. 2002), the mitochondria of most eukaryotic species retain their own genomes derived from their endosymbiotic bacterial ancestor. However,

the vast majority of the genes of the original endosymbiont were either lost if they were no longer necessary for the host–endosymbiont partnership or transferred to the nuclear genome during genetic and functional integration between the endosymbiont and the host. The end result is mitochondrial genomes of extant species that are streamlined in size and gene content compared to their bacterial relatives (Ettema and Andersson 2009). Importantly, even though some of this genomic streamlining occurred during eukaryogenesis, it has continued to occur after the diversification of the major extant eukaryotic lineages from the last eukaryotic common ancestor (LECA), yielding a considerable variation in gene content even among mitochondrial genomes of relatively closely related protist species (Eglit et al. 2024; Kamikawa et al. 2016; Moreira et al. 2024; Yazaki et al. 2022).

The most gene-rich mitochondrial genomes currently known are those of the discobid group Jakobida with 65–100 kb in length, which carry 61–66 protein genes and 30–34 RNA genes (Burger et al. 2013). The most gene-rich mitochondrial genome outside Jakobida so far was found in the CRuMs, representatives of which have 53–63 kb-long mitochondrial genomes carrying 50–62 protein genes and ~30 RNA-encoding genes (Kamikawa et al. 2016; Moreira et al. 2024). At the other end of the spectrum, the smallest mitochondrial genomes are found in dinoflagellates, apicomplexan parasites, and their photosynthetic close relatives, that retain only two or three protein genes and fragmented rRNA genes (Oborník and Lukeš 2015). The mitochondrial genomes of apicomplexan parasites are also the smallest in size (6–7 kb; Feagin 2000). In addition to their coding capacity, mitochondrial genomes show substantial diversity in structure among species (Burger et al. 2003; Gray et al. 1998). The vast majority of aerobic eukaryotes possess circularly- or linearly mapping mitochondrial genomes. However, the mitochondrial genomes of dinoflagellates are composed of a variety of linear DNA fragments generated by recombination (Kamikawa et al. 2007; Norman and Gray 2001; Slamovits et al. 2007). The mitochondria of some Euglenozoa also have unusual features, including gene fragmentation where each gene fragment is located within small circular DNAs (Dobáková et al. 2015; Kaur et al. 2020; Spencer and Gray 2011; Verner et al. 2015). Although somewhat unusual, inverted repeat sequences that encode genes (hereafter simply referred to as inverted repeats or IRs) are present in the mitochondrial genomes of a variety of distantly related protists including *Malawimonas jakobiformis* (NC_002553), *Proteromonas lacertae* (Pérez-Brocal et al. 2010), *Palpitomonas bilix* (Nishimura et al. 2016), and *Acavomonas peruviana* (Janouškovec et al. 2013; Tikhonenkov et al. 2014). To accurately infer the gene content and structure of the mitochondrial genome of LECA and the subsequent evolutionary "paths" from LECA to various extant species, we need broad sampling of mitochondrial genomes of phylogenetically diverse eukaryotes, particularly those from deep-branching lineages.

One important understudied "deep branching" protist lineage is the Ancyromonadida, a group of flagellates that are found in diverse aquatic and soil environments (Heiss et al. 2010; Yubuki et al. 2023). Members of Ancyromonadida are heterotrophic protists, characterized by two flagella, round-to-bean-shaped cells and a very small cell size of approximately $4 \pm 0.5\,\mu m$ in length. Electron microscopic analyses showed that these cells have a pellicle under the cell membrane and extrusomes (Heiss et al. 2010; Yubuki et al. 2023). Recent phylogenomic analyses have not settled on a robust placement of the Ancyromonadida in the eukaryote tree. Recent nuclear gene based analyses position Ancyromonadida as a lineage closely related to Malawimonadida and CRuMs (Eglit et al. 2024; Harada et al. 2024; Torruella et al. 2025; Yazaki et al. 2025), which, together with Amorphea comprised of Opisthokonta, Apusomonadida, Breviatea, and Amoebozoa, are proposed to form the huge assemblage called Opimoda (Derelle et al. 2015). The phylogenetic group including Ancyromonadida in Opimoda is referred to as "Opimoda+" (Williamson et al. 2025). The latest mitochondrial protein-based phylogenomic analyses placed the eukaryotic root position between Opimoda+ and Diphoda+, the latter of which comprises Diaphoretickes and Discoba (Williamson et al. 2025).

Recently, Gastineau et al. (2023) reported two incomplete linear mitochondrial genome fragments of Ancyromonadida: one from *Ancyromonas sigmoides* strain 1C2 and another from metagenomic samples. Given the incomplete mitochondrial genome fragments of *A. sigmoides* and the lack of any mitochondrial genome sequence in other ancyromonad lineages, the diversity and evolution of the mitochondrial genome structures and gene contents in Ancyromonadida remain poorly understood. To expand our understanding of mitochondrial genome diversity in eukaryotes, we have characterized the complete mitochondrial genome sequences of representatives of three ancyromonad species: *A. sigmoides*, *Nutomonas longa*, and two strains of *Fabomonas tropica*. Our analyses reveal that these Ancyromonadida mitochondrial genomes possess IRs and some ribosomal protein genes that have not previously been found in mitochondrial genomes. These findings update our understanding of diversity in mitochondrial genome structure and gene repertoire across the eukaryote tree, affording deeper insights into the evolutionary history of these organelles.

2 | Materials and Methods

2.1 | Culturing, Sequencing, and Assembling

Cells of *A. sigmoides* B-70, *N. longa* NCFW, and *F. tropica* NYK3C were provided by Dr. Aaron A. Heiss (Glücksman et al. 2013; Heiss et al. 2010). The protist DNA extracted with the Plant DNA Extraction Kit (Jena Biosciences) was subjected to the library preparation with the Illumina TruSeq Nano DNA Library Prep Kit for 350 bp inserts according to the manufacturer's instructions and sequenced with the HiSeq 2500 (Illumina), resulting in 46.0 million, 44.1 million, and 44.5 million paired-end reads for *A. sigmoides* B-70, *N. longa* NCFW, and *F. tropica* NYK3C, respectively. Adapter trimming was conducted using Cutadapt version 1.1 (Martin 2011) as the default setting. Quality filtering was conducted with Trimmomatic v0.32 (Bolger et al. 2014), setting the SLIDINGWINDOW option to 20:20 and keeping the minimum length of 50 nt. The resulting reads were assembled using Velvet version 1.2.08 (Zerbino and Birney 2008), setting input reads, expected coverage, and hash length as the short paired-end reads, "auto," and 35–55, respectively. By BLAST-based homology searches (Camacho et al. 2009) using *Reclinomonas americana* mitochondrial proteins (NC_001823.1), mitochondrial genome-derived contigs

TABLE 1 | Characteristics of mitochondrial genomes from Ancyromonadida and phylogenetically related candidates.

	Accession	Genome size (bp)	GC (%)	ORFs	Coding ratio (%)	Inverted repeats (bp)	Coverage of inverted repeats/ single-copy region
Ancyromonas sigmoides B-70	LC835033	48,960	34.18	55	90.31	7108	149.77×/75.21×
Nutomonas longa NCFW	LC835035	54,750	34.45	54	88.96	7052	20.46×/10.06×
Fabomonas tropica NYK3C	LC835034	79,657	25.28	68	91.22	28,695	449.11×/259.30×
Fabomonas tropica SRT902	LC860085	80,057	25.39	68	91.46	28,898	681.40×/381.73×
Malawimonas jakobiformis	AF295546	47,328	26.11	49	91.13	9100	—
Malawimonas californiana	KP165387	36,792	26.79	39	89.72	608	—
Diphylleia rotans	NC_029886	62,563	34.40	58	68.80	0	—
Mantamonas sphyraenae	LC842150	53,045	26.54	63	90.98	0	—

were detected as fragments of mitochondrial DNAs. Taking overlapping sequences at the 5′ and 3′ ends and read coverage into consideration, the contigs were manually assembled into single circularly mapping DNAs. Quality-filtered reads were then mapped onto the DNAs by HISAT2 (Kim et al. 2019) using default settings to confirm that the manual assemblies were correct (Figure S1).

We also sequenced the mitochondrial genome of the CRuM species *Mantamonas sphyraenae* SRT306. However, because the same organisms' mtDNA was recently published (Moreira et al. 2024), we do not discuss it in detail. The circularly mapping, 53,045 bp mitochondrial genome of *M. sphyraenae* was sequenced and assembled using the same methods as those described herein for *A. sigmoides* B-70, *N. longa* NCFW, and *F. tropica* NYK3C.

Cells of *F. tropica* SRT902 were cultivated as previously described (Harada et al. 2024). The whole DNA extracted by the phenol–chloroform method was subjected to the library preparation with the VAHTS Universal DNA Library Prep Kit for Illumina V3 ND607 and sequenced with the HiSeq X (Illumina), resulting in 112 million paired-end reads. Library construction and sequencing were conducted at a biotech company (AZENTA Japan Corp., Tokyo, Japan). The raw reads of *F. tropica* SRT902 were trimmed with fastp v0.19.7 (Chen et al. 2018) with the "-q 20 -u 80" options and then assembled with SPAdes v3.15.3 (Nurk et al. 2017) using the "-meta -k 13,21,29,37,45,53 ,61,69,77,85,93,101,109,117,125" options. Mitochondrial contigs were detected by BLAST-based homology search using *Jakoba libera* mitochondrial proteins (NC_021127.1). The contigs were manually assembled into single mapping DNA. Quality-filtered reads were then mapped onto the assembly by HISAT2 (Kim et al. 2019) as the default settings to confirm that the manual assembly was correct.

2.2 | Gene Annotation

The mitochondrial genome sequences of four Ancyromonadida strains and *M. sphyraenae* were annotated using the MFannot (Lang et al. 2023) web server with the standard mitochondrial genetic code (NCBI's translation Table 1). Infernal v1.1.5 (Nawrocki and Eddy 2013) and covariance model of mt-tmRNA (RF02544) were used to search for transfer-messenger RNA (tmRNA). Functional annotation of each open reading frame (ORF) assigned by MFannot was confirmed by PSI-BLAST against clustered_nr as of September 2024 (Altschul et al. 1997). If annotations of genes by MFannot and PSI-BLAST were inconsistent with each other, they were further investigated and their annotation was treated as provisional. ORFs not confidently annotated in MFannot and PSI-BLAST with *E*-values below 1e−4 were subjected to the following, more detailed annotation. For functionally unknown ORFs, we performed conserved domain searches using InterProScan v5.69 (Jones et al. 2014). The 3D structure of each ORF was predicted by the AlphaFold3 web server (https://alphafoldserver.com) v2024.08.19 (Abramson et al. 2024) and the predicted tertiary structures were used as queries in a similarity search using the Foldseek web server (https://search.foldseek.com) as of September 2024 (van Kempen et al. 2024) with an *E*-value cutoff of 1e−4.

Furthermore, an all-against-all BLAST analysis was performed among the four Ancyromonadida mitochondrial genomes. We focused on *atp4*, *sdh3*, and *tatC* in *A. sigmoides*, which were absent in the other three Ancyromonadida mitochondrial genomes. The tertiary structures of these three proteins were predicted using AlphaFold3 and compared with the structures of functionally unknown ORFs in the other three mitochondrial genomes using TM-align (Zhang and Skolnick 2005) with a TM-score cutoff of 0.5 to search for potential homologues. We also performed additional homology searches using HHpred (against

the Pfam-A database v37) and Phyre 2.2 with a threshold of 95% of HHpred probability and Phyre 2.2 confidence scores (Powell et al. 2025; Zimmermann et al. 2018). Additionally, we assessed transmembrane (TM) domains in functionally unknown ORFs using TMHMM2.0 (Krogh et al. 2001), DeepTMHMM1.0 (Hallgren et al. 2022), and Phobius1.01 (Käll et al. 2004). Genome maps were drawn by OGDRAW (Greiner et al. 2019) and edited manually. Synteny in the mitochondrial genomes of Ancyromonadida was evaluated manually by comparing gene contents and orders across the three species.

With the above-mentioned tools, we reinvestigated the gene contents of two previously reported Ancyromonadida mitochondrial genomes, Ancyromonadida sp. SAT37 (PALT01000012.1) and *A. sigmoides* 1C2 annotated in Gastineau et al. (2023). We found one gene of Ancyromonadida sp. SAT37 was annotated as *rps10* in Gastineau et al. (2023), which was not detected in the mitochondrial genomes of either *A. sigmoides* 1C2 (Gastineau et al. 2023) or B-70 (this study). No tools we employed identified it as *rps10*; therefore, we treat Ancyromonadida sp. SAT37 as a mitochondrial genome lacking *rps10*.

3 | Results

3.1 | Mitochondrial Genome Structure

The complete mitochondrial genome of *A. sigmoides* B-70 was determined to be a circularly mapping molecule, with a genome size of 48,960 bp (Figure 1; Table 1). This genome exhibited a G + C content of 34%, and coding regions accounted for 90% of the genome sequence. The mitochondrial small subunit rRNA (SSU rRNA) gene in *A. sigmoides* B-70 showed 99.8% identity to the previously sequenced strain 1C2. Despite this high degree of similarity, the mitochondrial genome of B-70 was found to be 7071 bp larger than the published mitochondrial sequence from strain 1C2, likely due to the incompleteness of the latter genome (Gastineau et al. 2023). Notably, the B-70 mitochondrial genome contains IRs spanning 7108 bp, which nearly match the size discrepancy observed in strain 1C2, supporting the notion that the latter may be missing this sequence. In fact, the gene content and synteny of the mitochondrial genomes of B-70 and 1C2 were identical except for the presence/absence of the IR regions. In the raw read mapping analysis, if the IRs are indeed repeat regions, the read coverage on the IRs would be approximately two times higher than that of single-copy regions. Consistent with this prediction, the coverage of the IR region and the single-copy regions were 149.77× and 75.21×, respectively, in B-70 (Table 1).

The *A. sigmoides* B-70 mitochondrial genome carries two rRNA genes and 24 tRNA genes: the annotation program MFannot did not detect a conserved 5S rRNA gene. The mitochondrial genome encodes 38 proteins with known functions when counting duplicated genes as one (Figure 1), and no introns were detected in any of the detected ORFs. The mitochondrial gene content of *A. sigmoides* B-70 is identical to that of the strain 1C2. Besides possessing genes encoding eight small subunit ribosomal proteins (rps) and seven large subunit ribosomal proteins (rpl), it retains genes encoding subunits of complex I (*nad1-4, 4L, and 5-11*), complex II (*sdh2* and *sdh3*), complex III (*cob*), complex IV (*cox1-3*), ATP synthase complex (*atp4, 6, 8, and 9*), and the twin arginine translocator (*tatC*). The IRs encode 9 tRNA (*trnA, I, L, P, Q, S, T, V, and W*), 7 rps (*rps2, 3, 12, 13, 14, 16, and 19*), and 5 rpl (*rpl1, 5, 6, 14, and 16*) genes.

The mitochondrial genome of *N. longa* NCFW was also a circularly mapping molecule, with a genome size of 54,750 bp (Figure 1; Table 1). The genome retains two rRNA genes and 24 tRNA genes. The *N. longa* mitochondrial genome encodes 37 proteins with known functions. Introns were not detected in any ORFs. Besides the 17 ribosomal proteins comprised of 11 rps and 6 rpl, the *N. longa* mitochondrial genome has genes encoding subunits of complex I (*nad1-4, 4L, and 5-11*), complex II (*sdh2*), complex III (*cob*), complex IV (*cox1-3*), and ATP synthase (*atp6, 8, and 9*), but lacks *sdh3*, *atp4*, and *tatC* that are present in the mitochondrial genome of *A. sigmoides*. Seven tRNA (*trnA, P, Q, S, T, V, and W*), 7 rps (*rps2, 3, 7, 12, 13, 14, and 19*), and 4 rpl (*rpl5, 6, 14, and 16*) genes were found to be present within the IRs of *N. longa*.

The sequenced mitochondrial genomes of two strains of *F. tropica*, NYK3C and SRT902, were determined as circularly mapping molecules with genome sizes of 79,657 bp and 80,057 bp, respectively (Figure 1; Table 1). Mitochondrial SSU rRNA genes of the two *F. tropica* strains shared 98% identity with each other. Both genomes exhibited 25% G + C content and 91% coding capacity. The two *F. tropica* strains exhibit larger IRs, measuring 28,695 bp and 28,898 bp, than those in the other Ancyromonadida species sequenced in this study (Table 1). The gene content and order of the two *F. tropica* strains are identical to each other. They encode two rRNA genes and 24 tRNA genes. The *F. tropica* mitochondrial genomes encode 37 proteins. Introns were not detected in any ORFs. Besides the ribosomal proteins, the protein gene contents are identical among strains of *N. longa* and *F. tropica*. Their gene contents differed only in presence or absence of *rps5*, *rps10*, *rpl5*, *rpl10*, and *tatC*. In the IRs, 14 tRNA (*trnC, D, E, F, K, L, M, N, P, Q, S, and Y*) including two copies each of *trnM* and *trnS*, 2 rps (*rps8 and 10*), 1 rpl (*rpl2*), 2 rRNA (*rnl and rns*), 3 nad (*nad4L, 6, and 8*), *cob*, 3 cox (*cox1-3*), 3 atp (*atp6, 8, and 9*), and *tatC* genes, as well as 7 unidentified ORFs (orf126, 180, 182, 306, 390, 785, and 1047) are present.

The mitochondrial gene synteny in Ancyromonadida is conserved. Specifically, there are conserved operon-like gene clusters with adjacent genes encoding functionally related subunits (Figures 1 and 2A,B). For example, we found that *nad11-nad1-nad5-nad4-nad2*, *nad10-nad9*, *rps13-rps19-rps3-rpl16-rpl14*, *cox1-cox3*, and *nad7-sdh2* are conserved among all the genomes analyzed in this study. Collectively, the four genomes sequenced here represent the first complete mitochondrial genomes from Ancyromonadida.

3.2 | Functionally Uncharacterized ORFs

Out of all the functionally annotated genes with MFannot and/or PSI-BLAST with significant statistical reliability (*E*-value < 1e−4), only *rps5* of *N. longa* and *rpl10* of *F. tropica* strains were annotated by more detailed analyses using InterProScan. In addition to this prediction, the amino acid sequences of the ORFs that were not confidently annotated in MFannot and PSI-BLAST were subjected to an AlphaFold/FoldSeek analysis, which

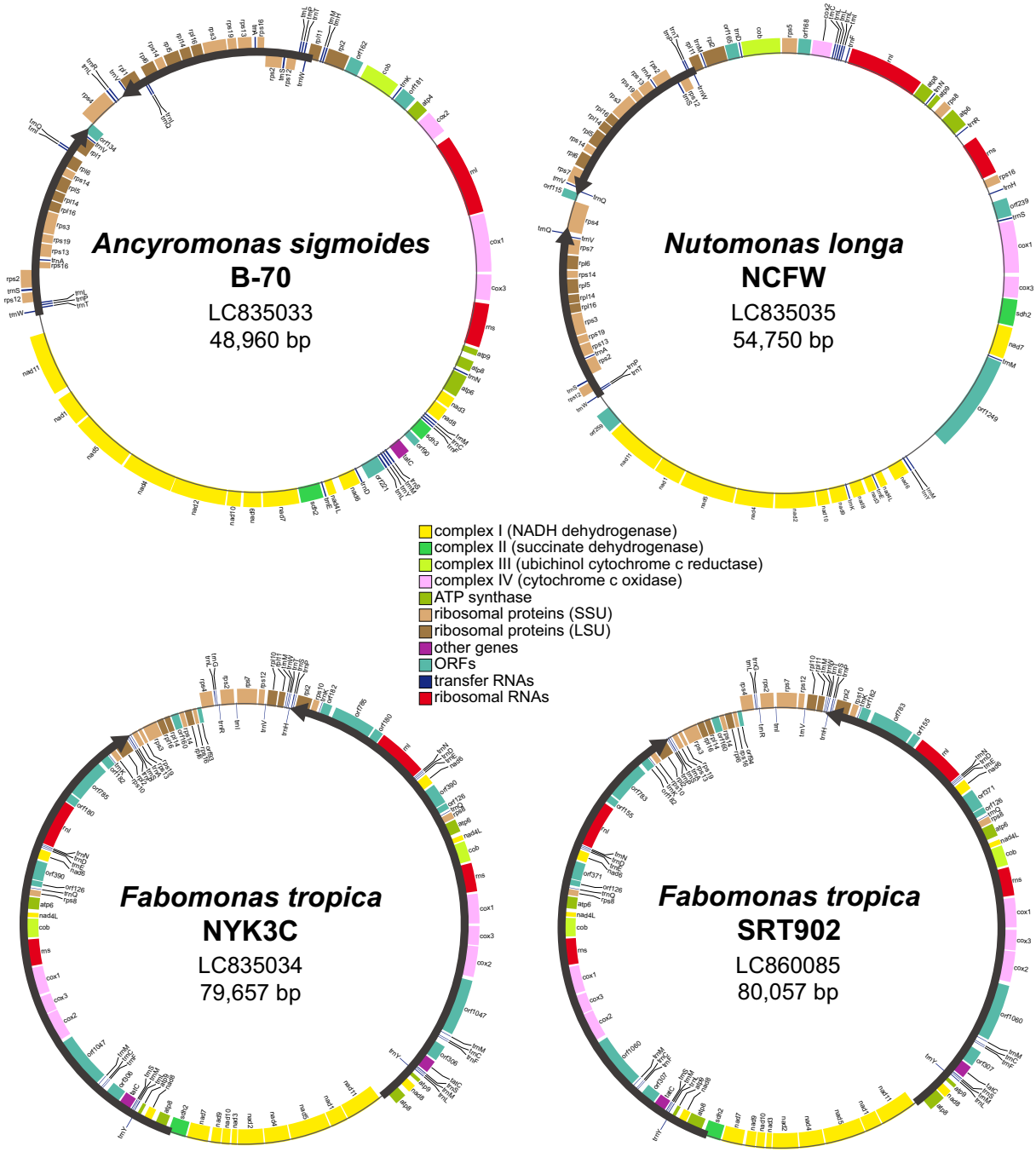

FIGURE 1 | Complete maps of four Ancyromonadida species mitochondrial genomes. Coding regions are colored based on their function (see color key in figure). Black bold arrows indicate inverted repeats. Each map was drawn by OGDRAW and edited manually.

confirmed that the tertiary structures predicted by AlphaFold3 for the proteins encoded by *N. longa* orf198 (Nuto01.10) and *F. tropica* NYK3C/SRT902 orf188/orf189 (Fabo01.15/SRT902.15) hit the tertiary structures of Rps5 and Rpl10 with *E*-values of 3.8e−5 and 6.1e−6/6.2e−5, respectively (Table S1). Additionally, HHpred and Phyre 2.2 searches detected significant homology of *F. tropica* NYK3C/SRT902 orf247 (Fabo01.47/SRT902.47) to *tatC* (Table S1). We compared the tertiary structures of *N. longa* mitochondrial Rps5 and *F. tropica* mitochondrial Rpl10 with those of experimentally determined *Homo sapiens* nuclear-encoded mitochondrial homologues and *Escherichia coli* homologues (Figure 3), confirming their similarities. *N. longa* Rps5 matched the Human homologue with root mean squared deviations (RMSDs) of 3.3 Å and 1.6 Å in the N- and C-terminal regions, respectively; *F. tropica* Rpl10 matched the *E. coli* homologue with an RMSD of 2.58 Å. To date, this is the only known case of Rps5 being encoded on a mitochondrial genome. The amino acid sequence of *N. longa* Rps5 was aligned with nuclear-encoded mitochondrial Rps5 of phylogenetically diverse eukaryotes and bacterial Rps5 (Figure S2); some conserved amino acid residues are present in the mitochondrion-encoded Rps5 of *N. longa*. Similarly, the structure-based annotation also detected candidate homologues of several other ORFs, although they were not statistically significant (highlighted in cyan in Table S1). For example, it seems probable that *A. sigmoides* orf181 (Ancy01.4) and *F. tropica* orf182 (Fabo01.11) both encode

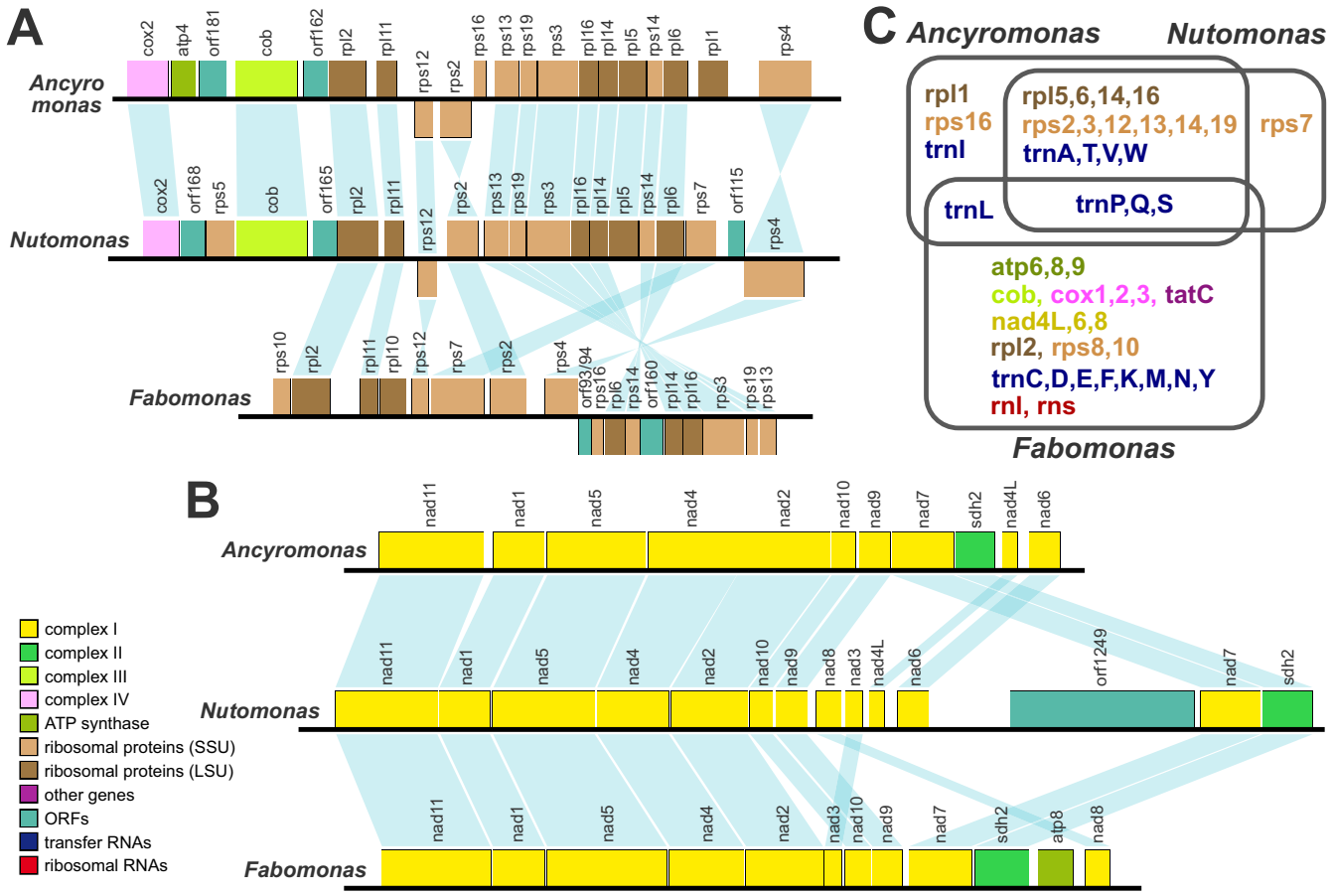

FIGURE 2 | Conservation of synteny and gene repertoire of inverted repeats. Conservation of synteny around ribosomal genes (A) and complex I genes (B) is highlighted in a blue band. tRNA genes are omitted in the synteny analysis. (C) Gene repertoire of inverted repeats.

Rps5. Furthermore, *N. longa* orf168 (Nuto01.9) might encode Atp4. Using amino acid sequences of *N. longa* Rps5 and *F. tropica* Rpl10 as queries, homology-based surveys also did not detect any homologues of these proteins in the mitochondrial genomes of three Ancyromonadida species.

MFannot annotated the C-terminal region of the 900 amino acid long *N. longa* Orf1249 as Sdh3, a subunit of mitochondrial complex II which has a typical length of 100–200 amino acids. However, our subsequent analyses with PSI-BLAST, InterProScan, AlphaFold/Foldseek, HHpred, and Phyre 2.2 did not find significant similarity to Sdh3, leaving Orf1249 functionally unidentified. The predicted 3D structure of Orf1249 showed alpha helices aligned along the same axial direction (Figure 4). Three *in silico* tools, DeepTMHMM, TMHMM2, and Phobius, also predicted some of the helices as TM domains. Since Sdh3 is known to be a membrane protein containing a TM domain as part of Complex II, it is possible, but not likely that the initial MFannot assignment of Orf1249 as Sdh3 could be correct. The tertiary structures of *A. sigmoides* Sdh3 and *N. longa* Orf1249 were aligned using TM-align, but the RMSD was 5.22Å, indicating no significant structural similarity. Another possibility is that because Orf1249 is much longer than typical Sdh3, it could correspond to Sdh3 fused with another protein or other proteins. Gene fusions are observed in the mitochondrial genomes of Opimoda+, for example, *cox1/cox3* in *Gefionella okellyi* (a member of Malawimonadida) and *cox1/cox2* in *Acanthamoeba castellanii* (a member of Amoebozoa) (Lonergan and Gray 1996; Valach et al. 2014). However, due to lack of definitive sequence and structural similarity evidence, precise annotation of orf1249 remains difficult.

In any case, genes encoding functionally annotated proteins larger than 1000 amino acids—for example, *rpoB* and *rpoC*, which are conserved in Jakobida species—are rare in mitochondrial genomes. It is interesting that the mitochondrial genomes of Ancyromonadida, which have relatively high coding density, retain large ORFs of unknown function. At present, potential functions of the other uncharacterized ancyromonad mitochondrial ORFs, including the large ORFs in *F. tropica*, remain unknown.

4 | Discussion

We determined the complete genome sequences of mitochondria in four ancyromonad strains from three species, *A. sigmoides*, *N. longa*, and *F. tropica*. As in some other mitochondrial genomes found in diverse eukaryotes, the mitochondrial genomes of ancyromonads are A+T rich, circularly mapping molecules that are densely packed by coding regions. Almost all the genes found in the mitochondrial genomes of Ancyromonadida are those that have previously been reported. However, we report an *rps5* gene that has never been found in any previously

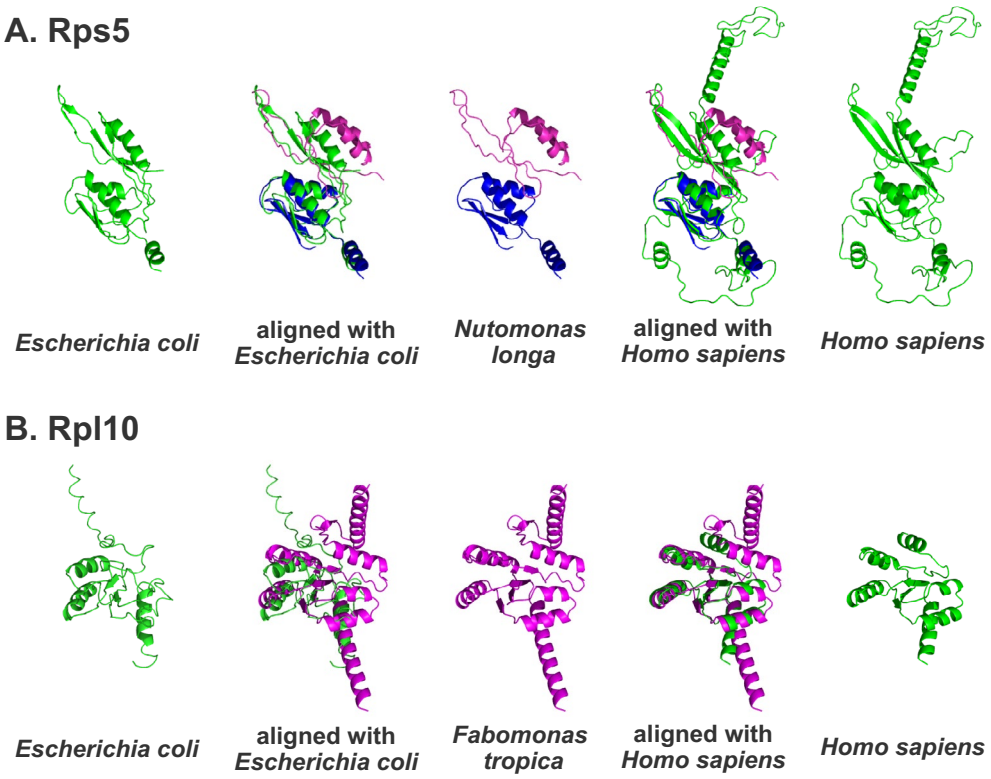

FIGURE 3 | Structural comparison of Rps5 and Rpl10. (A) The N-terminal (1–77 amino acid residues) and C-terminal (119–198 amino acid residues) regions of *Nutomonas longa* Rps5 are shown in magenta and blue, respectively. The left and right in green show the experimentally determined structure of *Escherichia coli* Rps5 (chain E in pdb_00006nqb) and *Homo sapiens* nuclear-encoded mitochondrial Rps5 (126–430 amino acid residues of chain AD in pdb_00003j9m), respectively. (B) *Fabomonas tropica* NYK3C Rpl10 is shown in magenta. The left and right in green show the experimentally determined structure of *E. coli* Rpl10 (chain I in pdb_00008upo) and *H. sapiens* nuclear-encoded mitochondrial Rpl10 (77–197 amino acid residues of chain I in pdb_00003j9m), respectively. Structures were aligned by TM-align.

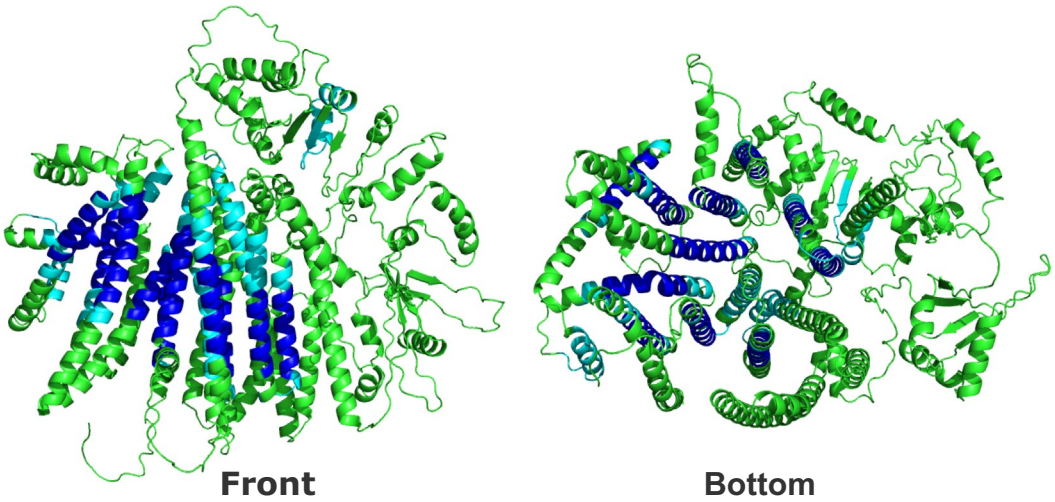

FIGURE 4 | Predicted 3D structure and TM helices of *Nutomonas longa* Orf1249. The structure of *N. longa* Orf1249 predicted by AlphaFold3 is shown in the front view and bottom view on the left and right, respectively. Transmembrane (TM) helices predicted by three *in silico* tools are shown in blue. TM helices predicted by at least one of three tools are shown in cyan.

published mitochondrial genomes, as well as an *rpl1* gene that has never been found in any Opimoda+ mitochondrial genomes other than Ancyromonadida. In addition to the functionally annotated genes, the Ancyromonadida mitochondrial genomes carry many ORFs that do not appear to have significant similarity to functionally annotated proteins in any databases. Unfortunately, even predicted tertiary structure-based analyses for all the ORFs provide no clue to estimate the functions of their encoded proteins, with some exceptions (Table S1). Nevertheless, it is intriguing that 13 out of the 29 ORFs were predicted to have one or more TM domains (Table S1), suggesting that they might encode membrane-localized proteins such as

transporters of metabolites. There are three possibilities for each of the functionally unannotated ORFs: (i) they might be non-coding sequences, (ii) they might be newly produced genes that possess no homologues in other organisms, or (iii) they might be divergent homologues of functionally known proteins. The high coding density of these genomes suggests that (i) is unlikely. In any case, with the expansion of available mitochondrial genome sequence data and the improvement of homology detection methods, annotation of these ORFs might be possible in the future. Alternatively, if these are truly "new genes," biochemical experiments will be needed to determine their functions.

4.1 | Ancestral Genome Structures and the Evolution of Mitochondria

All of the four complete mitochondrial genomes of Ancyromonadida have IRs. In *A. sigmoides* and *N. longa*, these IRs are 7108 and 7052 bp, respectively, sharing similar gene repertoires and orders (Figures 1 arrows and 2C). On the other hand, the two *F. tropica* strains, NYK3C and SRT902, have much larger IRs that are 28,695 and 28,898 bp long, respectively, with gene contents and orders identical to each other. However, almost all the genes on the IRs are not shared between *Ancyromonas*/*Nutomonas* and *Fabomonas*, and only three tRNA genes are shared among them (Figure 2C). A recent phylogenomic analysis suggested that *A. sigmoides* and *N. longa* are more closely related to each other than to *F. tropica* (Brown et al. 2018; Gastineau et al. 2023; Harada et al. 2024; Torruella et al. 2025; Williamson et al. 2025; Yazaki et al. 2025), which is consistent with the patterns of similarities/differences in the IRs of these taxa. The close relationship and the similarity of IRs between *A. sigmoides* and *N. longa* suggest the last common ancestor of the two genera possessed IRs with a similar gene repertoire and gene order. Nevertheless, the stark differences in size and gene arrangement between IRs of *A. sigmoides*/*N. longa* and *F. tropica* strains make it difficult to infer ancestral features of IRs in ancyromonads. As mentioned above, the IRs might have been present in the mitochondrial genome of the last common ancestor of Ancyromonadida followed by vertical inheritance to the extant species, and replacements of genes in the repeats. Alternatively, it is possible that IRs might have emerged independently after divergence in the *A. sigmoides*/*N. longa* lineage and in the *F. tropica* lineage.

More broadly, IRs in mitochondrial genome have been reported in diverse eukaryotes (Čechová et al. 2018; Wynn and Christensen 2019), including fungi, plants, metazoa, and deep-branching lineages such as Malawimonadida, Nebulidia in Provora, *Meteora*, a heterolobosean strain BB2, *Microheliella*, *Palpitomonas*, *P. lacertae*, and *A. peruviana* (Christinaki et al. 2023; Čutová et al. 2020; Eglit et al. 2024; Janouškovec et al. 2013, 2017; Jiang et al. 2023; Nishimura et al. 2016; Pérez-Brocal et al. 2010; Stern and Palmer 1984; Tikhonenkov et al. 2014, 2022; Valach et al. 2014; Yang et al. 2017; Yazaki et al. 2022). In some species of fungi, lineage specific gain of IRs are known to result from plasmid integration or recombination, resulting in structural diversity in mitochondrial genomes (Bágeľová Poláková et al. 2021; Christinaki et al. 2023; Clark-Walker et al. 1981; Ferandon et al. 2008; Férandon et al. 2013; Gerhold et al. 2010; Locker et al. 1974; Salavirta et al. 2014). It is also known that plastid genomes in some photosynthetic algae have lost either copy of inverted repeats containing rRNA operons multiple times in evolution (Choi et al. 2019; Kamikawa et al. 2015; Kayama et al. 2020; Krämer et al. 2024; Matsuo et al. 2022; Palmer and Thompson 1982), demonstrating that IRs can be lost. Currently, except for the fungi, it is unclear what mechanisms are underpinning the conversions between IR-containing and IR-lacking mitochondrial genomes and how "easy" these transitions are. Furthermore, it remains uncertain whether the mechanisms known in the fungi are conserved in all the above-mentioned eukaryotic lineages including Ancyromonadida. It would be noteworthy that no plasmid DNAs and no mitochondrial proteins involved in plasmid integration and recombination, known to be keys to gain/loss of IRs in fungi (Christinaki et al. 2023; Férandon et al. 2013), have been reported in Malawimonadida, Nebulidia in Provora, *Meteora*, a heterolobosean strain BB2, *Microheliella*, *Palpitomonas*, *P. lacertae*, and *A. peruviana*.

Along similar lines as the scenarios discussed above for Ancyromonadida, IRs might have emerged early in eukaryotic evolution followed by vertical inheritance, gene gains or losses within IRs, and multiple independent complete losses of IRs. Alternatively, IRs may have been acquired independently in multiple lineages, as seen in fungi. Since these are not mutually exclusive alternatives, a blend of both scenarios might have led to the punctate distribution of IRs observed across the eukaryotic tree. Since few species and lineages closely related to the above-mentioned deep-branching protists that retain IRs have been sequenced, the extent to which IR-containing genome structures are conserved among closely related species remains uncertain, except for the Ancyromonadida. The detection of IRs in all the Ancyromonadida mitochondrial genomes completely sequenced in this study is intriguing in light of the not-yet fully resolved phylogenetic position of Ancyromonadida (Brown et al. 2018; Eglit et al. 2024; Harada et al. 2024; Torruella et al. 2025; Williamson et al. 2025; Yazaki et al. 2025); Representatives of Malawimonadida, one of the lineages potentially related to Ancyromonadida, also possess IRs in all their characterized mitochondrial genomes, although there are no genes in IRs shared by the two lineages. Nevertheless, if the sister group relationship between Ancyromonadida and Malawimonadida is true, the mitochondrial IRs found in Ancyromonadida and Malawimonadida might stem from vertical inheritance from their common ancestor followed by modification of their gene contents.

4.2 | Evolutionary History of Mitochondrial Gene Repertoires

Our analyses of *A. sigmoides*, *N. longa*, and *F. tropica* have revealed variation in the mitochondrial gene repertoires of Ancyromonadida (Figure 5). Whereas confidently detectable *rps5* is exclusively present in the mitochondrial genome of *N. longa*, *rps10* and *rpl10* are exclusively detected in those of *F. tropica*. Similarly, *rpl1*, *atp4*, and *sdh3* are unique to *Ancyromonas* spp. In contrast, *rps7* and *rps8* are uniquely undetectable in the *Ancyromonas* mitochondrial genomes, whereas the lack of detectable *rpl5* and *tatC* is unique to the *F. tropica* and *N. longa* mitochondrial genomes, respectively. However, it is likely that

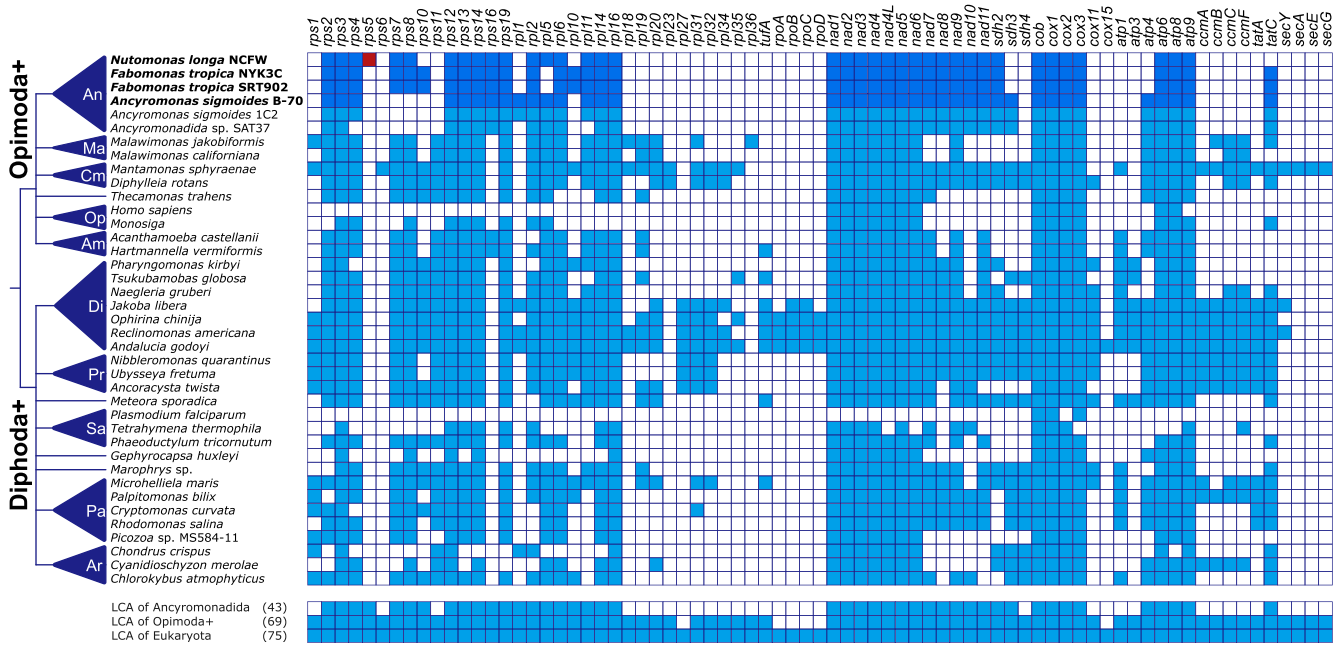

FIGURE 5 | Repertoire of protein-coding genes in mitochondrial genomes across the eukaryote tree of life. Presence and absence of corresponding genes of various eukaryotes is shown by closed and open boxes, respectively. Genes of the mitochondrial genomes newly sequenced in this study are highlighted in dark blue and the presence of *rps5* is indicated with the red box. The figure is updated based on several previous studies (Eglit et al. 2024; Kamikawa et al. 2016; Moreira et al. 2024; Yazaki et al. 2022). Phylogenetic relationships of eukaryotes are based on Williamson et al. (2025). The predicted gene contents of the last common ancestor (LCA) of Opimoda+, LCA of Ancyromonadida, and LCA of Eukaryota are shown at bottom, and numbers in parentheses are the total numbers of genes in each ancestral mitochondrial genome. Abbreviations of phyla: Am, Amoebozoa; An, Ancyromonadida; Ar, Archaeplastida; Cm, CRuMs; Di, Discoba; Ma, Malawimonadida; Op, Opisthokonta; Pa, Pancryptista; Pr, Provora; Sa, SAR.

some known proteins may have remained undetected in our annotation due to their high amino acid sequence divergence from their homologues in databases. Among the proteins detected using BLAST only in a subset of Ancyromonadida species, the average identity to the top BLAST hits against the clustered_nr database is 32%, when *A. sigmoides* 1C2 is removed from the database. This level of sequence similarity falls within the "twilight zone" of homology detection (20%–35%), where accurate annotation becomes particularly challenging (Rost 1999). If we ignore the possibility of gene gain in mitochondrial genome evolution, then the last common ancestor of Ancyromonadida must have possessed at least all of the 43 protein-coding genes detected in the four mitochondrial genomes analyzed in this work. We believe it is reasonable to ignore gene gain by horizontal gene transfer in mitochondrial genomes, as only a few cases of such gain have been reported so far (He et al. 2016; Nishimura et al. 2020). During the diversification of the various lineages within Ancyromonadida, losses of genes from mitochondrial genomes or losses of detectable sequence homology must have happened differentially among species. Based on the closer relationship of *Ancyromonas* with *Nutomonas* than with *Fabomonas* (Brown et al. 2018; Gastineau et al. 2023; Harada et al. 2024; Torruella et al. 2025; Williamson et al. 2025; Yazaki et al. 2025), *rps5*, *rpl1*, *atp4*, and *sdh3* could have been lost from mitochondrial genomes (or diverged beyond recognition) multiple times independently in the evolution of Ancyromonadida. This scenario seems plausible given that these genes are known to have been transferred to the nuclear genomes multiple times in eukaryotic evolution (Butenko et al. 2024).

More broadly, phylogenetic relationships within Opimoda+ including the phylogenetic position of Ancyromonadida are not conclusive (Brown et al. 2018; Eglit et al. 2024; Harada et al. 2024; Torruella et al. 2025; Williamson et al. 2025; Yazaki et al. 2025). Therefore, it is difficult to trace the reductive evolution of mitochondrial genome repertoires lineage by lineage within Opimoda+. Nevertheless, it is likely that the last common ancestor of Opimoda+ retained at least all of the 69 protein-coding genes detected in the mitochondrial genomes of Amorphea, CRuMs, Malawimonadida, and Ancyromonadida. If so, then six ribosomal protein-coding genes, *rps1*, *rps11*, *rpl18-20*, *rpl31*, and three cytochrome c maturase system I protein-coding genes, *ccmB*, *C*, and *F*, have very likely been lost within the Opimoda+ along the lineage leading to the Ancyromonadida; these genes appear to have been lost multiple times in mitochondrial genome evolution (Butenko et al. 2024). Based on the same reasoning, the mitochondrial genome in LECA encoded at least 75 proteins, which are most of the known protein-coding genes found in mitochondrial genomes. According to the recent phylogenomic analyses that estimated the root of the eukaryotes on the branch splitting Opimoda+ and Diphoda+ (Williamson et al. 2025), an ancestral mitochondrial genome of Opimoda+ had lost just four RNA polymerase subunit genes (*rpoA-D*), *rpl27* and *cox15*, on the descendant branch from LECA prior to the diversification of Opimoda+. On the other side of the root, five ribosomal protein-coding genes, *rps5*, *rps6*, *rps16*, *rpl23*, and *rpl36*, and genes for Sec translocator system except for *secY* were lost along the lineage from LECA to the

common ancestor of extant Diphoda+. Since the original alphaproteobacteria-related endosymbionts that gave rise to mitochondria likely possessed hundreds or thousands of protein-coding genes (Ettema and Andersson 2009), it seems likely that many or most of these genes were lost from the mitochondrial genome prior to the LECA (Richards et al. 2024; Roger et al. 2017).

5 | Conclusions

In this study, we sequenced and analyzed the complete mitochondrial genomes of three Ancyromonadida species, revealing structural and genetic features that expand our understanding of mitochondrial genome evolution. We found that these genomes are circularly mapping molecules and contain IRs, which are patchily distributed across eukaryotes. Additionally, we identified *rps5*, a gene never detected in any mitochondrial genome. These findings provide new insights into the ancestral mitochondrial genome structure and gene repertoire. A more detailed picture of IR evolution and gene loss events of mitochondrial genomes along eukaryote phylogeny will be afforded by an improved understanding of the deep relationships within Opimoda+ as well as within Diphoda+ and on the availability of mitochondrial genomes from more diverse, deeply diverging groups. Thus, we suggest that these inferences be revisited again whenever new mitochondrial genomes from diverse understudied protists are determined.

Acknowledgments

We would like to thank Dr. Aaron A. Heiss at Kyungpook National University for providing the cell cultures. This work was in part supported by the Japan Society for Promotion of Sciences (JSPS) projects, 18J02091 (awarded to T.S.), 22KJ0401 (awarded to R.H.), 24K21929 (awarded to R.K.), 23K27226, and BPI06050 (awarded to Y.I.), a Discovery grant RGPIN-2022-05430 awarded to A.J.R. by the Natural Sciences and Engineering Research Council of Canada. R.H. was also supported by JSPS Overseas Research Fellowships. R.K. was also supported by a grant from Institute for Fermentation, Osaka (G-2024-1-011). A.Y. was also supported by the World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan.

Data Availability Statement

The five mitochondrial genomes sequenced in this study were deposited in the DDBJ database under accession nos. LC835033-5, LC842150, and LC860085. The protein tertiary structures predicted by AlphaFold3 and output files from InterProScan and Foldseek are available on Zenodo (https://doi.org/10.5281/zenodo.15079085).

References

Abrahamsen, M. S., T. J. Templeton, S. Enomoto, et al. 2004. "Complete Genome Sequence of the Apicomplexan, *Cryptosporidium parvum*." *Science* 304, no. 5669: 441–445. https://doi.org/10.1126/science.1094786.

Abramson, J., J. Adler, J. Dunger, et al. 2024. "Accurate Structure Prediction of Biomolecular Interactions With AlphaFold 3." *Nature* 630, no. 8016: 493–500. https://doi.org/10.1038/s41586-024-07487-w.

Altschul, S. F., T. L. Madden, A. A. Schäffer, et al. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25, no. 17: 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

Bágeľová Poláková, S., Ž. Lichtner, T. Szemes, M. Smolejová, and P. Sulo. 2021. "Mitochondrial DNA Duplication, Recombination, and Introgression During Interspecific Hybridization." *Scientific Reports* 11, no. 1: 12726. https://doi.org/10.1038/s41598-021-92125-y.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30, no. 15: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brown, M. W., A. A. Heiss, R. Kamikawa, et al. 2018. "Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group." *Genome Biology and Evolution* 10, no. 2: 427–433. https://doi.org/10.1093/gbe/evy014.

Burger, G., M. W. Gray, L. Forget, and B. F. Lang. 2013. "Strikingly Bacteria-Like and Gene-Rich Mitochondrial Genomes Throughout Jakobid Protists." *Genome Biology and Evolution* 5, no. 2: 418–438. https://doi.org/10.1093/gbe/evt008.

Burger, G., M. W. Gray, and L. B. Franz. 2003. "Mitochondrial Genomes: Anything Goes." *Trends in Genetics* 19, no. 12: 709–716. https://doi.org/10.1016/j.tig.2003.10.012.

Butenko, A., J. Lukeš, D. Speijer, and J. G. Wideman. 2024. "Mitochondrial Genomes Revisited: Why Do Different Lineages Retain Different Genes?" *BMC Biology* 22, no. 1: 15. https://doi.org/10.1186/s12915-024-01824-1.

Camacho, C., G. Coulouris, V. Avagyan, et al. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10: 421. https://doi.org/10.1186/1471-2105-10-421.

Čechová, J., J. Lýsek, M. Bartas, and V. Brázda. 2018. "Complex Analyses of Inverted Repeats in Mitochondrial Genomes Revealed Their Importance and Variability." *Bioinformatics* 34, no. 7: 1081–1085. https://doi.org/10.1093/bioinformatics/btx729.

Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. "Fastp: An Ultra-Fast All-In-One FASTQ Preprocessor." *Bioinformatics* 34, no. 17: i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Choi, I.-S., R. Jansen, and T. Ruhlman. 2019. "Lost and Found: Return of the Inverted Repeat in the Legume Clade Defined by Its Absence." *Genome Biology and Evolution* 11, no. 4: 1321–1333. https://doi.org/10.1093/gbe/evz076.

Christinaki, A. C., B. Theelen, A. Zania, et al. 2023. "Co-Evolution of Large Inverted Repeats and G-Quadruplex DNA in Fungal Mitochondria May Facilitate Mitogenome Stability: The Case of *Malassezia*." *Scientific Reports* 13, no. 1: 6308. https://doi.org/10.1038/s41598-023-33486-4.

Clark-Walker, G. D., C. R. McArthur, and K. S. Sriprakash. 1981. "Partial Duplication of the Large Ribosomal RNA Sequence in an Inverted Repeat in Circular Mitochondrial DNA From *Kloeckera africana*." *Journal of Molecular Biology* 147, no. 3: 399–415. https://doi.org/10.1016/0022-2836(81)90492-7.

Čutová, M., J. Manta, O. Porubiaková, et al. 2020. "Divergent Distributions of Inverted Repeats and G-Quadruplex Forming Sequences in *Saccharomyces cerevisiae*." *Genomics* 112, no. 2: 1897–1901. https://doi.org/10.1016/j.ygeno.2019.11.002.

Derelle, R., G. Torruella, V. Klimeš, et al. 2015. "Bacterial Proteins Pinpoint a Single Eukaryotic Root." *Proceedings of the National Academy of Sciences* 112, no. 7: E693–E699. https://doi.org/10.1073/pnas.1420657112.

Dobáková, E., P. Flegontov, T. Skalický, and J. Lukeš. 2015. "Unexpectedly Streamlined Mitochondrial Genome of the Euglenozoan *Euglena gracilis*." *Genome Biology and Evolution* 7, no. 12: 3358–3367. https://doi.org/10.1093/gbe/evv229.

Eglit, Y., T. Shiratori, J. Jerlström-Hultqvist, et al. 2024. "*Meteora sporadica*, a Protist With Incredible Cell Architecture, Is Related to Hemimastigophora." *Current Biology* 34, no. 2: 451–459. https://doi.org/10.1016/j.cub.2023.12.032.

Ettema, T. J. G., and S. G. E. Andersson. 2009. "The α-Proteobacteria: The Darwin Finches of the Bacterial World." *Biology Letters* 5, no. 3: 429–432. https://doi.org/10.1098/rsbl.2008.0793.

Feagin, J. E. 2000. "Mitochondrial Genome Diversity in Parasites." *International Journal for Parasitology* 30, no. 4: 371–390. https://doi.org/10.1016/S0020-7519(99)00190-3.

Ferandon, C., S. E. K. Chatel, B. Castandet, M. Castroviejo, and G. Barroso. 2008. "The *Agrocybe aegerita* Mitochondrial Genome Contains Two Inverted Repeats of the *nad4* Gene Arisen by Duplication on Both Sides of a Linear Plasmid Integration Site." *Fungal Genetics and Biology* 45, no. 3: 292–301. https://doi.org/10.1016/j.fgb.2007.10.012.

Férandon, C., J. Xu, and G. Barroso. 2013. "The 135 Kbp Mitochondrial Genome of *Agaricus bisporus* Is the Largest Known Eukaryotic Reservoir of Group I Introns and Plasmid-Related Sequences." *Fungal Genetics and Biology* 55: 85–91. https://doi.org/10.1016/j.fgb.2013.01.009.

Gastineau, R., S. Harðardóttir, C. Guilmette, et al. 2023. "Mitochondrial Genome Sequence of the Protist *Ancyromonas sigmoides* Kent, 1881 (Ancyromonadida) From the Sugluk Inlet, Hudson Strait, Nunavik, Québec." *Frontiers in Microbiology* 14: 1275665. https://doi.org/10.3389/fmicb.2023.1275665.

Gerhold, J. M., A. Aun, T. Sedman, P. Jõers, and J. Sedman. 2010. "Strand Invasion Structures in the Inverted Repeat of *Candida albicans* Mitochondrial DNA Reveal a Role for Homologous Recombination in Replication." *Molecular Cell* 39, no. 6: 851–861. https://doi.org/10.1016/j.molcel.2010.09.002.

Glücksman, E., E. A. Snell, and T. Cavalier-Smith. 2013. "Phylogeny and Evolution of Planomonadida (Sulcozoa): Eight New Species and New Genera *Fabomonas* and *Nutomonas*." *European Journal of Protistology* 49, no. 2: 179–200. https://doi.org/10.1016/j.ejop.2012.08.007.

Gray, M. W., B. F. Lang, R. Cedergren, et al. 1998. "Genome Structure and Gene Content in Protist Mitochondrial DNAs." *Nucleic Acids Research* 26, no. 4: 865–878. https://doi.org/10.1093/nar/26.4.865.

Greiner, S., P. Lehwark, and R. Bock. 2019. "OrganellarGenomeDRAW (OGDRAW) Version 1.3.1: Expanded Toolkit for the Graphical Visualization of Organellar Genomes." *Nucleic Acids Research* 47, no. W1: W59–W64. https://doi.org/10.1093/nar/gkz238.

Hallgren, J., K. D. Tsirigos, M. D. Pedersen, et al. 2022. "DeepTMHMM Predicts Alpha and Beta Transmembrane Proteins Using Deep Neural Networks." *bioRxiv*, 2022.04.08.487609. https://doi.org/10.1101/2022.04.08.487609.

Harada, R., Y. Hirakawa, A. Yabuki, et al. 2024. "Encyclopedia of Family A DNA Polymerases Localized in Organelles: Evolutionary Contribution of Bacteria Including the Proto-Mitochondrion." *Molecular Biology and Evolution* 41, no. 2: msae014. https://doi.org/10.1093/molbev/msae014.

He, D., C.-J. Fu, and S. L. Baldauf. 2016. "Multiple Origins of Eukaryotic *cox15* Suggest Horizontal Gene Transfer From Bacteria to Jakobid Mitochondrial DNA." *Molecular Biology and Evolution* 33, no. 1: 122–133. https://doi.org/10.1093/molbev/msv201.

Heiss, A. A., G. Walker, and A. G. B. Simpson. 2010. "Clarifying the Taxonomic Identity of a Phylogenetically Important Group of Eukaryotes: *Planomonas* Is a Junior Synonym of *Ancyromonas*." *Journal of Eukaryotic Microbiology* 57, no. 3: 285–293. https://doi.org/10.1111/j.1550-7408.2010.00477.x.

Janouškovec, J., D. V. Tikhonenkov, F. Burki, et al. 2017. "A New Lineage of Eukaryotes Illuminates Early Mitochondrial Genome Reduction." *Current Biology* 27, no. 23: 3717–3724.e5. https://doi.org/10.1016/j.cub.2017.10.051.

Janouškovec, J., D. V. Tikhonenkov, K. V. Mikhailov, et al. 2013. "Colponemids Represent Multiple Ancient Alveolate Lineages." *Current Biology* 23, no. 24: 2546–2552. https://doi.org/10.1016/j.cub.2013.10.062.

Jiang, M., Y. Ni, J. Li, and C. Liu. 2023. "Characterisation of the Complete Mitochondrial Genome of *Taraxacum mongolicum* Revealed Five Repeat-Mediated Recombinations." *Plant Cell Reports* 42, no. 4: 775–789. https://doi.org/10.1007/s00299-023-02994-y.

Jones, P., D. Binns, H.-Y. Chang, et al. 2014. "InterProScan 5: Genome-Scale Protein Function Classification." *Bioinformatics* 30, no. 9: 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Käll, L., A. Krogh, and E. L. L. Sonnhammer. 2004. "A Combined Transmembrane Topology and Signal Peptide Prediction Method." *Journal of Molecular Biology* 338, no. 5: 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.

Kamikawa, R., Y. Inagaki, and Y. Sako. 2007. "Fragmentation of Mitochondrial Large Subunit rRNA in the Dinoflagellate *Alexandrium catenella* and the Evolution of rRNA Structure in Alveolate Mitochondria." *Protist* 158, no. 2: 239–245. https://doi.org/10.1016/j.protis.2006.12.002.

Kamikawa, R., T. Shiratori, K. Ishida, H. Miyashita, and A. J. Roger. 2016. "Group II Intron-Mediated *Trans*-Splicing in the Gene-Rich Mitochondrial Genome of an Enigmatic Eukaryote, *Diphylleia rotans*." *Genome Biology and Evolution* 8, no. 2: 458–466. https://doi.org/10.1093/gbe/evw011.

Kamikawa, R., G. Tanifuji, M. Kawachi, H. Miyashita, T. Hashimoto, and Y. Inagaki. 2015. "Plastid Genome-Based Phylogeny Pinpointed the Origin of the Green-Colored Plastid in the Dinoflagellate *Lepidodinium chlorophorum*." *Genome Biology and Evolution* 7, no. 4: 1133–1140. https://doi.org/10.1093/gbe/evv060.

Kaur, B., K. Záhonová, M. Valach, et al. 2020. "Gene Fragmentation and RNA Editing Without Borders: Eccentric Mitochondrial Genomes of Diplonemids." *Nucleic Acids Research* 48, no. 5: 2694–2708. https://doi.org/10.1093/nar/gkz1215.

Kayama, M., K. Maciszewski, A. Yabuki, H. Miyashita, A. Karnkowska, and R. Kamikawa. 2020. "Highly Reduced Plastid Genomes of the Non-Photosynthetic Dictyochophyceans *Pteridomonas* spp. (Ochrophyta, SAR) Are Retained for tRNA-Glu-Based Organellar Heme Biosynthesis." *Frontiers in Plant Science* 11: 602455. https://doi.org/10.3389/fpls.2020.602455.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. "Graph-Based Genome Alignment and Genotyping With HISAT2 and HISAT-Genotype." *Nature Biotechnology* 37, no. 8: 907–915. https://doi.org/10.1038/s41587-019-0201-4.

Krämer, C., C. R. Boehm, J. Liu, et al. 2024. "Removal of the Large Inverted Repeat From the Plastid Genome Reveals Gene Dosage Effects and Leads to Increased Genome Copy Number." *Nature Plants* 10, no. 6: 923–935. https://doi.org/10.1038/s41477-024-01709-9.

Krogh, A., B. Larsson, G. von Heijne, and E. L. L. Sonnhammer. 2001. "Predicting Transmembrane Protein Topology With a Hidden Markov Model: Application to Complete Genomes." *Journal of Molecular Biology* 305, no. 3: 567–580. https://doi.org/10.1006/jmbi.2000.4315.

Lang, B. F., N. Beck, S. Prince, M. Sarrasin, P. Rioux, and G. Burger. 2023. "Mitochondrial Genome Annotation With MFannot: A Critical Analysis of Gene Identification and Gene Model Prediction." *Frontiers in Plant Science* 14: 1222186. https://doi.org/10.3389/fpls.2023.1222186.

Leger, M. M., M. Kolisko, R. Kamikawa, et al. 2017. "Organelles That Illuminate the Origins of *Trichomonas* Hydrogenosomes and *Giardia* Mitosomes." *Nature Ecology & Evolution* 1, no. 4: 0092. https://doi.org/10.1038/s41559-017-0092.

Locker, J., M. Rabinowitz, and G. S. Getz. 1974. "Tandem Inverted Repeats in Mitochondrial DNA of Petite Mutants of *Saccharomyces cerevisiae*." *Proceedings of the National Academy of Sciences* 71, no. 4: 1366–1370. https://doi.org/10.1073/pnas.71.4.1366.

Lonergan, K. M., and M. W. Gray. 1996. "Expression of a Continuous Open Reading Frame Encoding Subunits 1 and 2 of Cytochrome *c* Oxidase in the Mitochondrial DNA of *Acanthamoeba castellanii*."

Journal of Molecular Biology 257, no. 5: 1019–1030. https://doi.org/10.1006/jmbi.1996.0220.

Martin, M. 2011. "Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads." *EMBnet.Journal* 17, no. 1: 10–12. https://doi.org/10.14806/ej.17.1.200.

Matsuo, E., K. Morita, T. Nakayama, et al. 2022. "Comparative Plastid Genomics of Green-Colored Dinoflagellates Unveils Parallel Genome Compaction and RNA Editing." *Frontiers in Plant Science* 13: 918543. https://doi.org/10.3389/fpls.2022.918543.

Moreira, D., J. Blaz, E. Kim, and L. Eme. 2024. "A Gene-Rich Mitochondrion With a Unique Ancestral Protein Transport System." *Current Biology* 34, no. 16: 3812–3819.e3. https://doi.org/10.1016/j.cub.2024.07.017.

Nawrocki, E. P., and S. R. Eddy. 2013. "Infernal 1.1: 100-Fold Faster RNA Homology Searches." *Bioinformatics* 29, no. 22: 2933–2935. https://doi.org/10.1093/bioinformatics/btt509.

Nishimura, Y., K. Kume, K. Sonehara, et al. 2020. "Mitochondrial Genomes of *Hemiarma marina* and *Leucocryptos marina* Revised the Evolution of Cytochrome *c* Maturation in Cryptista." *Frontiers in Ecology and Evolution* 8: 140. https://doi.org/10.3389/fevo.2020.00140.

Nishimura, Y., G. Tanifuji, R. Kamikawa, A. Yabuki, T. Hashimoto, and Y. Inagaki. 2016. "Mitochondrial Genome of *Palpitomonas bilix*: Derived Genome Structure and Ancestral System for Cytochrome *c* Maturation." *Genome Biology and Evolution* 8, no. 10: 3090–3098. https://doi.org/10.1093/gbe/evw217.

Norman, J. E., and M. W. Gray. 2001. "A Complex Organization of the Gene Encoding Cytochrome Oxidase Subunit 1 in the Mitochondrial Genome of the Dinoflagellate, *Crypthecodinium cohnii*: Homologous Recombination Generates Two Different *cox1* Open Reading Frames." *Journal of Molecular Evolution* 53, no. 4: 351–363. https://doi.org/10.1007/s002390010225.

Nurk, S., D. Meleshko, A. Korobeynikov, and P. A. Pevzner. 2017. "metaSPAdes: A New Versatile Metagenomic Assembler." *Genome Research* 27, no. 5: 824–834. https://doi.org/10.1101/gr.213959.116.

Oborník, M., and J. Lukeš. 2015. "The Organellar Genomes of *Chromera* and *Vitrella*, the Phototrophic Relatives of Apicomplexan Parasites." *Annual Review of Microbiology* 69: 129–144. https://doi.org/10.1146/annurev-micro-091014-104449.

Palmer, J. D., and W. F. Thompson. 1982. "Chloroplast DNA Rearrangements Are More Frequent When a Large Inverted Repeat Sequence Is Lost." *Cell* 29, no. 2: 537–550. https://doi.org/10.1016/0092-8674(82)90170-2.

Pérez-Brocal, V., R. Shahar-Golan, and C. G. Clark. 2010. "A Linear Molecule With Two Large Inverted Repeats: The Mitochondrial Genome of the Stramenopile *Proteromonas lacertae*." *Genome Biology and Evolution* 2: 257–266. https://doi.org/10.1093/gbe/evq015.

Powell, H. R., S. A. Islam, A. David, and M. J. E. Sternberg. 2025. "Phyre2.2: A Community Resource for Template-Based Protein Structure Prediction." *Journal of Molecular Biology* in press. https://doi.org/10.1016/j.jmb.2025.168960.

Richards, T. A., L. Eme, J. M. Archibald, et al. 2024. "Reconstructing the Last Common Ancestor of All Eukaryotes." *PLoS Biology* 22, no. 11: e3002917. https://doi.org/10.1371/journal.pbio.3002917.

Roger, A. J., S. A. Muñoz-Gómez, and R. Kamikawa. 2017. "The Origin and Diversification of Mitochondria." *Current Biology* 27, no. 21: R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015.

Rost, B. 1999. "Twilight Zone of Protein Sequence Alignments." *Protein Engineering* 12, no. 2: 85–94. https://doi.org/10.1093/protein/12.2.85.

Salavirta, H., I. Oksanen, J. Kuuskeri, et al. 2014. "Mitochondrial Genome of *Phlebia radiata* Is the Second Largest (156 kbp) Among Fungi and Features Signs of Genome Flexibility and Recent Recombination Events." *PLoS One* 9, no. 5: e97141. https://doi.org/10.1371/journal.pone.0097141.

Slamovits, C. H., J. F. Saldarriaga, A. Larocque, and P. J. Keeling. 2007. "The Highly Reduced and Fragmented Mitochondrial Genome of the Early-Branching Dinoflagellate *Oxyrrhis marina* Shares Characteristics With Both Apicomplexan and Dinoflagellate Mitochondrial Genomes." *Journal of Molecular Biology* 372, no. 2: 356–368. https://doi.org/10.1016/j.jmb.2007.06.085.

Spencer, D. F., and M. W. Gray. 2011. "Ribosomal RNA Genes in *Euglena gracilis* Mitochondrial DNA: Fragmented Genes in a Seemingly Fragmented Genome." *Molecular Genetics and Genomics* 285, no. 1: 19–31. https://doi.org/10.1007/s00438-010-0585-9.

Stern, D. B., and J. D. Palmer. 1984. "Recombination Sequences in Plant Mitochondrial Genomes: Diversity and Homologies to Known Mitochondrial Genes." *Nucleic Acids Research* 12, no. 15: 6141–6157. https://doi.org/10.1093/nar/12.15.6141.

Tikhonenkov, D. V., J. Janouškovec, A. P. Mylnikov, et al. 2014. "Description of *Colponema vietnamica* sp.n. and *Acavomonas peruviana* n. gen. n. sp., Two New Alveolate Phyla (Colponemidia nom. nov. and Acavomonidia nom. nov.) and Their Contributions to Reconstructing the Ancestral State of Alveolates and Eukaryotes." *PLoS One* 9, no. 4: e95467. https://doi.org/10.1371/journal.pone.0095467.

Tikhonenkov, D. V., K. V. Mikhailov, R. M. R. Gawryluk, et al. 2022. "Microbial Predators Form a New Supergroup of Eukaryotes." *Nature* 612, no. 7941: 714–719. https://doi.org/10.1038/s41586-022-05511-5.

Torruella, G., L. J. Galindo, D. Moreira, and P. López-García. 2025. "Phylogenomics of Neglected Flagellated Protists Supports a Revised Eukaryotic Tree of Life." *Current Biology* 35, no. 1: 198–207.e4. https://doi.org/10.1016/j.cub.2024.10.075.

Valach, M., G. Burger, M. W. Gray, and B. F. Lang. 2014. "Widespread Occurrence of Organelle Genome-Encoded 5S rRNAs Including Permuted Molecules." *Nucleic Acids Research* 42, no. 22: 13764–13777. https://doi.org/10.1093/nar/gku1266.

van Kempen, M., S. S. Kim, C. Tumescheit, et al. 2024. "Fast and Accurate Protein Structure Search With Foldseek." *Nature Biotechnology* 42, no. 2: 243–246. https://doi.org/10.1038/s41587-023-01773-0.

Verner, Z., S. Basu, C. Benz, et al. 2015. "Malleable Mitochondrion of *Trypanosoma brucei*." In *International Review of Cell and Molecular Biology*, edited by K. W. Jeon, vol. 315, 73–151. Academic Press. https://doi.org/10.1016/bs.ircmb.2014.11.001.

Vosseberg, J., J. J. E. van Hooff, S. Köstlbacher, K. Panagiotou, D. Tamarit, and T. J. G. Ettema. 2024. "The Emerging View on the Origin and Early Evolution of Eukaryotic Cells." *Nature* 633, no. 8029: 295–305. https://doi.org/10.1038/s41586-024-07677-6.

Williams, B. A. P., R. P. Hirt, J. M. Lucocq, and T. M. Embley. 2002. "A Mitochondrial Remnant in the Microsporidian *Trachipleistophora hominis*." *Nature* 418, no. 6900: 865–869. https://doi.org/10.1038/nature00949.

Williamson, K., L. Eme, H. Baños, et al. 2025. "A Robustly Rooted Tree of Eukaryotes Reveals Their Excavate Ancestry." *Nature* 640, no. 8060: 947–981. https://doi.org/10.1038/s41586-025-08709-5.

Wynn, E. L., and A. C. Christensen. 2019. "Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution." *G3: Genes, Genomes, Genetics* 9, no. 2: 549–559. https://doi.org/10.1534/g3.118.200948.

Yang, J., T. Harding, R. Kamikawa, A. G. B. Simpson, and A. J. Roger. 2017. "Mitochondrial Genome Evolution and a Novel RNA Editing System in Deep-Branching Heteroloboseids." *Genome Biology and Evolution* 9, no. 5: 1161–1174. https://doi.org/10.1093/gbe/evx086.

Yazaki, E., R. Harada, R. Isogai, et al. 2025. "*Glissandra oviformis* n. sp.: A Novel Predatory Flagellate Illuminates the Character Evolution

Within the Eukaryotic Clade CRuMs." *Open Biology* 15: 250057. https://doi.org/10.1098/rsob.250057.

Yazaki, E., A. Yabuki, Y. Nishimura, T. Shiratori, T. Hashimoto, and Y. Inagaki. 2022. "*Microheliella maris* Possesses the Most Gene-Rich Mitochondrial Genome in Diaphoretickes." *Frontiers in Ecology and Evolution* 10: 1030570. https://doi.org/10.3389/fevo.2022.1030570.

Yubuki, N., G. Torruella, L. J. Galindo, et al. 2023. "Molecular and Morphological Characterization of Four New Ancyromonad Genera and Proposal for an Updated Taxonomy of the Ancyromonadida." *Journal of Eukaryotic Microbiology* 70, no. 6: e12997. https://doi.org/10.1111/jeu.12997.

Zerbino, D. R., and E. Birney. 2008. "Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs." *Genome Research* 18, no. 5: 821–829. https://doi.org/10.1101/gr.074492.107.

Zhang, Y., and J. Skolnick. 2005. "TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score." *Nucleic Acids Research* 33, no. 7: 2302–2309. https://doi.org/10.1093/nar/gki524.

Zimmermann, L., A. Stephens, S.-Z. Nam, et al. 2018. "A Completely Reimplemented MPI Bioinformatics Toolkit With a New HHpred Server at Its Core." *Journal of Molecular Biology* 430, no. 15: 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.