Turkish Journal of Zoology

Volume 49 | Number 4 Article 2

7-17-2025

Comparative mitogenomics of Clitellata reveals mitogenome organization can affect the mode of phylogeny

DENİZ MERCAN

AUSTIN HORENKAMP

MERVE NUR AYDEMİR

HABEŞ BİLAL AYDEMİR

KURT NEUBIG

See next page for additional authors

Recommended Citation

MERCAN, D, HORENKAMP, A, AYDEMİR, M. N, AYDEMİR, H. B, NEUBIG, K, ARSLAN, N, & ANDERSON, F. E (2025). Comparative mitogenomics of Clitellata reveals mitogenome organization can affect the mode of phylogeny. *Turkish Journal of Zoology 49* (4): 197-206. https://doi.org/10.55730/1300-0179.3225

Comparative mitogenomics of Clitellata reveals mitogenome organization can affect the mode of phylogeny

Authors
DENİZ MERCAN, AUSTIN HORENKAMP, MERVE NUR AYDEMİR, HABEŞ BİLAL AYDEMİR, KURT NEUBIG, NAİME ARSLAN, and FRANK E. ANDERSON

TÜBİTAK

Turkish Journal of Zoology

http://journals.tubitak.gov.tr/zoology/

Research Article

Turk J Zool
(2025) 49: 197-206
© TÜBİTAK
doi:10.55730/1300-0179.3225

Comparative mitogenomics of Clitellata reveals mitogenome organization can affect the mode of phylogeny

Deniz MERCAN[1,*], Austin HORENKAMP[2], Merve Nur AYDEMİR[3], Habeş Bilal AYDEMİR[4], Kurt NEUBIG[5], Naime ARSLAN[1], Frank E. ANDERSON[5]

[1]Department of Biology, Faculty of Sciences, Eskisehir Osmangazi University, Eskişehir, Turkiye
[2]School of Science, Faculty of Health and Environmental Sciences, Auckland University of Technology, Auckland, New Zealand
[3]Department of Genomics, Faculty of Aquatic Sciences, İstanbul University, İstanbul, Turkiye
[4]Department of Molecular Biology and Genetics, Faculty of Science and Lecture, Tokat Gaziosmanpaşa University, Tokat, Turkiye
[5]School of Biological Sciences, Southern Illinois University, Carbondale, USA

Received: 31.10.2024 • **Accepted/Published Online:** 13.05.2025 • **Final Version:** 17.07.2025

Abstract: Clitellata is a major clade of Annelida that includes almost all freshwater and terrestrial annelids as well as many marine species. Aquatic oligochaetes include the union of the two multispecies families: Naididae and Tubificidae. There has been much debate regarding the phylogeny and classification of the cosmopolitan and diverse Naididae. In this study, the total mitogenome sequences of *Potamothrix hammoniensis* (Naididae: Tubificinae), *Stylaria fossularis*, *Stylaria lacustris*, *Chaetogaster diaphanus*, *Chaetogaster diaphanus* sp. B, and *Slavina appendiculata* (Naididae: Naidinae) were sequenced and characterized and can be accessed from NCBI under accession numbers OQ654101, PP909790, PP909791, PP909792, PP909793, and PP893275, respectively. For comparative analysis, 101 clitellate mitogenome sequences (from 70 oligochaetes and 31 hirudineans) were downloaded from NCBI. Comparative analysis showed that nine anticodon-codon interactions do not have exact matches—a pattern consistent with the restriction of transcription by mito-tRNAs. Within Clitellata, nine different gene order patterns were determined, and four of them were found in Naididae. This suggests that mitochondrial gene order shows a cumulative effect in the evolution of mitogenome sequences. Based on phylogenetic analysis, it appears that stronger support for Clitellata can be achieved with additional sequences. Additional sequences sampled from several other families could be used for even more stability.

Key words: Comparative mitogenomics, mtDNA, Clitellata phylogeny, Naididae

1. Introduction

Clitellata is a major clade of Annelida that includes freshwater, terrestrial, and marine annelids (Erséus et al., 2020). Molecular analysis has shown that clitellates are a clade within the paraphyletic Polychaeta (Erséus et al., 2020; Lewin et al., 2024). Based on morphological and genetic evidence, it was assumed that Haplotaxidae and/or branchiobdellidans are members of Clitellata and are nonmonophyletic (Erséus et al., 2020). Molecular evidence has suggested that leeches may have descended from a lumbriculid-like ancestor (Erséus, 2005). Major classification changes in the aquatic oligochaetes over the past few years include the union of two multispecies families, Naididae and Tubificidae, into one family after it was determined that the former was polyphyletic and the latter was paraphyletic (Envall et al., 2006; Erséus and Gustavsson, 2002). There has been much discussion over the years regarding the phylogeny and classification of the cosmopolitan and diverse Naididae sensu lato (i.e. including Tubificidae and Opistocystidae). The genus *Potamothrix* Vejdovský et Mrázek, 1903, is a monophyletic, well-defined naidid group in the subfamily Tubificinae (Erséus et al., 2020; Timm, 2013). Twenty-seven valid species of *Potamothrix* have been described to date. One of them is the Ponto-Caspian species *Potamothrix hammoniensis* Michaelsen, 1901. This species has a significant impact on the intensity and direction of material exchange in the water/sediment interface through bioturbation (Geta et al., 2004; Svensson et al., 2001; Żbikowski et al., 2018). It is a crucial component of the food webs of benthic fish and predatory invertebrates (Thorhauge, 1975).

The genus *Stylaria* is monophyletic and currently considered to comprise two species. Although cryptic genetic diversity in *Stylaria fossularis* has been found, *St. lacustris* appears to lack cryptic genetic diversity (Erséus et al., 2017; Horenkamp, 2023). *Slavina appendiculata* is

* Correspondence: deniss-kara@hotmail.com

one of five species in the genus *Slavina* and is a freshwater naidid commonly found in coarse gravel, sand, and decaying plant material in shallow waters. *Chaetogaster diaphanus* is a species of carnivorous annelid found in slow-moving streams, usually attached to plant material. Recent research on carnivory amongst *Chaetogaster* has shown the possibility of cryptic diversity in *Chaetogaster*, including within *C. diaphanus* (Mack et al., 2023).

The mitogenome is a compact molecule with its own transcription and translation system. There are 13 protein-coding genes (PCGs) that are essential for oxidative phosphorylation (OXPHOS). The genes are transcribed and translated by the two mitogenomic rRNA subunits (rrnS and rrnL) and the 22 tRNAs (two isoacceptor tRNAs for *trnS* and *trnL* and one for each of the remaining) (Boore, 1999). The mitogenome also has highly diverse noncoding regulatory regions, one of which is an A+T-rich region, and the other is a set of intergenic spacers (IGSs) containing transcription breakpoints (Roberti et al., 2003). The unique character of the mitogenome is shaped by both the nuclear genome-encoded transcription factors and endosymbiotic origin, and it is maintained by highly specific mutation and selection processes (Andersson et al., 2003). Because of these effects, the mitogenome contains different rates of variation within (*cox1* PCG vs. *atp8* PCG) and between (PCGs vs. tRNAs) genomic compartments (Aydemir and Korkmaz, 2020).

Clitellata is represented in the NCBI GenBank database by 141 mitogenomes (released before April 2023). The majority of the clitellate samples in the NCBI database belong to terrestrial species with only four aquatic oligochaete mitogenomes (2.83%) represented in the NCBI database. *P. hammoniensis*, *St. fossularis*, *St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata* mitogenome data were added by us to the NCBI database for this study. The mitogenome data from 101 clitellates were downloaded and used for the comparison of mitogenomic characters and for phylogeny construction, by removing unverified INSDC data of representative families in 141 Clitellata mitogenomes. In particular, mitochondrial gene order and codon-anticodon interactions were evaluated.

2. Materials and methods

2.1. Specimen collection and DNA extraction

Representatives of *P. hammoniensis*, *St. fossularis*, *St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata* were collected from Lake Eğirdir in Isparta (Türkiye); Seven Mile Slough, Sacramento, California (USA); Lången Lake, Alingsås, (Sweden); Åkers Kanal, Uppland, (Sweden); and Ship Creek Anchorage, AK, (USA), respectively. The samples of *P. hammoniensis* were placed in 100% ethanol during collection and preserved at −20 °C in the NA Hydrobiology Laboratory at the Eskişehir Osmangazi University. Before DNA extraction, selected whole organisms were identified at the species level using morphological characteristics via various species identification keys (Reynolds, 1973; Smith, 1987; Timm, 2013). Total genomic DNA was extracted from the whole body of *P. hammoniensis* using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions with minor changes. The quality and quantity of DNA was assessed using a MN-917 MaestroNano spectrophotometer (Maestrogen Inc.), and the total extracted DNA was sequenced as 150 bp paired-end reads using the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) by Macrogen Inc. (Amsterdam, Netherlands) to generate 65 million high-quality reads (15 GB of raw data). Samples of the remaining five species were stored in 70% EtOH at −20 °C. Total genomic DNA of *St. fossularis*, *St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata* were extracted using the posterior part of the body following Neubig et al. (2012), and the quality and quantity of each DNA isolate was determined using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). High-throughput Illumina shotgun sequencing was performed on the total genomic DNA at the Roy J. Carver Biotechnology Center (Urbana, IL, USA).

2.2. Mitogenome assembly

The quality of raw next-generation sequencing (NGS) reads obtained for *P. hammoniensis* was assessed with FastQC version 0.11.4[1]. High-quality NGS reads were assembled using Geneious R9 using the map to reference option (Kearse et al., 2012). *Limnodrilus hoffmeisteri* (MW732144), *Tubifex tubifex* (MW690579), and *Nais communis* (MW770354) were used as reference mitogenomes for the assembly of the total mitogenome of *P. hammoniensis*. After consensus sequences were generated, BLAST searches were conducted. The assemblies with a high score match with naidid mitogenomes were fully annotated to provide a final assembly.

2.2.1. Mitogenome organization and annotation

Total mitogenome organization and characterization were determined using MITOS (Bernt et al., 2013). The locations and nucleotide boundaries of the PCGs were identified using ORF Finder[2] using the invertebrate mitochondrial genetic code. The nucleotide and amino acid sequences of all ORFs found via ORF Finder were verified in BLASTn

[1]Andrews S (2010). FastQC: a quality control tool for high throughput sequence data [online]. Website http://www.bioinformatics.babraham.ac.uk/projects/fastqc [accessed 20 March 2023].

[2]National Center for Biotechnology Information (2023). Open Reading Frame Finder [online]. Website: https://www.ncbi.nlm.nih.gov/orffinder [accessed 10 March 2023].

and BLASTp, respectively. Conserved domains of PCGs were determined using BLAST CD-search (Marchler-Bauer et al., 2017). Identification of tRNA genes was conducted with the ARWEN (Laslett and Canbäck, 2008) web server and recognized manually as sequences having the appropriate anticodon and typical cloverleaf secondary structure. rRNA genes were identified with alignment scores of closely related species and checked for their capability to fold into the correct secondary structures with conserved domain topology. Comparative secondary structures of rrnL and rrnS were produced through homology modeling using the published secondary structures of *Whitmania laevis* as templates using XRNA version 1.2.0b (Ye et al., 2015).

In the whole mitogenome, repeated sequence motifs were identified using Tandem Repeat Finder (Benson, 1999) and putative secondary structures were constructed using MFOLD (Zuker, 2003).

The mitogenome and mitogenomic compartments of *P. hammoniensis* were aligned with other clitellate species in the NCBI database to ensure that the annotation was performed correctly, with each gene checked via BLAST searches.

2.3. Comparative analysis of clitellate mitogenomes

The complete mitogenome sequences of *P. hammoniensis*, *St. fossularis*, *St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata* obtained from this study, and 101 clitellate mitogenome sequences from NCBI were used for comparative analyses (Table S1). Nucleotide and amino acid compositions of genes and mitogenomes were calculated for each species.

For comparison of PCGs, the relative synonymous codon usage (RSCU) value of genes was calculated with MEGA 7 (Kumar et al., 2016). To estimate the selective forces on the PCGs, the ratio of the nonsynonymous substitution rate (dN) to the synonymous substitution rate (dS) was calculated with DNAsp version 6 (Rozas et al., 2017). tRNA genes were compared with the base of INUC% (percentage of identical nucleotides) values, and the conservation level of preferred anticodons was determined.

Gene rearrangement events in Clitellata were evaluated using CREx with a polychaete, *Nereis zonata* (NC_053360), as a reference (Bernt et al., 2007).

A mitogenome-based phylogeny for Clitellata was constructed with a dataset of 13 PCGs using an echiuroid (*Urechis caupo* (Urechidae)) as an outgroup. Individually translation-aligned PCGs were concatenated with MAFFT using the L-INS-i algorithm in PhyloSuite version 1.2.3 (Katoh et al., 2005). The optimal nucleotide substitution model was estimated by PartitionFinder version 1.1.1 using the Bayesian information criterion (BIC) and the greedy algorithm with branch lengths estimated as unlinked (Lanfear et al., 2012). The best partitioning schemes and associated substitution model (GTR+G+I) were used in all phylogenetic estimations. Maximum likelihood (ML) was used to infer phylogenetic trees using IQTREE, applying the proposed substitution model with 1000 bootstrap replicates (Stamatakis, 2006). Visualization of the trees was performed using FigTree version 1.4.2 (Cummings, 2004).

3. Results and discussion

3.1. Mitogenome characterization of Naididae

The *P. hammoniensis* mitogenome assembly was 14,881 bp long and was completely annotated. The partial mitogenomes of *St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, *Sl. appendiculata*, and *St. fossularis* were 14,813; 15,724; 15,694; 15,699, and 15,719 bp long, respectively. All genes are dispersed and transcribed in the heavy chain, like other annelids (Oceguera-Figueroa et al., 2016; Weigert et al., 2016) (Figure 1; Table S2). The mitogenome sequences of *P. hammoniensis*, *St. fossularis*, *St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata* can be accessed from NCBI under accession numbers OQ654101, PP909790, PP909791, PP909792, PP909793 and PP893275, respectively.

As seen in other invertebrates, these mitogenomes showed an A+T bias of 67.8– 70.9%). Several adjacent genes overlapped from two (trnY-trnG, trnM-rrnS, trnA-trnS2 for all species) to 29 (trnQ-*nd6*, *nd2-cox1* for *C. diaphanus*) nucleotides (Table S2). In the mitogenomes, the overall lengths of PCGs, tRNAs, and rRNAs were 11928, 1389, and 1982 bp, respectively. Conserved *atp6-atp8* overlap was missing in this mitogenome, and accelerated evolution seems to lead to both differentiation of gene order and accumulation of intergenic regions. The conserved transcriptional sequence WHWGHTW identified in Ecdysozoa was detected in the *P. hammoniensis* mitogenome, 14 bp upstream of the trnT gene (Aydemir et al., 2023).

All PCGs were found to start with the standard ATG (methionine) codon, except for *nd6* (ATT) in *P. hammoniensis*; *nd1* (GTG) in *St. lacustris*; *nd2* (CTT), *nd1* (GTG), and *nd3* (ATT) in *St. fossularis*; *nd2* (TTT), *nd4L* (ATA), and *nd1* (GTG) in *Sl. appendiculata*; *nd3* (ATA) and *nd1* (GTG) in *C. diaphanus* and *nd2* (ATC), *nd1* (GTG), *nd3* (ATA), and *atp8* (TTA) in *C. diaphanus* sp. B. While *cox1*, *cox3*, *cytB*, *atp6*, *nd5*, *nd4L*, and *nd4* genes had complete TAA triplets as stop codons, *cox2*, *atp8*, *nd6*, *nd1*, *nd3*, and *nd2* genes had incomplete T-stop codons that are completed via polyadenylation processes. Like the total mitogenome, all PCGs had negative AT- and negative GC-skew values. From the ORF Finder analysis, the *P. hammoniensis* mitogenome is capable of coding 780 different ORFs that start with any sense codon and are longer than 30 nucleotides (Supplementary file

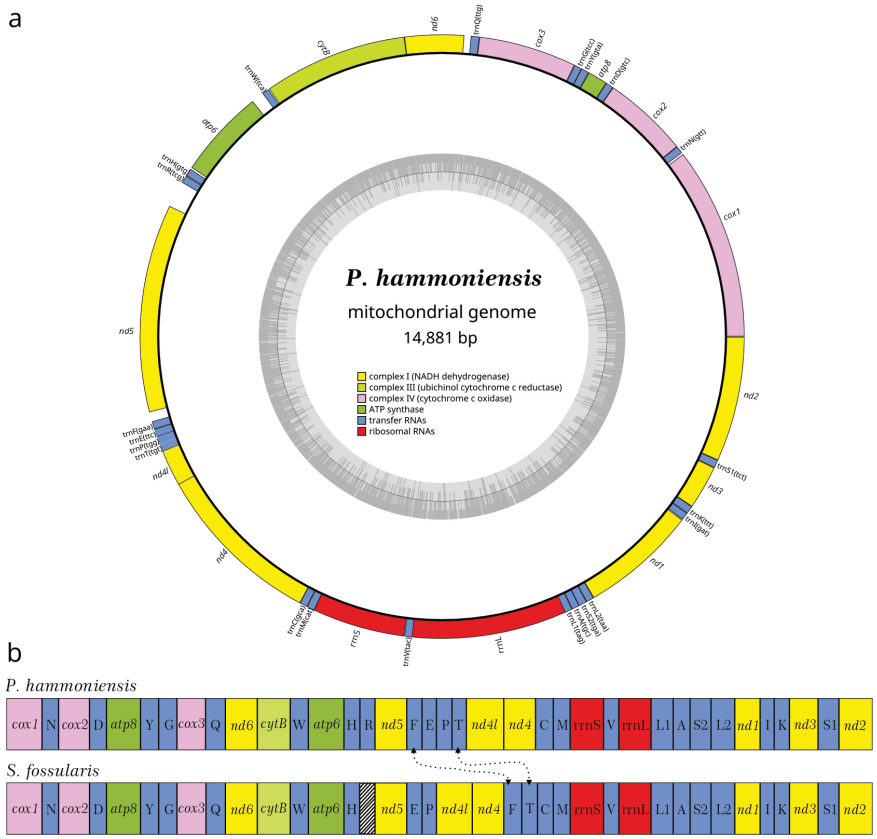

Figure 1. a. Mitogenome organization of *P. hammoniensis* drawn using OGDraw (Greiner et al., 2019). All genes are encoded in heavy (J-) strand. Gaps between genes indicate IGSs. **b.** Different gene order patterns in Naididae mitogenomes. Two tRNA genes (trnF and trnP) were rearranged with TDRL events in Naididae (between Tubificinae and Naidinae) mitogenomes. *St. fossularis* has the same mitochondrial gene order pattern with other Naidinae species sequenced in this study (*St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata*).

S1). Interestingly, 445 of these 780 ORFs are potentially encoded in the light chain, which does not code any known mitochondrial gene. Based on BLAST searches, 39 of 780 potential ORFs aligned with protein sequences in the nonredundant protein database (Supplementary file S2). Thirteen of the 39 were formal mitochondrial PCGs (*atp6*, *atp8*, *cox1*, *cox2*, *cox3*, *cytB*, *nd1*, *nd2*, *nd3*, *nd4*, *nd4L*, *nd5*, *nd6*) and five of the 39 were different framed forms of *cox1* and *cox2* sequences. Eighteen of the 39 aligned amino acid sequences were potentially coded in the light chain, eight of which seemed to be an alternative form of some *nd* genes (*nd1*, *nd3*, *nd4* and *nd5*) in different organisms (Supplementary file S2). Similarly, 726, 727, 757, 770 and 736 potential ORF sequences were detected in the *St. fossularis, St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata* mitogenomes, respectively.

The average length of tRNAs was 63.13 bp and varied from 56 bp in trnT to 69 bp in trnQ (Table S2). All tRNAs constructed had a conserved clover-leaf structure except trnS1 and trnT, both of which had a DHU-replacement (Figure S1). A reduced secondary structure pattern is common in invertebrates for trnS1, but trnT reduction is unique to *P. hammoniensis* in Clitellata. trnE, trnP, trnS1, and trnS2 tRNAs had negative AT- and negative GC- skew values. Unlike the total mitogenome and PCGs, tRNA genes tended to have at least one positive skewness value (Table S2).

The rrnS and rrnL genes were 737 and 1245 bp long, respectively, and they had positive AT- and negative GC-skew values (Table S2). Both rRNA genes had conserved putative secondary structures (Figures 2a and 2b). The putative secondary structures of rrnS and rrnL genes comprised three domains with 30 helices and five domains (domain III is absent in invertebrates) with 49 helices, respectively.

Noncoding regions of the mitogenome of *P. hammoniensis* were dispersed between 11 consecutive gene clusters and had a total length of 472 bp, varying

from 2 (trnH-trnR and rrnS-trnV) to 208 (trnR-*nd5*) bp long. Intergenic regions did not exhibit significant matches in the BLAST search with low query coverage values (e > 0.01). In Clitellata, the putative control region is generally located in the *atp6-nd5* gene cluster (trnW-*atp6*-trnH-trnR-*nd5*-trnF), and the *P. hammoniensis* mitogenome had 39-bp-long total intergenic sequences in this region (Table S2). In the *nd5*-trnF intergenic region, transcription polarity was dramatically changed. This region probably has the replication origin that can construct a stable secondary structure with significant ΔG value (−13.60 kcal/mol) (Figure S2).

3.2. Comparative analysis of Clitellata mitogenomes

The size of complete Clitellata mitogenomes varied from 14,407 bp in *Erpobdella octoculata* (Hirudinea: Erpobdellidae) to 18,528 bp in *Acanthobdella peledina* (Hirudinea: Acanthobdellidae), with an average of 15,096 bp in length. Total mitogenomes of Clitellata species were biased to A+T and generally have negative AT- and GC-skew values. Exceptionally, 13 species from Hirudinea and two species from Oligochaeta had moderately positive GC-skew values, with most of Oligochaeta (32 of 46) mildly biased to positive AT- skewness (Table S3). Average nucleotide compositions were as follows: T = 35.3%, A = 33.4%, C = 17.5%, and G = 13.8%.

Total PCG size varied from 10,979 bp (*Haemadipsa tianmushana*) to 11,244 bp (*Theromyzon tessulatum*), with an average of 11,102.1 bp long. Like the total mitogenome data, total PCGs of Clitellata species were biased to A+T and generally had negative AT- and GC-skew values.

Amino acid preferences of Clitellata were represented mostly by hydrophobic amino acids as expected with an A+T nucleotide bias: leucine (14.75%), serine (9.80%), isoleucine (8.86%), and methionine (8.37%). Cysteine (1.04%), arginine (1.60%), and glutamine (1.74%) amino acid preferences had the lowest levels (Table S4).

RSCU values from whole sampling are given in Figure 3. All stop codons were removed to avoid the bias of incomplete stop codons. All 4-fold degenerate codons preferred A in the third codon position. All 2-fold degenerate codons were biased to keep A or T nucleotides consistent with the A+T bias of the total mitogenome. On the contrary, anticodons of complementary tRNAs are conserved among Clitellata and are biased to G and T nucleotides in the first position. Anticodons of 9 tRNAs start with G or C nucleotides and therefore some interactions do not exhibit exact matches between the tRNA anticodon and the mRNA codon, as previously documented in the trnP interaction in hymenopteran insects (Aydemir and Korkmaz, 2020). The situation in Clitellata appears more dramatic. This pattern is explained by the fact that the transcription of mitogenomes can be partially restricted by mitochondrial tRNAs (Shtolz and Mishmar, 2023). On the other hand, this is one reason why the most conserved interaction position is the second codon (anticodon) position (Błażej et al., 2018).

The average rrnS gene length is 763.4 bp and ranges from 688 (*Hirudinaria manillensis* and *Poecilobdella manillensis*) to 801 bp (*Perionyx excavatus*). Average nucleotide compositions of rrnS are as follows: T = 29.7%,

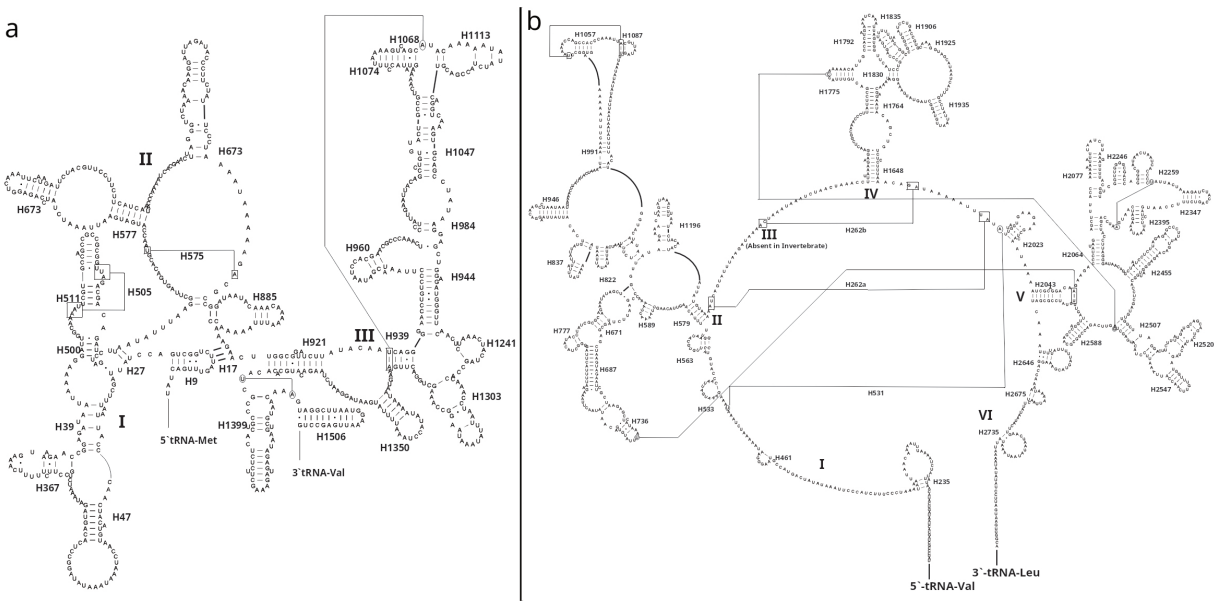

Figure 2. The putative secondary structures of (a) rrnS and (b) rrnL of *P. hammoniensis*. Structures were predicted and helixes were numbered with reference *W. laevis* secondary structures (Ye et al., 2015). Roman numbers refer to domain names. Dashes indicate Watson-Crick base pairing and dots indicate G-U base pairing.

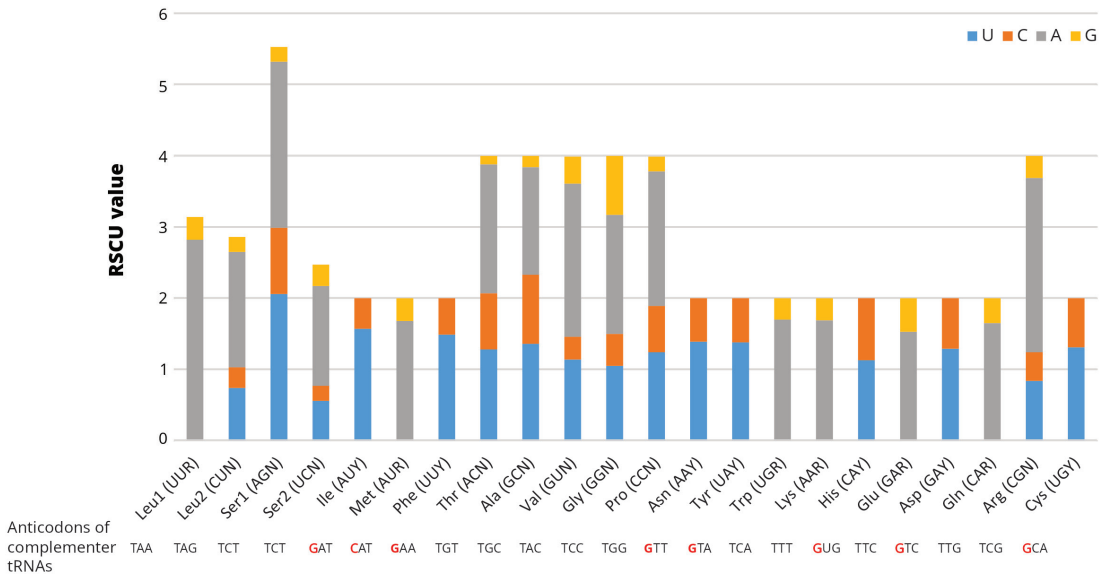

Figure 3. RSCU values and preferred anticodon pattern among Clitellata. Colors indicate preferred nucleotides at the 3rd codon position. Red-marked tRNA anticodons emphasize that do not exhibit an exact-match between related tRNA anticodon and preferred mRNA codons.

A = 38.5%, C = 16.1%, and G = 15.7%. The rrnS gene has a positive AT-skew (except for *Poecilobdella javanica*) but variable GC-skew (32 of 78 have positive values, and 45 of 78 have negative values) (Table S5). The size of the rrnL gene varied from 1027 bp in *Enchytraeus irregularis* to 12,812 bp in *Amynthas deogyusanensis,* with an average of 1206.4 bp in length. Average nucleotide compositions of rrnL are calculated as T = 31.0%, A = 39.8%, C = 15.1%, and G = 14.1% (Table S6). As seen with the rrnS gene, the rrnL gene of Clitellata generally has positive AT- and variable GC-skew values.

The length of tRNA genes varied from 41 (trnR of *W. pigra*) to 83 (trnS of some Hirudinea organisms) with an average of 63.4 bp in length (Table S7). Average nucleotide compositions of tRNAs are calculated as follows: T = 33.3%, A = 35.9%, C = 14.7%, and G = 16.1%.

The results of dN/dS calculations indicated that all mitochondrial PCGs are subjected to negative selection (dN/dS < 0.5) in Clitellata (Figure 4). Mutations that may occur in the *cox1* gene are strongly and rapidly purified and following that, the *cox3*, *cytB*, *nd1* and *cox2* genes are subjected to densified negative selection. It is expected for *cox* and *cytB* genes because of their critical roles in OXPHOS (Castellana et al., 2011), but purifying selection on the *nd1* gene is unexpected, especially despite the relaxation of negative selection observed in other *nd* subunits. The most likely reason may be that the *nd1* gene product functions as a binding site for the hydrophobic quinone during OXPHOS. This is essential as it interfaces between hydrophobic and hydrophilic subunits, and it is crucial to eliminate mutations in the *nd1* gene to stabilize this interaction (Garvin et al., 2015). On the other hand, the *atp8*, *nd6*, *nd2*, and *nd4L* genes had the highest dN/dS values and seem to have evolved relatively rapidly among mitochondrial PCGs (Figure 4).

From the CREx analysis, nine different gene order patterns were observed within Clitellata (Figure 5). A minimum of three and a maximum of 20 rearrangement breakpoints were detected between the reference mitogenome and the clitellate mitogenomes (Table S8). Accordingly, the transposition and tandem-duplication-random-loss (TDRL) event of trnG-trnY-*atp8*, the TDRL event of trnM-trnD gene clusters and the transposition of trnC were characterised. The most divergent gene order was determined in *Olavius algarvensis* and generally four different patterns were determined in Naididae. Except for Naididae (with five representatives) and Lumbriculidae (*Lumbriculus variegatus*), the oligochaete species shared a common mitochondrial gene order. Within Hirudinea, two unique gene orders were generally detected in two main groups (Erpobdelliformes + Oceanobdelliformes and Hirudiniformes) with a few exceptions. Interestingly, the mitochondrial gene order of *P. hammoniensis* was identical to *Theromyzon tessulatum* (Hirudinea: Glossiphoniidae) (Figure 5).

3.3. Phylogenetic analysis of Clitellata

In the phylogenetic analysis of the newly generated mitogenomes combined with all available clitellate mitogenomes on GenBank, *St. lacustris*, *St. fossularis*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata*

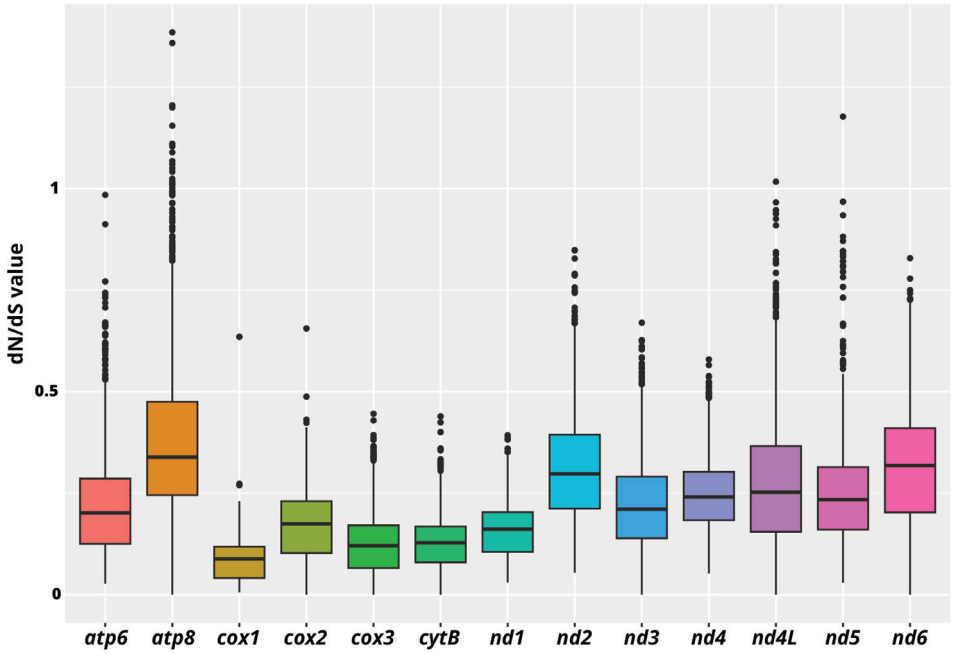

Figure 4. Comparison of dN/dS values of mitochondrial PCGs of Clitellata. For each gene, the bottom and top of the line indicates the minimum and maximum values respectively. All mitochondrial genes seem to be under negative purifying selection (dN/dS < 1).

clustered with the other naidines with high support (Figure 5). These findings are congruent with previous studies, with Naidinae forming a sister clade to Tubificinae, with near maximum support on every branch. In previous studies, within the species used in this study, Naididae is recovered as a sister clade to most other clitellate taxa, but the mitogenome tree recovers Enchytraeidae (represented by *E. irregularis*) in that position (Erséus et al., 2020). Mitogenomes from representatives of Phreodrilidae and Propappidae could aid in recovering a similar tree to the phylogenomic tree in Erseus et al., 2020. In this analysis, the leeches are recovered as the sister taxon to Lumbriculidae, which matches a previous phylogenomic study (Erséus et al., 2020). The unusual relationships recovered in our mitogenomic tree may reflect phylogenetic reality, but they could also be a result of long-branch attraction between Hirudinida and Naididae, which itself may be due in part to the relatively poor sampling of mitogenomes from Clitellata to date—there are no published mitogenomes available for several major lineages represented in Erséus et al. (2020). Levels of bootstrap support for many taxa, particularly within Megascolecidae, are low, so relationships between members of this family and other families are less trustworthy.

This is the most complete mitogenome study using all available clitellate mitogenomes to date, with six newly created mitogenomes for a total of 90 species. Previous studies of clitellate evolution examined the evolution of carnivory within the group, and the timing and nature of transitions among different habitat types (i.e. marine, freshwater, and terrestrial) (Erséus et al., 2020; Horn et al., 2019; Mack et al., 2023; Rousset et al., 2008). The mitogenome data here could aid future studies of clitellate phylogeny and evolution, but accurate phylogenetic inferences will likely require sampling of mitogenomes from additional clitellate lineages, including Branchiobdellidae, Capilloventridae, Haplotaxidae, Parvidrillidae, Phreodrillidae and Propappidae. To conduct a more thorough mitogenome study, additional specimens from the following clades should also be obtained: Alluroididae, Capilloventridae, Crassiclitellata, Haplotaxidae, Moniligastridae, Narapidae, Parvidrillidae, and Randiellidae, as well as additional members of Naididae.

4. Conclusion

The total mitogenomes of *P. hammoniensis*, *St. fossularis*, *St. lacustris*, *C. diaphanus*, *C. diaphanus* sp. B, and *Sl. appendiculata* were sequenced as representatives of Naididae and added to the NCBI database with accession numbers OQ654101, PP909790, PP909791, PP909792, PP909793 and PP893275, respectively. Based on comparative mitogenomics of Clitellata, some important findings are obtained. While Hirudinea mitogenomes show more variable results in terms of dN/dS ratios and RSCU values, other clitellate mitogenomes seem to be

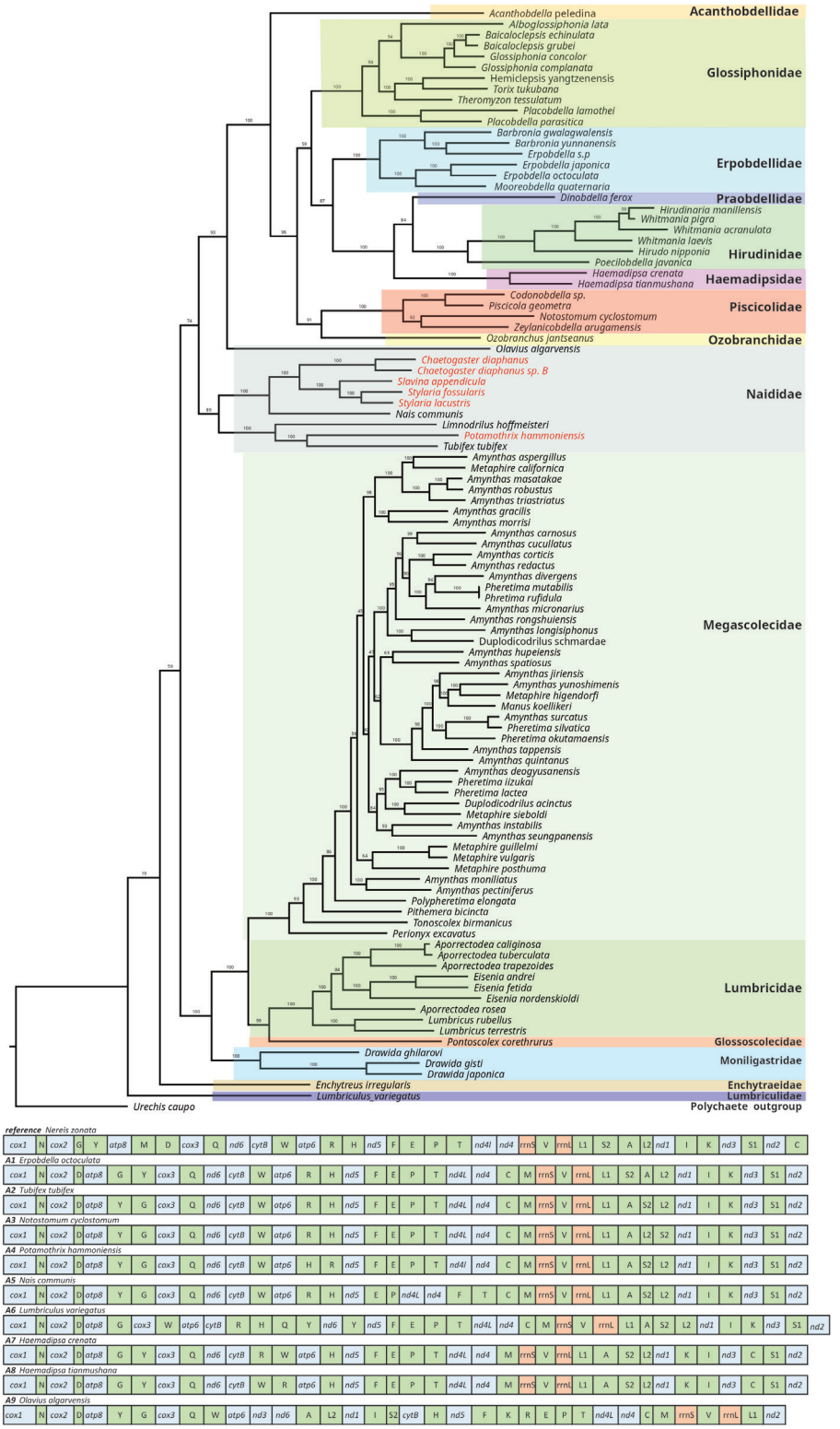

Figure 5. ML tree created in IQTREE displaying the phylogenetic relationships between different families within Clitellata. Bootstrap support values are shown at the base of each clade with >80% showing high support. Species names in red represent the newly sequenced clitellate mitogenomes. Nine different gene order patterns observed in Clitellata were given at the end of the figure.

more saturated. As a result of phylogenetic analysis of mitogenome sequences, not only point mutations but also mitochondrial gene orders reflect phylogenetic relationships within Clitellata. Gene rearrangement frequencies of Naididae appear to be quite diverse and the sampling needs to be expanded by the addition of new sequences.

Acknowledgments
This project was supported in part by NSF S-STEM 1564969 awarded to Kurt Neubig and Frank E. Anderson, which also provided a scholarship to support Austin Horenkamp. We would like to acknowledge Christer Erséus for providing the tissue samples of *Stylaria*, *Chaetogaster*, and *Slavina* used in the study.

References

Andersson SGE, Karlberg O, Canbäck B, Kurland CG (2003). On the origin of mitochondria: a genomics perspective. Philosophical Transactions of the Royal Society B Biological Sciences 358: 165-177. https://doi.org/10.1098/rstb.2002.1193

Aydemir MN, Aydemir HB, Budak M, Kızıltepe B, Çelebi MŞ et al. (2023). A novel, conserved and possibly functional motif "WHWGHTW" in mitochondrial transcription across Bilateria. Mitochondrion 68: 72-80. https://doi.org/10.1016/j.mito.2022.11.004

Aydemir MN, Korkmaz EM (2020). Comparative mitogenomics of Hymenoptera reveals evolutionary differences in structure and composition. International Journal of Biological Macromolecules 144: 460-472. https://doi.org/10.1016/j.ijbiomac.2019.12.135

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27 (2): 573-580. https://doi.org/10.1093/nar/27.2.573

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C et al. (2013). MITOS: improved *de novo* metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69 (2): 313-319. https://doi.org/10.1016/j.ympev.2012.08.023

Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M et al. (2007). CREx: inferring genomic rearrangements based on common intervals. Bioinformatics 23 (21): 2957-2958. https://doi.org/10.1093/bioinformatics/btm468

Błażej P, Wnętrzak M, Mackiewicz D, Mackiewicz P (2018). Optimization of the standard genetic code according to three codon positions using an evolutionary algorithm. PLoS ONE 13 (8): e0201715. https://doi.org/10.1371/journal.pone.0201715

Boore JL (1999). Animal mitochondrial genomes. Nucleic Acids Research 27 (8): 1767-1780. https://doi.org/10.1093/nar/27.8.1767

Castellana S, Vicario S, Saccone C (2011). Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein–coding genes. Genome Biology and Evolution 3: 1067-1079. https://doi.org/10.1093/gbe/evr040

Cummings MP (2004). FigTree. In: Dictionary of Bioinformatics and Computational Biology. Chichester, UK: John Wiley & Sons Ltd. https://doi.org/10.1002/9780471650126.dob0904

Envall I, Källersjö M, Erséus C (2006). Molecular evidence for the non-monophyletic status of Naidinae (Annelida, Clitellata, Tubificidae). Molecular Phylogenetics and Evolution 40 (2): 570-584. https://doi.org/10.1016/j.ympev.2006.03.021

Erséus C (2005). Phylogeny of oligochaetous Clitellata. Hydrobiologia 535: 357-372. https://doi.org/10.1007/s10750-004-4426-x

Erséus C, Envall I, De Wit P, Gustavsson LM (2017). Molecular data reveal a tropical freshwater origin of Naidinae (Annelida, Clitellata, Naididae). Molecular Phylogenetics and Evolution 115: 115-127. https://doi.org/10.1016/j.ympev.2017.07.016

Erséus C, Gustavsson L (2002). A proposal to regard the former family Naididae as a subfamily within Tubificidae (Annelida, Clitellata). Hydrobiologia 485: 253-256. https://doi.org/10.1023/A:1021366204441

Erséus C, Williams BW, Horn KM, Halanych KM, Santos SR et al. (2020). Phylogenomic analyses reveal a Palaeozoic radiation and support a freshwater origin for clitellate annelids. Zoologica Scripta 49 (5): 614-640. https://doi.org/10.1111/zsc.12426

Garvin MR, Bielawski JP, Sazanov LA, Gharrett AJ (2015). Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. Journal of Zoological Systematics and Evolutionary Research 53 (1): 1-17. https://doi.org/10.1111/jzs.12079

Geta R, Postolache C, Vădineanu A (2004). Ecological significance of nitrogen cycling by tubificid communities in shallow eutrophic lakes of the Danube Delta. Hydrobiologia 524: 193-202. https://doi.org/10.1023/B:HYDR.0000036133.92034.69

Greiner S, Lehwark P, Bock R (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47 (W1): W59-W64. https://doi.org/10.1093/nar/gkz238

Horenkamp AJ (2023). An analysis of species-level diversity of the freshwater clitellate genus *Stylaria* and an analysis of novel mitochondrial genomes for 5 Naidid Annelids. Master's thesis, Southern Illinois University Carbondale, Carbondale, IL, USA.

Horn KM, Williams BW, Erséus C, Halanych KM, Santos SR et al. (2019). Na+/K+-ATPase gene duplications in clitellate annelids are associated with freshwater colonization. Journal of Evolutionary Biology 32 (6): 580-591 https://doi.org/10.1111/jeb.13439

Katoh K, Kuma K, Toh H, Miyata T (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33 (2): 511-518. https://doi.org/10.1093/nar/gki198

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12): 1647-1649. https://doi.org/10.1093/bioinformatics/bts199

Kumar S, Stecher G, Tamura K (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33 (7): 1870-1874. https://doi.org/10.1093/molbev/msw054

Lanfear R, Calcott B, Ho SYW, Guindon S (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29 (6): 1695-1701. https://doi.org/10.1093/molbev/mss020

Laslett D, Canbäck B (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24 (2): 172-175. https://doi.org/10.1093/bioinformatics/btm573

Lewin TD, Liao IJ-Y, Luo Y-J (2024). Annelid comparative genomics and the evolution of massive lineage-specific genome rearrangement in bilaterians. Molecular Biology and Evolution 41(9): msae172. https://doi.org/10.1093/molbev/msae172

Mack JM, Klinth M, Martinsson S, Lu R, Stormer H et al. (2023). Cryptic carnivores: intercontinental sampling reveals extensive novel diversity in a genus of freshwater annelids. Molecular Phylogenetics and Evolution 182: 107748. https://doi.org/10.1016/j.ympev.2023.107748

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Research 45 (D1): D200-D203. https://doi.org/10.1093/nar/gkw1129

Neubig KM, Whitten WM, Williams NH, Blanco MA, Endara L et al. (2012). Generic recircumscriptions of Oncidiinae (Orchidaceae: Cymbidieae) based on maximum likelihood analysis of combined DNA datasets. Botanical Journal of the Linnean Society 168 (2): 117-146. https://doi.org/10.1111/j.1095-8339.2011.01194.x

Oceguera-Figueroa A, Manzano-Marín A, Kvist S, Moya A, Siddall ME, Latorre A (2016). Comparative mitogenomics of leeches (Annelida: Clitellata): genome conservation and *Placobdella*-specific *trnD* gene duplication. PLoS ONE 11 (5): e0155441. https://doi.org/10.1371/journal.pone.0155441

Reynolds JW (1973). Review of Aquatic Oligochaeta of the World, by Barry G. M. Jamieson and R. O. Brinkhurst]. Systematic Zoology 22 (2): 196-197. https://doi.org/10.2307/2412106

Roberti M, Loguercio Polosa P, Bruni F, Musicco C, Gadaleta MN, Cantatore P (2003). DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of *Drosophila melanogaster* mitochondrial DNA. Nucleic Acids Research 31: 1597-1604. https://doi.org/10.1093/nar/gkg272

Rousset V, Plaisance L, Erséus C, Siddall ME, Rouse GW (2008). Evolution of habitat preference in Clitellata (Annelida). Biological Journal of the Linnean Society 95 (3): 447-464. https://doi.org/10.1111/j.1095-8312.2008.01072.x

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology Evolution 34 (12): 3299-3302. https://doi.org/10.1093/molbev/msx248

Shtolz N, Mishmar D (2023). The metazoan landscape of mitochondrial DNA gene order and content is shaped by selection and affects mitochondrial transcription. Communications Biology 6: 93. https://doi.org/10.1038/s42003-023-04471-4

Smith ME (1987). Guide to the freshwater aquatic microdrile oligochaetes of North America. RO. Brinkhurst. Journal of the North American Benthological Society 6 (1): 78-79. https://doi.org/10.2307/1467528

Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22 (21): 2688-2690. https://doi.org/10.1093/bioinformatics/btl446

Svensson JM, Enrich-Prast A, Leonardson L (2001). Nitrification and denitrification in a eutrophic lake sediment bioturbated by oligochaetes. Aquatic Microbial Ecology 23 (2): 177-186. https://doi.org/10.3354/ame023177

Thorhauge F (1975). Reproduction of *Potamothrix hammoniensis* (Tubificidae, Oligochaeta) in Lake Esrom, Denmark. A field and laboratory study. Archiv für Hydrobiologie 76 (4): 449-474.

Timm T (2013). The genus *Potamothrix* (Annelida, Oligochaeta, Tubificidae): a literature review. Estonian Journal of Ecology 62 (2): 121-136. https://doi.org/10.3176/eco.2013.2.04

Weigert A, Golombek A, Gerth M, Schwarz F, Struck TH et al. (2016). Evolution of mitochondrial gene order in Annelida. Molecular Phylogenetics and. Evolution 94 (Part A): 196-206. https://doi.org/10.1016/j.ympev.2015.08.008

Ye F, Liu T, Zhu W, You P (2015). Complete mitochondrial genome of *Whitmania laevis* (Annelida, Hirudinea) and comparative analyses within *Whitmania* mitochondrial genomes. Belgian Journal of Zoology 145 (2): 114-128. https://doi.org/10.26496/bjz.2015.52

Żbikowski J, Mimier D, Żbikowska E (2018). Reproduction of *Potamothrix hammoniensis* (Oligochaeta) in shallow eutrophic lakes. Oceanological and Hydrobiological Studies 47 (2): 181-189. https://doi.org/10.1515/ohs-2018-0017

Zuker M (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31 (13): 3406-3415. https://doi.org/10.1093/nar/gkg595

Supplementary materials

Supplementary tables (Tables S1, S2, S3, S4, S5, S6, S7, and S8), supplementary figures (Figures S1 and S2), and supplementary files S1 and S2 are available at the following link:
https://aperta.ulakbim.gov.tr/record/286044