Evolutionary History of Mitochondrial Genomes in Discoba, Including the Extreme Halophile *Pleurostomum flabellatum* (Heterolobosea)

Khaoula Ettahi¹, Duckhyun Lhee¹, Ji Yeon Sung², Alastair G.B. Simpson³,⁴, Jong Soo Park ²,⁵,*, and Hwan Su Yoon ¹,*

¹Department of Biological Sciences, Sungkyunkwan University, Suwon, South Korea
²Department of Oceanography, Kyungpook Institute of Oceanography, School of Earth System Sciences, Kyungpook National University, Daegu, South Korea
³Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada
⁴Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Nova Scotia, Canada
⁵Research Institute for Dok-do and Ulleung-do Island, Kyungpook National University, Daegu, South Korea

*Corresponding authors: E-mails: jongsoopark@knu.ac.kr; hsyoon2011@skku.edu.

Accepted: 10 November 2020

**Abstract**

Data from Discoba (Heterolobosea, Euglenozoa, Tsukubamonadida, and Jakobida) are essential to understand the evolution of mitochondrial genomes (mitogenomes), because this clade includes the most primitive-looking mitogenomes known, as well some extremely divergent genome information systems. Heterolobosea encompasses more than 150 described species, many of them from extreme habitats, but only six heterolobosean mitogenomes have been fully sequenced to date. Here we complete the mitogenome of the heterolobosean *Pleurostomum flabellatum*, which is extremely halophilic and reportedly also lacks classical mitochondrial cristae, hinting at reduction or loss of respiratory function. The mitogenome of *P. flabellatum* maps as a 57,829-bp-long circular molecule, including 40 coding sequences (19 tRNA, two rRNA, and 19 orfs). The gene content and gene arrangement are similar to *Naegleria gruberi* and *Naegleria fowleri*, the closest relatives with sequenced mitogenomes. The *P. flabellatum* mitogenome contains genes that encode components of the electron transport chain similar to those of *Naegleria* mitogenomes. Homology searches against a draft nuclear genome showed that *P. flabellatum* has two homologs of the highly conserved Mic60 subunit of the MICOS complex, and likely lost Mic19 and Mic10. However, electron microscopy showed no cristae structures. We infer that *P. flabellatum*, which originates from high salinity (313 ‰) water where the dissolved oxygen concentration is low, possesses a mitochondrion capable of aerobic respiration, but with reduced development of cristae structure reflecting limited use of this aerobic capacity (e.g., microaerophily).

**Key words:** *Pleurostomum flabellatum*, Heterolobosea, *Naegleria* with cristae, MICOS complex, microaerophily.

**Significance**

*Pleurostomum flabellatum* (Discoba: Heterolobosea) is a microbial eukaryote that thrives in near-saturated brine at 40 °C, where dissolved oxygen concentration is several times lower than in typical seawater or freshwater—this, together with a lack of obvious mitochondrial cristae hinted that *P. flabellatum* could be an anaerobe. We sequenced the 57,829-bp mitochondrial genome of *P. flabellatum*, revealing a similar complement of electron transport chain protein-coding genes to the related aerobe *Naegleria*, and identified two homologs of the mitochondrial cristae organization protein Mic60 in preliminary nuclear genome data. In addition to demonstrating *P. flabellatum*’s capacity for aerobic respiration (possibly as a microaerophile), we add a mitochondrial genomic perspective to the picture of the evolution of Discoba’s remarkable mitogenome diversity.
Introduction

Endosymbiosis is an important evolutionary process that has greatly impacted the evolution of life and continues to shape eukaryotic cells and their genomes (Lang et al. 1997). Mitochondria are eukaryotic organelles that originated from an endosymbiotic \(\alpha\)-proteobacterium, perhaps \(\sim 1.8\) billion years ago (Bhattacharya et al. 2004; Gray 2012). Because then, mitochondrial genomes (mitogenomes) evolved specific DNA structure and genetic functions (Gray et al. 1989, 1999; Lang et al. 1997; Gray 2012) through massive gene loss, by a combination of endosymbiotic gene transfer (EGT) and gene elimination (Adams and Palmer 2003). Endosymbiotic gene transfer occurs when a gene migrates from the organellar genome to the host nucleus (and is subsequently lost from the organellar genome). In contrast, gene elimination occurs when a gene is no longer required by the host—endosymbiont holobiont. This history of evolutionary change began before the diversification of the major eukaryotic lineages, leading to substantial variation in gene content among the mitogenomes of extant eukaryotes (Hancock et al. 2010).

During mitochondrial evolution, many homologous genes have frequently been lost in separate branches of the eukaryotic tree, making the pattern and the prevalence of mitogenome gene loss more complicated to understand (Gray 1998; Lang et al. 1999; Adams and Palmer 2003). Reconstructing an evolutionary history of mitogenomes in a given clade requires mitogenome sequences from taxa that represent the variety of the investigated clade, as well as an accurate phylogeny of the group (Chihade et al. 2000).

The taxon Discoba consists of four main lineages—Jakobida, Euglenozoa, Heterolobosea, and Tsukubamonadida—that have various unique mitogenome features (Gray et al. 2004; Simpson et al. 2006; Hampel et al. 2009). Jakobida has the most bacteria-like and gene-rich mitogenomes known. For instance, the mitogenome of *Andalucia godoyi* comprises 100 functionally assignable genes, with 66 and 34 genes encoding proteins and structural RNAs, respectively (Burger et al. 2013; Gray et al. 2020). Several unique and/or complex features characterize the mitogenomes of Euglenozoa (Kinetoplastea, Diplonemea, and Euglenida) (Roy et al. 2007; Dobáková et al. 2015). For example, kinetoplastid mitogenomes are present as the “kinetoplast” (thus, “kDNA”), which typically includes two different DNA molecule classes called maxicircles and minicircles (Lukeš et al. 2002).

Heterolobosea (Discoba), meanwhile, is a major heterotrophic protist group containing \(\sim 150\) described species. In addition to mesophiles that live in conventional freshwater, seawater, or soil systems, this group is notable for including many species that are halophiles, acidophiles, thermophiles, or anaerobes (Pánek et al. 2014; Park and Simpson 2015). It seems that anaerobic lineages arose at least three times in Heterolobosea. Psalteriomonadidae, Creneis, and Dactylomonas represent distinct lineages that possess mitochondrion-related organelles that lack cristae and are known or suspected to function anaerobically, for example, as hydrogenosomes (Park et al. 2007; de Graaf et al. 2009; Barberá et al. 2010; Hanousková et al. 2019). It is suggested that the lack of cristae in anaerobic protists such as psalteriomonads is related to a loss of the mitochondrial contact site and cristae organizing system (MICOS) (see fig. 5 in Muñoz-Gómez, Slamovits, Dacks, Baier, et al. 2015), localized in the cristae junctions. The MICOS complex of yeast consists of six proteins: Mic10, Mic12, Mic19, Mic26, Mic28, and Mic60. The most ancient component of the MICOS subunits may be Mic60, which seems to have originated from \(\alpha\)-proteobacterium, whereas Mic10 is the most widespread component among eukaryotes (Muñoz-Gómez, Slamovits, Dacks, and Wideman 2015; Muñoz-Gómez, Slamovits, Dacks, Baier, et al.; Gray et al. 2020). Some Discoba genomes were previously reported to encode Mic10, Mic19, and Mic60 homologs (fig. 5; Muñoz-Gómez, Slamovits, Dacks, Baier, et al.; Gray et al. 2020). The well-studied heterolobosean *Naegleria gruberi* is a free-living amoeba, closely related to the human pathogen *Naegleria fowleri* that is the causative agent of the deadly human disease primary amoeboencephalitis. Muñoz-Gómez, Slamovits, Dacks, Baier, et al. (2015) examined the phylogenetic distribution of MICOS subunits and identified two MICOS subunits (Mic60 and Mic19) in *N. gruberi*. Furthermore, *N. gruberi* was suggested to contain genes related to anaerobic energy metabolism (e.g., \([FeFe]\)-hydrogenase, Fritz-Laylin et al. 2010), although a more recent study suggests that hydrogen production occurs exclusively in the cytoplasm (Tsaoiss et al. 2014). Thus, there may be a broad and complex diversity of functions in aerobic, anaerobic, or microaerobic Heterolobosea.

*Pleurostomum flabellatum* is an obligate extremely halophilic heterolobosean. It can grow optimally in more than 30\(^\circ\)C salinity water at 40 \(^\circ\)C, where theoretical oxygen saturation (1.26 mg l\(^{-1}\)) is 3.5 times lower than in freshwater (4.47 mg l\(^{-1}\)) at 40 \(^\circ\)C (Battino et al. 1983). This organism is reported to lack mitochondrial cristae (Park et al. 2007) and is sometimes listed as an additional (fourth) lineage of anaerobic heteroloboseans (Pánek et al. 2014; Hanousková et al. 2019). However, nothing is known about the function of mitochondrion-related organelles in *P. flabellatum* (Park et al. 2007). Codon usage by extremophiles, namely halophiles and thermophiles, has been reported to be significantly different from non-extremophile organisms, at least in prokaryotes (Khan and Patra 2018). Because *P. flabellatum* is both a candidate anaerobe and confirmed extreme halophile, it is a particularly interesting species to examine to better understand genome evolution in anaerobic and/or halophilic eukaryotes.

In this study, we assembled the mitogenome of *P. flabellatum* and analyzed its gene composition and gene arrangements. We performed comparative and phylogenetic analysis.
analyses of the \textit{P. flabellatum} mitogenome together with those of other discoids, including six heteroloboseids. Also, we studied codon usage bias to detect possible variation between heteroloboseid mitogenomes and performed a GC skewness analysis to compare inferred replication patterns of mitochondria in Discoba. \textit{P. flabellatum} is sister to \textit{Naegleria} spp. in our phylogenetic analysis of mitogenome data and is likely a microaerophile rather than an anaerobe based on the mitogenome gene repertoire. The apparent absence of cristae-like structures may be due to the lack of genes encoding several MICOs components.

**Materials and Methods**

**Isolation, Cultivation, and Ultrastructure**

\textit{Pleurostomum flabellatum} strain EHF1 (10–14 \textmu m-long) was initially isolated from 3130\textsubscript{salin} salinity water collected from the Soosung solar saltern located at Seosin on the west coast of the Republic of Korea (36°09’36”N, 126°40’44”E; Park et al. 2007). The culture of \textit{P. flabellatum} was maintained in 200\textsubscript{salin} salinity media, incubated at 37°C and subcultured every 4 weeks for 12 years. In brief, 0.1 ml of fluid from the culture was used to inoculate 5 ml of 200\textsubscript{salin} liquid media, made by dilution of Medium V (300 g NaCl, 7.6 g KCl, 17.8 g MgCl\textsubscript{2}, 1.8 g MgSO\textsubscript{4} \cdot 7H\textsubscript{2}O, 1.3 g CaCl\textsubscript{2} l\textsuperscript{-1} water; see Park 2012) with sterile distilled water, and supplemented with Marine Broth 2216 (final concentration of 0.5%; Difco) plus autoclaved barley grains to grow indigenous prokaryotes in the culture as prey. Before DNA extraction, approximately 2.75 l of well-grown culture (15,000 \textit{P. flabellatum} cells per ml) in liquid media (200\textsubscript{salin} salinity) was prepared. For ultrastructural sectioning, the culture of \textit{P. flabellatum} was grown in 250\textsubscript{salin} salinity media. All procedures for transmission electron microscopy (TEM) were as described in Park et al. (2007).

**DNA Extraction and Sequencing**

After the culture was pre-filtered using a 38-\mu m mesh sieve, cells were harvested using a 10-\mu m pore size membrane filter (Advantec, Tokyo, Japan). A DNeasy Plant Mini Kit (Qiagen, Santa Clarita, CA) was used for DNA extraction. A sequencing library was constructed using the Truseq Nano DNA Prep Kit (Insert 550-bp) based on the manufacturer’s manual. Samples were sequenced as 150-bp paired-end reads on the Illumina Novaseq6000 platform at DNA-Link Inc. (Seoul, Korea).

**Mitogenome Assembly and Annotation**

The mitogenome of \textit{P. flabellatum} was assembled using the \textit{de novo} assembler NOVOPlasty 2.7.2 (Dierckxsens et al. 2017) from the Illumina whole-genome sequencing data. Here, the mitochondrial gene sequences from the closely related species \textit{N. gruberi} (GenBank accession number = AF288092) was used as the seed sequence for NOVOPlasty. The processed reads were mapped to the assembly to identify potential assembly errors, followed by coverage evaluation. Open reading frames (orfs) were identified by Geneious 8.1.2 (https://www.geneious.com, last accessed November 1, 2018), with the universal genetic code, as this code is used in closely related species within Heterolobosea. Annotation of putative protein-coding sequences (CDSs) was inferred first with BLASTX (e-value \( \leq 1.0 \text{e}^{-05} \)) by searching against the non-redundant (nr) protein database for sequence similarities. Less well-preserved genes (i.e., that were not annotated with BLASTX) were subjected to HMMER searches (http://hmmer.janelia.org, last accessed November 1, 2018) derived from models of all available mitochondrion-encoded proteins. Multiple protein-alignment examinations were further used to approve or reject orfs with significant or close-to-significant sequence similarity.

Large subunit (mL) and small subunit (mS) ribosomal RNA (rRNA) genes were identified by RNAmmer 1.2 (www.cbs.dtu.dk/services/RNAmmer, last accessed November 1, 2018) and transfer RNA (tRNA) genes were detected using ARAGORN (http://mbio-serv2.mbioekol.lu.se/ARAGORN, last accessed November 1, 2018). Structural comparison of mitogenomes was performed using progressive Mauve (Darling et al. 2010). The circular mitogenome map was visualized using DNAPlotter 17.0.1 (Carver et al. 2009). Genome sequences were deposited in the NCBI GenBank database under the following accession number MT843578.

**Phylogenetic Analysis**

For the phylogenetic analysis of Discoba, 20 complete mitogenomes were obtained from GenBank as follows: seven heteroloboseids, including \textit{P. flabellatum} (this study), \textit{Naegleria gruberi} (Fritz-Laylin et al. 2010, 2011; AF288092), \textit{Naegleria fowleri} (KX589092; Herman et al. 2013), Acrasis kona (KJ679272; Fu et al. 2014), \textit{Stachyamoeba lipophora} (KP165388, Valach et al. 2014), \textit{Pharyngomonas kirbyi} (NC_034798; Park and Simpson 2011; Yang et al. 2017), and the undescribed amoeba “Heteroloboseid sp. BB2” (KY379823; Yang et al. 2017); \textit{Tsukubamonadida} (AB854048; Kamikawa et al. 2014); five euglenozoans: \textit{Diplonema papillatum} (KJ679272; Fu et al. 2014), \textit{Diplonema ambulator} (MF436761, Valach et al. 2017), \textit{Rynchopus euleides} (EU123537; Marande et al. 2005), \textit{Diplonema ambulator} (MF436761, Valach et al. 2017), \textit{Rynchopus euleides} (EU123537; Marande et al. 2005), \textit{Trypanosoma rangeli} (KJ083830, Stoco et al. 2014), and \textit{Anomogamus deanei} (KU778684); and seven jakobids: \textit{Jakoba bahamiensis} (KC353354; Burger et al. 2013), \textit{Jakoba libera} (KC353355; Burger et al. 2013), \textit{Reclinomonas americana} (KC353356; Lang et al. 1997), \textit{Histiona aroides} (KC353353; Burger et al. 2013), \textit{Seculamonas ecudoriensis} (KC353355; Burger et al. 2013), \textit{Andalucia godoyi} (KC353352; Burger et al. 2013), and \textit{Ophirina amphinema} (LC369600; Yabuki et al. 2018).
Conserved protein sequences shared by the discobid groups were retrieved (14 proteins total: nad1-5, nad7-9, nad4L, cox1-3, atp6, rps12). The amino acid sequences were aligned and concatenated using MAFFT 7.1.10 (Katoh and Standley 2013). Aligned mitogenome genes were concatenated for multigene phylogenetic analysis. The phylogenetic trees were inferred from the resulting concatenated amino acid alignment. A maximum likelihood (ML) tree was constructed using IQ-TREE 1.6.7 with LG+F+I+G4 as the best-fit model of sequence evolution (as chosen by ModelFinder) and with robustness evaluated with a 1000-replicate UFBoot bootstrap approximation (Minh et al. 2013; Flouri et al. 2015; Nguyen et al. 2015; Kalyaanamoorthy et al. 2017). A Bayesian analyses was carried out using PhyloBayes v8.28 (Lartillot et al. 2009) under the CAT+F+I+G4 model. Two independent chains were run with discrete gamma distribution with four categories, and a burn-in of 1000.

Mitogenome Replication and Codon Bias

The Relative Synonymous Codon Usage (RSCU) values from mitogenome CDSs were calculated to characterize synonymous codon usage, and correspondence analysis (COA) was performed, both using codonW (http://codonw.sourceforge.net/culong.html, last accessed November 15, 2018).

Cumulative GC-skew values were examined to find the origin point (oriC) of replication with window/step settings of 1000/300 (Eppinger et al. 2004). This method measures the asymmetric strand distribution of G and C using the formula GC-skew = [G − C]/[G + C] and accumulates the obtained values (Perna and Kocher 1995). The GC skewness has been used in several studies to assess the strand bias of nucleotide composition of different eukaryotic mitogenomes (e.g., Burger et al. 2013; Jackson and Reyes-Prieto 2014). This strand bias varies gradually along a genome, and the region with the lowest GC skew value indicates where the origin of replication might be located. GC-skew values for selected discobids with circular mitogenomes were obtained using fasta_g_c_skew.py program (https://github.com/shenwei356/bio_scripts, last accessed November 15, 2018).

Homology Search

Proteins homologous to MICOS components were searched for using BLAST against the list of MICOS protein sequences from Muñoz-Gómez, Slamovits, Dacks, Baier, et al. (2015), that is, Mic10, Mic12, Mic19, Mic25, Mic26, Mic27, and Mic60. The sequences that were hit with an e-value lower than 0.10 were considered as significant. The two Mic60 sequences were compared against protein-coding regions within the P. flabellatum transcript assembly set (N50 = 3,485), using BLAST searches with an e-value cut-off ≤1.0e-05. Then, we searched for protein signatures for the two Mic60-like sequences were deposited in the NCBI GenBank (accession numbers: MW019459 and MW019460).

Results and Discussion

Pleurostomum flabellatum Contains Classical Mitochondrial Genome

General Features

We assembled the complete mitogenome of P. flabellatum into a single circular map of 57.8-kbp (fig. 1). This size is within the previously reported range for heterolobosean mitogenomes (49.5–75.7 kbp; supplementary table S1, Supplementary Material online), except for the early-diverged Heteroloboseid sp. BB2, which has a larger mitogenome (119.3-kbp), mostly due to a 49-kbp inverted repeat (IR) region (Yang et al. 2017). The proportion of coding regions in the P. flabellatum mitogenome (86.1%) is similar to that of other heteroloboseans (81–93.2%), and jakobids (78.5–94.1%) (Burger et al. 2013; Yabuki et al. 2018), but higher than in Tsukubamonas (65.7%) (supplementary table S1, Supplementary Material online). The P. flabellatum mitogenome is AT-rich (71.1%), which is also consistent with other heteroloboseids (AT content = 69.3–87.5%) (supplementary table S1, Supplementary Material online).

The mitogenome of P. flabellatum contains 40 CDSs, two rRNAs, 20 tRNAs. In addition to CDSs with known functions, the P. flabellatum mitogenome contains 19 predicted rRNA, including orf145, which also occurs in Naegleria mitogenomes (supplementary fig. S1, Supplementary Material online). The orf145 from P. flabellatum shares 59.3% amino acid sequence similarity with N. gruberi and 58.8% with N. fowleri. However, the function of orf145 remains unknown. The gene set of the P. flabellatum mitogenome is similar to those of other heteroloboseans (rRNAs: 11–24, CDSs: 26–42) (supplementary tables S2 and S3, Supplementary Material online). All the genes are located on the same strand except ccmF and trnY (fig. 1).

Codon Usage

Codon usage in halophilic (P. flabellatum and P. kirbyi) and thermophilic (Heteroloboseid sp. BB2) heterolobosean mitogenomes was compared with that of the mesophiles N. gruberi, N. fowleri, A. kona, and S. lipophora. A supplementary table S4, Supplementary Material online, summarizes the use of amino acids and relative synonymous codon usage (RSCU) values in the CDS sequences. The RSCU values of all heteroloboseans revealed that AT-rich codons are favored over synonymous codons with lower A/T content (supplementary fig. S2, Supplementary Material online).
The examination of the correspondence (COA) was carried out based on RSCU values of heterolobosean mitogenomes. Correlation analysis was conducted on the difference in codon patterns represented by the first and second major COA axis (supplementary fig. S3, Supplementary Material online). Heterolobosean mitogenomes showed low variability in codon usage, with no specific patterns between extremophiles and mesophiles. Overall, we found no significant difference between the mitogenome codon usage of heterolobosean extremophiles and non-extremophiles.

**Phylogeny and Evolutionary Inferences Based on Mitochondrial Proteins**

The Discoba organismal phylogeny was inferred from the concatenated amino acid data set of 14 conserved mitochondrial proteins (fig. 2A). Our phylogeny is consistent with the immediate sister relationship between the Heterolobosea and Euglenozoa clades to the exclusion of Jakobida and Tsukubamonadida (Hampl et al. 2009; Kamikawa et al. 2014; Yang et al. 2017), although the heterolobosean clade itself was poorly supported (BS = 62%; PP = 0.9).
**FIG. 2.**—Phylogenetic tree of the Discoba based on mitogenome data and synteny comparisons among the Heterolobosea. (A) Phylogenetic relationships between heteroloboseids and other discobids deduced from the analysis of 14 concatenated mitochondrial genes from 20 discobids. The tree was inferred based on amino acid sequences using maximum-likelihood (ML) under the LG+F+I+G4 model and Bayesian phylogenetic analysis under the CAT+GTR model. Numbers at nodes indicate UFBoot ML bootstrap approximation percentages and Bayesian posterior probability, slash separated and “*” means not recovered. The tree is rooted with the Jakobida as sister to other Discoba following Yang et al. (2017) and Yabuki et al. (2018). Habitats of heteroloboseid taxa are indicated below each species name. Structures of euglenozoan mitogenomes are listed following Faktorová et al. (2016). (B) Synteny comparison among five Tetramitia species (Naegleria gruberi, N. fowleri, Pleurostomum flabellatum, Acrasis kona, and Stachyamoeba lipophora) using the Mauve program.

**Pleurostomum** and **Naegleria** are clustered with strong statistical support (BS = 100%; PP = 1) and fall along with **Acrasis** and **Stachyamoeba** into a robust Tetramitia clade (BS = 100%; PP = 1). This **Pleurostomum**+**Naegleria** cluster is consistent with previous inferences from 18S rRNA genes (Park et al. 2007, 2012; Harding et al. 2013; Tyml et al. 2017; Jhin and Park 2019). Heteroloboseid BB2 and **Pharyngomonas** are monophyletic in the ML tree with low support.
Mitogenome Rearrangement

The alignment of heterolobosean mitogenomes was inspected to investigate the arrangement of the *P. flabellatum* mitogenome (fig. 2B). Synteny appeared to be highly conserved between *P. flabellatum* and *Naegleria* spp., where a total of 13 syntenic blocks are well preserved in the same order. *P. flabellatum* has more regions containing orfs of unknown function (figs. 1 and 2B). In contrast, several rearrangements are evident among *P. flabellatum*, *A. kona*, and *S. lipophora*, which suggest that the shuffling has occurred after the split of these lineages. In terms of mitogenome size, *A. kona* is close to *P. flabellatum* and *Naegleria* spp. but has a very different organization and gene content (Fu et al. 2014) (fig. 2B). Conserved synteny blocks in *A. kona* and *S. lipophora* counts for 10 and 12 blocks, respectively. In addition, syntenies of *P. flabellatum* with *P. kirbyi* showed four low homology blocks where the first two *P. flabellatum* synteny blocks are inverted in *P. kirbyi* (supplementary fig. S5, Supplementary Material online). The structures of the heterolobosea mitogenomes are thus quite diverse.

Characterization of GC Skewness in Discoba

There is a negative AT skew over the *P. flabellatum* mitogenome (−0.094) indicating a slight bias towards the use of T (38.9%) over A (32.2%), as well as a positive GC-skew (+0.214) showing a preference toward the use of G (38.9%) over A (32.2%), as well as a positive GC-skew (+0.214) showing a preference toward the use of G (38.9%) over A (32.2%). This skewness is also usually found in circular plasmids, known to replicate by unidirectional replication of the rolling circle (RCR) (Arakawa et al. 2009).

However, *R. americana* and *H. aroides* display prominent bimodal GC-skew curves—more distinctly in *R. americana*—which is characteristic of the classical bidirectional theta mode (fig. 3; Burger et al. 2013). The GC-skew graphs of the remaining discobids (*P. kirbyi*, BB2, *A. godoyi*, *S. ecuadoriensis*) mitogenomes are insufficient to draw meaningful inferences about the replication mechanism (fig. 3; Burger et al. 2013). Interestingly, similar GC trajectories have been observed in bacterial genomes (Saha et al. 2014) and have been shown to be induced by genomic rearrangement types mixing leading and lagging sequences that influence trends in local base usage.

Distribution of Gene Losses in Discoba

Heterolobosea *sensu lato* is divided into two subgroups: Tetramitida (represented here by four genera) and Pharyngomonas + BB2 (fig. 2), albeit the latter grouping was paraphyletic in our Bayesian analysis (Cavalier-Smith and Nikolaev 2008; Harding et al. 2013; Yang et al. 2017; Petú et al. 2018). The cluster Pharyngomonas + BB2 includes two extra CDS genes (*rpl10*, *rpl32*) and three rRNAs (*trnA*, *trnC*, and *trnG*). Four additional CDSs are distributed unequally over the mitogenomes of the Tetramitida genera (fig. 4, supplementary tables S2 and S3, Supplementary Material online); three cytochrome c matrasure subunits; the ABC-transporter *ccmA* and heme transfer proteins *ccmC* and *ccmF*, along with ribosomal protein S4 (*rps4*). The twin-arginine translocase tatC was recently identified starts from specific replication origins by a unidirectional theta-type intermediate and ends at the replication terminus, nearly opposite of the origin (Lobry 1996; Hines and Ray 2011). Likewise, the kinetoplast DNA molecules—both minicircles and maxicircles—replicate from specific replication origins via unidirectional theta-type intermediates (Hines and Ray 2011).

Here, we analyzed the cumulative GC-skew plots of the circular-mapping mitogenomes of discobids (fig. 3). Heterolobosean GC-skew trajectories and the cumulative GC-skew curves map simply to organismal phylogeny (fig. 3). Cumulative GC-skew of the *P. flabellatum* mitogenome shows a homogenous and positive skew curve along the sequence, resembling the situation in *A. kona*, *Naegleria* spp., and *S. lipophora*. A similar pattern is seen in some jacobids; *J. bahamiensis* and *O. amphimena*, as well as *Tsukubamonas* (*T. globosa*) (fig. 3B; Burger et al. 2013).

The twin-arginine translocase tatC was recently identified starts from specific replication origins by a unidirectional theta-type intermediate and ends at the replication terminus, nearly opposite of the origin (Lobry 1996; Hines and Ray 2011). Likewise, the kinetoplast DNA molecules—both minicircles and maxicircles—replicate from specific replication origins via unidirectional theta-type intermediates (Hines and Ray 2011).

Here, we analyzed the cumulative GC-skew plots of the circular-mapping mitogenomes of discobids (fig. 3). Heterolobosean GC-skew trajectories and the cumulative GC-skew curves map simply to organismal phylogeny (fig. 3). Cumulative GC-skew of the *P. flabellatum* mitogenome shows a homogenous and positive skew curve along the sequence, resembling the situation in *A. kona*, *Naegleria* spp., and *S. lipophora*. A similar pattern is seen in some jacobids; *J. bahamiensis* and *O. amphimena*, as well as *Tsukubamonas* (*T. globosa*) (fig. 3B; Burger et al. 2013). This linear skewness was also observed in the mitogenomes of the glaucophyte *Cyanoptyche gloeocystis* (Jackson and Reyes-Prieto 2014), and in the yeast *Candida glabrata*, which this latter mitogenome was experimentally shown to replicate by a rolling circle (Maleszka et al. 1991; Burger et al. 2013). This skewness is also usually found in circular plasmids, known to replicate by unidirectional replication of the rolling circle (RCR) (Arakawa et al. 2009).

However, *R. americana* and *H. aroides* display prominent bimodal GC-skew curves—more distinctly in *R. americana*—which is characteristic of the classical bidirectional theta mode (fig. 3; Burger et al. 2013). The GC-skew graphs of the remaining discobids (*P. kirbyi*, BB2, *A. godoyi*, *S. ecuadoriensis*) mitogenomes are insufficient to draw meaningful inferences about the replication mechanism (fig. 3; Burger et al. 2013). Interestingly, similar GC trajectories have been observed in bacterial genomes (Saha et al. 2014) and have been shown to be induced by genomic rearrangement types mixing leading and lagging sequences that influence trends in local base usage.
as present among both Tetramitia and Pharyngomonada (Petrü et al. 2018, see fig. 4), having been overlooked in the original description of the Pharyngomonas kirbyi mitogenome (Yang et al. 2017). We reconstructed single-gene phylogenies to examine the evolutionary history of these five genes among the eukaryotes, none of which showed strong evidence against a common origin of mitochondrial forms across eukaryotes (supplementary figs. S6–S10, Supplementary Material online). The presence of ccmA, C, F, rps4, and tatC genes in some but not all mitogenomes in various distantly related eukaryote lineages suggests that these genes were lost at several independent points during eukaryotic/mitogenome evolution.

Electron Transport Chain Components in Pleurostomum flabellatum

The typical mitochondrial of aerobic eukaryotes contains the electron transport chain (ETC; Complexes I–IV) and an ATP-synthase (Complex V) within the inner mitochondrial
FIG. 4.—A comparison of mitogenome gene contents of three lineages of Discoba (Heterolobosea, Euglenozoa, and Tsukubamonas). (A) Venn diagram representation of the variable portions of the protein-coding gene repertoires of available heterolobosean mitogenomes, that is, excluding the protein-coding genes present in all Heteroloboseid mitogenomes (nad1-6, nad, 8,11, nad4L, cytB, cox1,2,3, atp1,3,6,8,9, rps13,14, rpl5,6,14). For each taxon, the current sets of functionally assignable CDSs in each species mitogenomes are shown in parentheses. The gene rpl32 (in red) was lost in Pharyngomonas kirbyi. The gene tatC (in red) was identified in Pharyngomonas kirbyi by Petrú et al. (2018), however, there is no evidence of its presence in the heterolobosean sp. BB2. (B) Venn diagram representing the minimal inferred gene content of the mitogenome of the last common heterolobosean ancestor, along with representative euglenozoans and Tsukubamonas. The gene in red (rps12) is present in Trypanosoma but absent from Diplonema. (C) Inferred gene losses are shown on tree branches in red. Numbers in square boxes indicate the minimum inferred number of genes in ancestral species. Parentheses after genus names show the number of genes retained by extant mitogenomes. Gene loss events were predicted on the phylogeny of discobids, as extracted from figure 2.
**FIG. 5.**—Mitochondria cristae morphotype and MICOS proteins distribution in the Discoba. (A) TEM micrographs of *P. flabellatum* (N, nucleus; M, mitochondrion; scale bar: 500-nm). (B) Distribution of MICOS subunits across the Discoba with mitochondria cristae morphotype. The presence of certain MICOS components (Mic10, Mic60, Mic19) are shown by gray rectangles. Other lineages examined in this analysis and characterized by "EST data" had only EST data sets available. The question mark indicates unavailable/uncomplete nuclear genome data.
membrane (Gray 2012). Some of the proteins that make up this complex system are invariably encoded on the mitogenome in such aerobes, even when the mitogenome is unusually small and the ETC is reduced (Flegontov et al. 2015). The P. flabellatum and Naegleria mitogenomes contain the same genes that encode electron transport chain components (supplementary table S2, Supplementary Material online), indicating that Pleurostomum mitochondria must be either aerobic or microaerophilic, contrary to some suggestions in the literature (Park et al. 2007; Pánek et al. 2014; Hanousková et al. 2019).

**Mic60 Suggests the Potential for Mitochondrial Cristae Formation**

We identified two genes coding putative homologs of the highly conserved Mic60 protein (GenBank accession numbers: MW019459 and MW019460) from the draft nuclear genome as well as transcriptome data of P. flabellatum (unpublished). Mic60 protein sequences of P. flabellatum contain the mitoflin domain architecture found in N. gruberi Mic60 (Naegru_Mic60_64657 XP_002683319.1, see supplementary fig. S11, Supplementary Material online). The transcript sequences include an extension of approximately 800-bp and poly-A tail in 5’- and 3’-UTR regions, respectively, suggesting active function of Mic60-Mic60 genes. This implies that P. flabellatum, which lacks typical cristae, nonetheless shares at least the Mic60 subunit with cristae-bearing heteroloboseans.

The evolutionary history of mitochondria in eukaryotes generally showed a correlation between the occurrence of cristate mitochondria and the presence encoding MICOS subunits, although there are some exceptions likely due to the lack of high-quality genome data (Muñoz-Gómez, Slamovits, Dacks, Baier, et al. 2015; Huynen et al. 2016). Mic60/Mic60 has been inherited from a putative Mic60 and two Mic10 paralogs (Eichenberger et al. 2013; Park et al. 2007) and may retain only Mic60, hinting at least two putative, but actively transcribed Mic60/Mic60 homologs. These findings suggest that the Mic60 subunit alone is not enough to form the cristae junction. We conclude that P. flabellatum has an unusual mitochondrial structure due to its tolerance to low oxygen conditions, which prevail under extreme hypersalinity. Further analysis of the transcriptome and nuclear genome of P. flabellatum will help elucidate the functions of this organelle within the cell, and these examinations may provide useful insights into generating energy under suboxic conditions.

**Conclusions**

Comparison of the complete P. flabellatum mitogenomic sequence with other discobid taxa shows that the proportion of coding regions and the AT content are comparable to other heteroloboseans. There is coherence between mitochondrial gene loss and phylogenetic position in Heterolobosea. For instance, the P. flabellatum and its close relative Naegleria mitogenomes share considerable similarity in terms of mitogenome content, architecture, and syntenry. Heterolobosean mitochondrion-encoded genes show minimal variation in codon usage, including the two halophilic heteroloboseids (i.e., Pharyngomonas and Pleurostomum). P. flabellatum cumulative GC-skew reveals a positive skew plot along with other heteroloboseans except for early-diverging P. kirbyi and Heterolobosei sp. BB2.

The mitogenome P. flabellatum includes the same genes that encode components of the electron transport chain as the Naegleria mitogenome, indicating that P. flabellatum mitochondria are either aerobic or microaerophilic. Although no cristae structure has been shown by electron microscopy, the draft nuclear genome sequence of P. flabellatum revealed at least two putative, but actively transcribed Mic60/Mic60 homologs. These findings suggest that the Mic60 subunit alone is not enough to form the cristae junction. We conclude that P. flabellatum has an unusual mitochondrial structure due to its tolerance to low oxygen conditions, which prevail under extreme hypersalinity. Further analysis of the transcriptome and nuclear genome of P. flabellatum will help elucidate the functions of this organelle within the cell, and these examinations may provide useful insights into generating energy under suboxic conditions.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.

**Acknowledgments**

This study was supported by the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (MOF) (20180430), the National Research Foundation of Korea (NRF-2017R1A2B3001923), the Next-generation BioGreen21 Program (PJ01389003) from the RDA ( Rural Development Administration), Korea to Hwan Su Yoon, and the National Research Foundation of Korea (NRF-2016R1A6A1A05011910 and 2019R1A2C2002379) to Jong Soo Park.
Data Availability
The mitogenome sequence has been deposited at GenBank under the accession number MT843578, as well as two two Mic60/Mic60-like sequences (accession numbers: MW019459 and MW019460).

Literature Cited


Associate editor: Geoff McFadden

Mitochondrial Genome of the Extreme Halophile Pleurostomum flabellatum