



Buried treasure in a public repository: Mining mitochondrial genes of 32 annelid species from sequence reads deposited in the Sequence Read Archive (SRA)

Genki Kobayashi

Research Center for Creative Partnerships, Ishinomaki Senshu University, Ishinomaki, Miyagi, Japan

ABSTRACT

Background. The mitochondrial genomes (mitogenomes) of metazoans generally include the same set of protein-coding genes, which ensures the homology of mitochondrial genes between species. The mitochondrial genes are often used as reference data for species identification based on genetic data (DNA barcoding). The need for such reference data has been increasing due to the application of environmental DNA (eDNA) analysis for environmental assessments. Recently, the number of publicly available sequence reads obtained with next-generation sequencing (NGS) has been increasing in the public database (the NCBI Sequence Read Archive, SRA). Such freely available NGS reads would be promising sources for assembling mitochondrial protein-coding genes (mPCGs) of organisms whose mitochondrial genes are not available in GenBank. The present study aimed to assemble annelid mPCGs from raw data deposited in the SRA.

Methods. The recent progress in the classification of Annelida was briefly introduced. In the present study, the mPCGs of 32 annelid species of 19 families in clitellates and allies in Sedentaria (echiurans and polychaetes) were newly assembled from the reads deposited in the SRA. Assembly was performed with a recently published pipeline mitoRNA, which includes cycles of Bowtie2 mapping and Trinity assembly. Assembled mPCGs were deposited in GenBank as Third Party Data (TPA) data. A phylogenetic tree was reconstructed with maximum likelihood (ML) analysis, together with other mPCGs deposited in GenBank.

Results and Discussion. mPCG assembly was largely successful except for *Travisia forbesii*; only four genes were detected from the assembled contigs of the species probably due to the reads targeting its parasite. Most genes were largely successfully obtained, whereas *atp8*, *nad2*, and *nad4l* were only successful in 22–24 species. The high nucleotide substitution rates of these genes might be relevant to the failure in the assembly although *nad6*, which showed a similarly high substitution rate, was successfully assembled. Although the phylogenetic positions of several lineages were not resolved in the present study, the phylogenetic relationships of some polychaetes and leeches that were not inferred by transcriptomes were well resolved probably due to a more dense taxon sampling than previous phylogenetic analyses based on transcriptomes. Although NGS data are generally better sources for resolving phylogenetic relationships of both higher and lower classifications, there are ensuring

Submitted 21 March 2023
Accepted 21 October 2023
Published 29 November 2023

Corresponding author
Genki Kobayashi,
genkikobayashi5884@gmail.com

Academic editor
Barbara Langille

Additional Information and
Declarations can be found on
page 13

DOI 10.7717/peerj.16446

© Copyright
2023 Kobayashi

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

needs for specific loci of the mitochondrial genes for analyses that do not require high resolutions, such as DNA barcoding, eDNA, and phylogenetic analysis among lower taxa. Assembly from publicly available NGS reads would help design specific primers for the mitochondrial gene sequences of species, whose mitochondrial genes are hard to amplify by Sanger sequencing using universal primers.

Subjects Bioinformatics, Evolutionary Studies, Genomics, Molecular Biology, Zoology

Keywords RNA-seq, Polychaetes, Cox1, Mitochondria, Hirudineans, Annelida, Sedentaria

INTRODUCTION

The mitochondrial genomes (mitogenomes) of metazoans generally include the same set of 13 protein-coding genes (PCGs), namely, *atp6*, *atp8*, *cox1–3*, *cytb*, *nad1–6*, and *nad4L*. The mitochondrial PCGs (mPCGs) are homologous and thus the nucleotide sequences of mitochondrial genes flanked by conservative regions are therefore easily compared using sequences obtained with universal primers (e.g., the *cox1* gene; [Folmer et al., 1994](#)). In addition, such gene sequences are applied to various molecular techniques, such as DNA barcoding, eDNA, and phylogenetic analyses. However, the mitochondrial gene sequences of some lineages are hard or impossible to amplify using universal primers with standard protocols due to nucleotide substitutions in conservative sites ([Carr et al., 2011](#); [Sun, Kupriyanova & Qiu, 2012](#)) and the insertion of group II introns between the annealing sites of universal primers ([Bernardino et al., 2017](#); [Kobayashi, Itoh & Kojima, 2022c](#)). Such problems can be resolved by adjusting conditions for polymerase chain reaction (PCR) and/or designing more specific primers for a target group. Also, the mitochondrial gene sequences in sequence reads generated by next-generation sequencing (NGS), such as genome skimming and RNA sequencing (RNA-seq), can help in determining the nucleotide sequences of such lineages. Genome skimming is applicable to ethanol-fixed specimens that are not suitable for RNA-seq, which requires high-quality and specifically fixed samples.

The NGS data openly available in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) can also be useful sources for obtaining the mitochondrial genes of species whose mitogenomes are not yet publicly available. RNA-seq and genome skimming are powerful tools for resolving phylogenetic relationships among higher taxa ([Kocot et al., 2011](#); [Kocot et al., 2017](#); [Weigert et al., 2014](#); [Laumer et al., 2019](#); [Ma et al., 2019](#); [Martín-Durán et al., 2021](#); [Tan et al., 2021](#); [Taite et al., 2023](#)). Consequently, NGS data are accumulating for several phyla to elucidate inter-familial relationships that are hard to be resolved with limited genes. However, the annotations of each assembled gene used for the phylogenetic analyses are often not available in public databases (e.g., GenBank) and are not readily used by other researchers. A recent study developed a pipeline “mitoRNA” for assembling mitochondrial genes from RNA-seq reads ([Forni et al., 2019](#)). This pipeline includes two steps for assembling mitochondrial genes from transcriptome data: (1) mapping all transcriptome reads on reference mitogenomes with Bowtie2 ([Langmead & Salzberg, 2012](#)); (2) assembling mapped reads with Trinity ([Grabherr et al., 2011](#)). The resultant contigs are then used as references for successive

iterations of these steps automatically. Newly assembled mitochondrial PCGs from public NGS data enable analyses with a wider taxon sampling than those solely based on published mitogenomes.

Mainly, two types of NGS data, transcriptomes and genomic data, would be a candidate for the assembly of mitochondrial genes. A lot of the transcriptome data of Annelida, one of the major components of marine benthos, are available in the SRA; searching with the keywords “Annelida” and “RNA-seq” yielded 2,162 records in 2022. Furthermore, the number of deposited transcriptomes has been increasing as the search with the same words returned 2,524 records in October 2023. In addition to the application of transcriptomes to evolutionary developmental biology, the transcriptomes of various annelid lineages are obtained to infer deep phylogenetic relationships (Weigert *et al.*, 2014; Andrade *et al.*, 2015; Struck *et al.*, 2015; Helm *et al.*, 2018; Erséus *et al.*, 2020; Martín-Durán *et al.*, 2021; Tilic *et al.*, 2022) since the phylogenetic relationships of higher taxa in annelids were not sufficiently resolved in phylogenies based on several genes. Transcriptomes are also used for the phylogenetic analyses of interfamilial relationships of closely related families and intrafamilial relationships of annelids (Lemer *et al.*, 2015; Novo *et al.*, 2016; Anderson *et al.*, 2017; Stiller *et al.*, 2020; Tilic *et al.*, 2020b; Shekhovtsov *et al.*, 2022). Also, there has been an increase in the number of studies that employ genome skimming for phylogenetic analyses based on mitogenomes in recent years (Richter *et al.*, 2015; Bernardino *et al.*, 2017; Zhang *et al.*, 2018; Zhao *et al.*, 2022; Struck *et al.*, 2023). Although the publicly available genomic data of annelids are still limited, those of annelids are rapidly accumulating led by large projects, *e.g.*, the Darwin Tree of Life Project (<https://www.darwintreeoflife.org/>), which aims to sequence genomes of all known species of eukaryotes in Britain and Ireland.

In the present study, the mPCGs of 32 annelid species of 19 families of clitellates and allies in Sedentaria (echiurans and some polychaetes) were newly assembled from the NGS reads, mainly obtained by transcriptome analysis (Erséus *et al.*, 2020), deposited in the NCBI SRA database. Also, the phylogenetic analyses based on 13 mPCGs are conducted using clitellates and allies.

MATERIALS & METHODS

Although the paraphyletic status of “Polychaeta” has been recognized for decades, the higher classification of Annelida is not yet stable (for recent reviews, see Weigert & Bleidorn, 2016; Struck, 2019). There has been notable progress in the classification of Annelida for the last decade and a half. Errantia and Sedentaria were resurrected (Struck *et al.*, 2011) and a clade name Pleistoannelida was proposed for Sedentaria + Errantia (any of which were without a taxonomic rank) (Struck, 2011). Andrade *et al.* (2015) rejected Pleistoannelida mainly because they considered changes in the phylogenetic positions of several lineages shown by Andrade *et al.* (2015) made the original definition of Pleistoannelida inappropriate. Later, Struck (2019) adjusted the definition of Pleistoannelida to incorporate the entities of Errantia and Sedentaria shown by recent studies. Rouse, Pleijel & Tilic (2022) followed the view regarding Pleistoannelida in Andrade *et al.* (2015) although the changes in Struck (2019) were not cited in Rouse, Pleijel & Tilic

(2022). Weigert & Bleidorn (2016) proposed Palaeoannelida for the clade Magelonidae + Oweniidae. Recently, the considerably revised classification for Annelida was proposed in a book by researchers mainly working on polychaetes (Rouse, Pleijel & Tilic, 2022), including the redefinition of the class Polychaeta. Polychaeta sensu Rouse, Pleijel & Tilic (2022) comprises Errantia and Sedentaria and thus it includes leeches and oligochaetes but not early branching lineages that are the members of conventional polychaetes (e.g., Magelonidae, Oweniidae, Chaetopteridae, and Amphinomidae). The common name “polychaetes” for Polychaeta sensu Rouse, Pleijel & Tilic (2022) therefore includes leeches and oligochaetes but not magelonids, oweniids, chaetopterids, or amphinomids. This situation is different from the conventional usage of polychaetes and can lead to confusion when using the word “polychaetes” (it is avoided when using “pleistoannelids” for Errantia + Sedentaria and polychaetes for conventional “Polychaeta”). Furthermore, the taxonomic grouping of clitellates is also controversial: Rouse, Pleijel & Tilic (2022) proposed the order Oligochaeta, consisting of Aeolosomatidae, *Hrabeiella*, and the suborder Clitellata, within Polychaeta sensu Rouse, Pleijel & Tilic (2022), whereas Clitellata was conventionally higher than Oligochaeta (see a recent proposed order-level classification of oligochaetes; Schmelz et al., 2021). For these reasons, further discussion on the classification of Annelida would be required to establish a stable and widely accepted classification of Annelida in the future. In the present study, common names (not as formal taxonomic groupings) polychaetes, oligochaetes, and clitellates are used to indicate the conventional grouping (i.e., not polychaetes, etc. in Rouse, Pleijel & Tilic, 2022) to avoid confusion explained above.

The paired-end raw reads of 32 annelid species of Sedentaria were downloaded from the SRA with fasterq-dump v3.0.0 (Table 1). Adapter trimming and quality filtering ($Q > 30$) were performed using fastp v0.20.1 (Chen et al., 2018). The first five million reads were extracted from filtered reads with the “head” command of SeqKit v0.12.0 (Shen et al., 2016) to save computation time of assembly since the results of a preliminary assembly using all reads (32 million) of *Travisia forbesii* did not differ between assembly using five million reads. A pipeline mitoRNA (Forni et al., 2019), which uses Bowtie2 v2.4.5 (Langmead & Salzberg, 2012) for mapping reads on references and Trinity v2.14.0 (Grabherr et al., 2011) for assembling mapped reads, was used to assemble mPCGs from the reads with default settings. The mitogenomes of *Abarenicola claparedi oceanica*, *Notomastus* sp., *Urechis unicinctus*, *Nais communis*, *Olavius algarvensis*, *Tubifex tubifex*, *Acanthobdella peledina*, and *Ozobranchus jantseanus* were used as references for mitoRNA. The mPCGs were searched from assembled contigs with the “nhmmer” command (Wheeler & Eddy, 2013) (using the --max option) in HMMER v3.3.2 (<http://hmmer.org/>) using HMM files constructed with the “hmmbuild” command based on the alignment files of annelid mPCGs (same dataset as Kobayashi, Itoh & Kojima, 2022c). Then, searched sequences were used as references for GetOrganelle v1.7.5.1 (Jin et al., 2020) and/or NOVOPlasty v4.2.1 (Dierckxsens, Mardulyn & Smits, 2017), which were used to assemble the complete length of each gene from the contigs generated by mitoRNA. The average coverage of each gene was checked with the “bbmap” command implemented in BBtools (<https://sourceforge.net/projects/bbmap/>). The obtained genes were deposited in DDBJ/EMBL/GenBank through DNA Data Bank of Japan as TPA (Third Party Data) (Table S1 for accession numbers of each gene sequence),

Table 1 Accession numbers of sequence reads downloaded from NCBI SRA. Newly assembled mitochondrial protein-coding gene sequences with TPA accession numbers are listed in [Table S1](#).

Family	Species	SRR
Opheliidae	<i>Ophelina acumulata</i>	SRR10997422
	<i>Thoracophelia mucronata</i>	SRR2017631
Thalassematidae	<i>Bonellia viridis</i>	SRR2017645
Trichobranchidae	<i>Trichobranchus roseus</i>	SRR11434466
Arenicolidae	<i>Arenicola marina</i>	SRR2005653
	<i>Abarenicola pacifica</i>	SRR10997426
Scalibregmatidae	<i>Scalibregma inflatum</i>	SRR8799334
Travisiidae	<i>Travisia forbesii</i> ^a	SRR9888046
Aeolosomatidae	<i>Aeolosoma</i> sp.	SRR11559519
Hrabeiellidae	<i>Hrabeiella periglandulata</i>	SRR10997424
Randiellidae	<i>Randiella</i> sp.	SRR10997431
Parvidrilidae	<i>Parvidrilus meyssonnieri</i>	SRR8799336
Capilloventridae	<i>Capilloventer australis</i>	SRR8799324
Phreodrilidae	Phreodrilidae sp. A	SRR10997437
Naididae	<i>Albanidrilus</i> sp.	SRR10997452
	<i>Chaetogaster diaphanus</i>	SRR10997419
	<i>Bathydrilus rohdei</i>	SRR8799332
	<i>Olavius</i> sp.	SRR8799329
	<i>Potamothrix</i> nr <i>heuscheri</i>	SRR10997432
Propappidae	<i>Rhyacodrilus pigueti</i>	SRR8799325
Enchytraeidae	<i>Propappus volki</i>	SRR5353250
	<i>Grania simonae</i>	SRR10997449
Lumbriculidae	<i>Enchytraeus crypticus</i>	SRR10997417
	<i>Lumbriculus variegatus</i>	SRR8842488
	<i>Kincaidiana</i> sp.	SRR10997445
Branchiobdellidae	<i>Cirrodrilus suzukii</i>	SRR8842480
	<i>Branchiobdella kobayashii</i>	SRR8799326
	<i>Holtodrilus truncatus</i>	SRR8842481
	<i>Bdellodrilus illuminatus</i>	SRR8842477
	<i>Triannulata magna</i>	SRR8842482
Cylicobdellidae	Cylicobdellidae sp.	SRR8842484
Haemopidae	<i>Haemopsis sanguisuga</i>	SRR10997447

Notes.

^aGene sequences were not used for phylogenetic analysis.

[PRJDB14830](#) (BioProject), and [SAMD00561016–SAMD00561048](#) (BioSample). The species and SRA accession numbers used in this study are summarized in [Table 1](#).

The assembled mPCGs were aligned with the corresponding genes of part of Sedentaria (clitellates, echiurans, and part of polychaetes in Sedentaria) whose mitogenomes are deposited in GenBank and were used for phylogenetic analyses ([Table 2](#)) (88 OTUs). Siboglinidae was used as an outgroup. This subset was selected since the mitogenomes of some families of Sedentaria, such as Serpulidae, Fabriciidae, and Spionidae, show an extremely high nucleotide substitution rate in their mitogenomes ([Seixas et al.](#),

2017; Tilic, Atkinson & Rouse, 2020a; Sun et al., 2021; Ye et al., 2021). *Travisia forbesii* was excluded from subsequent analyses since only four genes were successfully assembled. The mPCGs of *Capitella teleta* were searched by nhmmer using genome data downloaded from EnsembleMetazoa (<https://metazoa.ensembl.org/index.html>) and included in the phylogenetic analysis. The amino acid sequences were translated using invertebrate mitochondrial code with the “translate” command with the “-M” option (translate initial codon at the beginning to M) in SeqKit v0.12.0 (Shen et al., 2016). The amino acid sequences were aligned with MAFFT v7 (Katoh & Standley, 2013) using the G-INS-I option. The aligned amino acid sequences were checked manually. The “tranalign” command implemented in EMBOSS v6.6.0.0. (Rice, Longden & Bleasby, 2000) was used to align nucleotide sequences based on the aligned amino acid sequences. Ambiguous positions were excluded using trimAl (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009) with the “gappout” option. The maximum likelihood (ML) analysis was conducted for amino acid sequences with IQ-TREE v1.6.12 (Nguyen et al., 2014) using 1,000 ultrafast bootstrap (UFBoot) replicates. The interpretations of UFBoot values (ufBS) are different from the normal bootstrap values and $ufBS \geq 95\%$ would be reliable (Minh et al., 2021). The NEXUS partition files were prepared to input sequence data into IQ-TREE. The best-fit substitution model for each gene was selected with ModelFinder (Kalyaanamoorthy et al., 2017), implemented in IQ-TREE (Data S1). FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used to draw phylogenetic trees.

The pairwise genetic distances (p -distances) of aligned nucleotide sequences of each gene were calculated with the “distmat” command of EMBOSS (Rice, Longden & Bleasby, 2000).

A degenerated primer, modified from HCO2198 (Folmer et al., 1994), was designed considering the alignment of the *cox1* gene (HCO-cliteall: 5'-CTTCNGGRTGNCCRAARAAAYCA-3'), which can be used as a pair of LCO-annelid (Kobayashi, Itoh & Kojima, 2022c).

RESULTS

Assembly of mitochondrial protein-coding genes

mPCG assembly was largely successful except for *Travisia forbesii*; only four genes (*cox1*, *cox2*, *cytb*, and *nad5*) of this species were detected from assembled contigs. For 31 other species, *cox1* and *cytb* were successfully detected in all of them (Table S1; partly assembled genes are listed in Table S2). Also, *cox2*, *cox3*, *nad1*, *nad3*, *nad4*, *nad5*, and *nad6* were obtained in ≥ 29 species. On the other hand, *atp8*, *nad2*, and *nad4l* were only successful in 22–24 species. The medians of pairwise differences of genes among species were higher than 50% in *atp8* and *nad6* (Fig. 1).

Phylogenetic analysis based on amino acid sequences

In a polychaete lineage, Traviidae + Scalibregmatidae was monophyletic ($ufBS = 100\%$) and sister to a clade including Arenicolida (Arenicolidae and Maldanidae; $ufBS = 100\%$) and Terebellida (Pectinariidae, Alvinellidae, Ampharetidae, Terebellidae, Melinnidae, and Trichobranchidae; $ufBS = 100\%$) ($ufBS = 99\%$) (Fig. 2). The support value between families in Terebellida was high for the position of Pectinariidae, which was sister to the

Table 2 Mitogenome sequences obtained from GenBank and used for phylogenetic analysis, in addition to newly assembled gene sequences.

Family	Species	GenBank accession No.
Moniligastridae	<i>Drawida gisti</i>	MN539609
	<i>Drawida japonica</i>	KM199288
Megascolecidae	<i>Amyntas aspergillus</i>	KJ830749
	<i>Metaphire vulgaris</i>	KJ137279
	<i>Perionyx excavatus</i>	EF494507
	<i>Tonoscolex birmanicus</i>	KF425518
Lumbricidae	<i>Aporrectodea rosea</i>	MK573632
	<i>Eisenia balatonica</i>	MK642872
	<i>Lumbricus terrestris</i>	U24570
Rhinodrilidae	<i>Pontoscolex corethrurus</i>	KT988053
Naididae	<i>Limnodrilus hoffmeisteri</i>	MW732144
	<i>Nais communis</i>	MW770354
	<i>Olavius algarvensis</i>	LR992058
	<i>Tubifex tubifex</i>	MW690579
Acanthobdellidae	<i>Acanthobdella peledina</i>	MZ562997
Haemadipsidae	<i>Haemadipsa crenata</i>	MW711186
Hirudinidae	<i>Hirudo medicinalis</i>	KU672396
	“ <i>Hirudo nipponia</i> ”	KC667144
	<i>Hirudo verbana</i>	KU672397
	<i>Poecilobdella javanica</i>	MN542781
	“ <i>Poecilobdella manillensis</i> ”	KC688268
	<i>Whitmania acranulata</i>	KM655838
	<i>Whitmania laevis</i>	KC688269
	<i>Whitmania laevis</i>	KM655839
	Erpobdellidae	<i>Erpobdella japonica</i>
<i>Erpobdella testacea</i>		MT584166
“ <i>Erpobdella octoculata</i> ”		KC688270
Glossiphoniidae		<i>Glossiphonia complanata</i>
	<i>Glossiphonia concolor</i>	MT628565
	<i>Haementeria officinalis</i>	LT159848
	<i>Hemiclepsis yangtzenensis</i>	MN106285
	<i>Placobdella lamothei</i>	LT159849
	<i>Placobdella parasitica</i>	LT159850
Piscicolidae	<i>Codonobdella</i> sp.	MZ202177
	<i>Piscicola geometra</i>	BK059172
	<i>Zeylanicobdella arugamensis</i>	KY474378
Ozobranchidae	<i>Ozobranchus jantseanus</i>	KY861060
Travisiidae	<i>Travisia sanrikuensis</i>	LC677172
Ampharetidae	<i>Auchenoplax crinita</i>	FJ976041

(continued on next page)

Table 2 (continued)

Family	Species	GenBank accession No.
	<i>Decemunciger</i> sp.	KY742027
	<i>Eclysippe vanelli</i>	EU239687
Alvinellidae	<i>Paralvinella sulfincola</i>	FJ976042
Melinnidae	<i>Melinna cristata</i>	MW542504
Trichobranchidae	<i>Terebellides stroemii</i>	EU236701
Terebellidae	<i>Neoamphitrite affinis</i>	MZ326700
	<i>Pista cristata</i>	EU239688
	<i>Thelepus plagiostoma</i>	MW557377
Pectinariidae	<i>Pectinaria gouldii</i>	FJ976040
Arenicolidae	<i>Abarenicola claparedi oceanica</i>	LC707921
Maldanidae	<i>Clymenella torquata</i>	AY741661
Cepitellidae	<i>Notomastus</i> sp.	LC661358
Opheliidae	<i>Armandia</i> sp.	LC661359
Thalassematidae	<i>Urechis caupo</i>	AY619711
	<i>Urechis unicinctus</i>	EF656365
Siboglinidae	<i>Lamellibrachia luymesii</i>	KJ789163
	<i>Sclerolinum brattstromi</i>	KJ789167

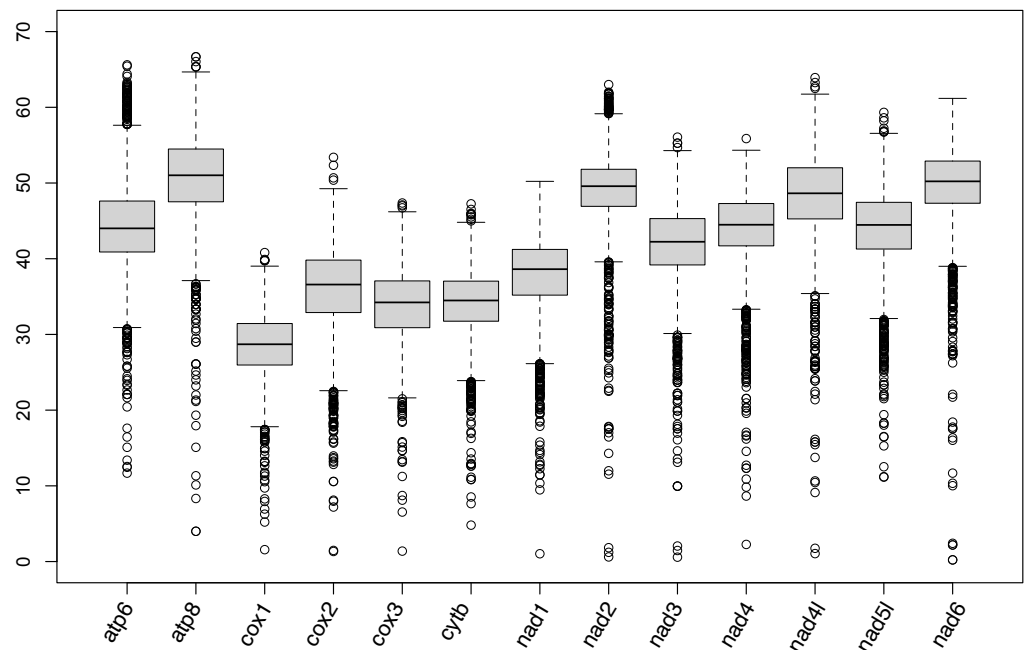


Figure 1 Box plot showing the percentage of the pairwise genetic distances of protein-coding genes of mitogenomes.

Full-size  DOI: [10.7717/peerj.16446/fig-1](https://doi.org/10.7717/peerj.16446/fig-1)

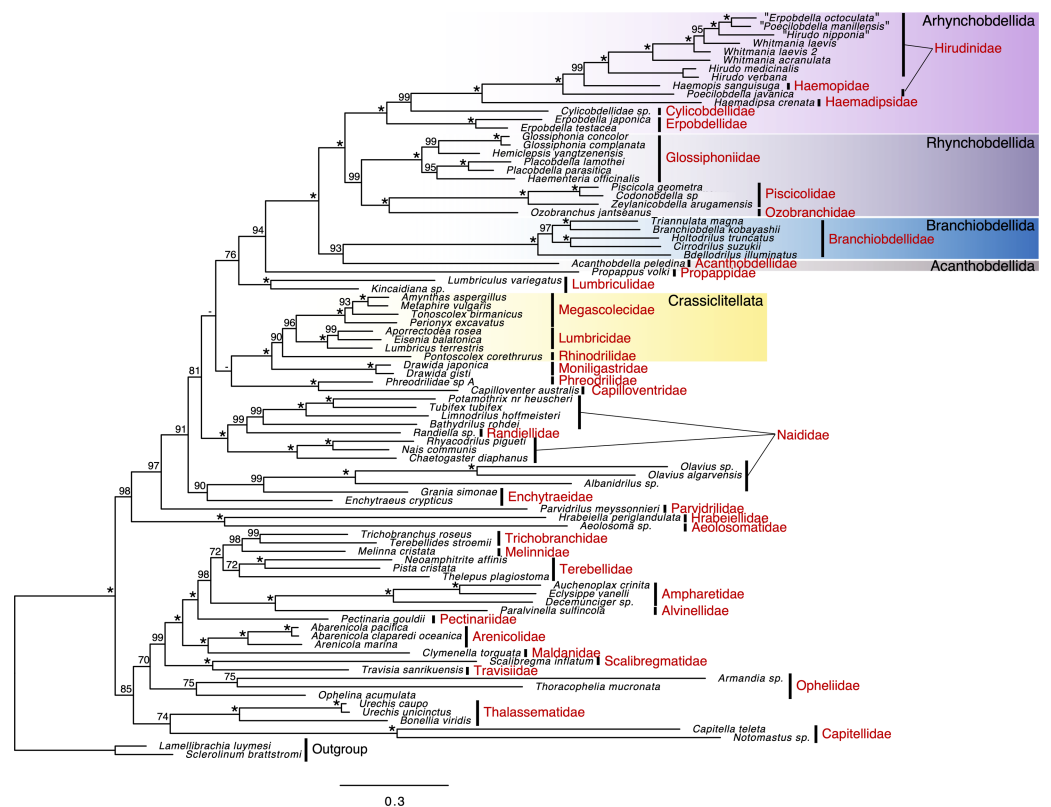


Figure 2 Maximum likelihood phylogeny of a subset of Sedentaria based on the dataset, including amino acid sequences of 13 protein-coding genes (PCGs) after ambiguous positions were excluded (3,504 characters). Numbers above the branches represent the ultrafast bootstrap values (ufBS; $\geq 95\%$ would be reliable; Minh et al., 2021). Asterisks indicate highly supported branches (ufBS = 100%).

Full-size [DOI: 10.7717/peerj.16446/fig-2](https://doi.org/10.7717/peerj.16446/fig-2)

remaining Terebellida (ufBS = 98%), Trichobranchidae + Mellinidae (ufBS = 98%), Ampharetidae + Alvinellidae (ufBS = 100%). Monophyletic Capitellidae (*Capitella teleta* and *Notomastus* sp.; ufBS = 100%) was sister to Thalamematidae (ufBS = 100%) although the support value was low (ufBS = 74%).

The sister relationship between monophyletic Aphanoneura (Aeolosomatidae and Hrabieilidae; ufBS = 100%) and monophyletic clitellates (ufBS = 97%) was highly supported (ufBS = 98%). Parvidrilidae was sister to the moderately supported lineage of other clitellates (ufBS = 91%) although the support value was moderate (ufBS = 91%). The support values at deep nodes of oligochaetes and inside Naididae were largely low. A part of Naididae (*Olavius* spp. and *Albanidrilus*) was included in Enchytraeidae (*Enchytraeus crypticus* and *Grania simonae*) (ufBS = 90%). Randiellidae was nested within Naididae and the monophyly of this cluster was fully supported (ufBS = 100%). Monophyly of Phreodrilidae and Capilloventridae was fully supported, although the position of this clade was ambiguous. Propappidae was sister to the fully supported clade of leeches (ufBS = 94%), and Lumbriculidae was sister to this clade although the support value was low (ufBS = 76%).

UFboot support values in leeches were generally high. Branchiobdellida + Acanthobdellida (ufBS = 93%) was sister to Hirudinida (ufBS = 100%). Although Branchiobdellidae was monophyletic (ufBS = 100%), Branchiobdellinae (*Branchiobdella kobayashii* and *Cirrodrilus suzukii*) was not monophyletic: *B. kobayashii* was sister to Cambarincolinae (*Triannulata magna*) (ufBS = 100%) and *C. suzukii* was sister to *Holtodrilus truncatus* (ufBS = 100%). Each of Rhynchobdellida and Arhynchobdellida was monophyletic (ufBS = 99% and 100%, respectively). In Rhynchobdellida, Piscicolidae + Ozobranchidae clade (Oceanobdelliformes) (ufBS = 100%) was sister to Glossiphoniidae (ufBS = 99%). In Arhynchobdellida, Erpobdellidae (Erpobdelliformes) was sister to Hirudiniformes (Cyclicobdellidae + Haemadipsidae + Haemopidae + Hirudinidae) (ufBS = 100%). Cyclicobdellidae was sister to the other Hirudiniformes (ufBS = 99%). The sister relationship between Haemadipsidae and the Hirudinidae + Haemopidae clade (ufBS = 99%) was fully supported. *Poecilobdella javanica* (Hirudinidae) was sister to the highly supported lineage (ufBS = 99%) of Haemopidae + other hirudinids (ufBS = 100%), resulting in the paraphyletic nature of Hirudinidae.

DISCUSSION

In the present study, the complete mPCGs of 11 families were registered in GenBank for the first time (Aeolosomatidae, Branchiobdellidae, Capilloventridae, Cyclicobdellidae, Enchytraeidae, Hrabeiellidae, Randiellidae, Scalibregmatidae, Parvidrilidae, Phreodrilidae, and Propappidae). The mPCGs of the part of Sedentaria were well-assembled from the reads of RNA-seq or anchored hybrid enrichment (AHE) deposited in the SRA. Sequence reads generated by genome skimming were not used in the present study because of suitable datasets of the target group found in two previous studies using RNA-seq or AHE ([Phillips et al., 2019](#); [Erséus et al., 2020](#)). Since genome skimming aims to skim the nucleotide sequences of the high copy fraction of the genome ([Straub et al., 2012](#)), prevalent organelle genome assemblers usually succeed in assembling the mitochondrial genome from the sequence reads generated by genome skimming. Assembly using the pipeline mitoRNA introduced in the present study may therefore be unnecessary for genomic data. The *atp8*, *nad2*, and *nad4l* genes showed high nucleotide substitution rates ([Fig. 1](#)). This may have resulted in the poor performances of nhmmer due to a divergence between query contigs and reference mitogenomes. These genes may be obtained when the mitogenomes of more closely related species or the species become available as the reference for assembly in the future. The low success rate, however, did not necessarily coincide with high nucleotide substitution rates (e.g., the *nad6* gene was assembled in 31 species although it shows a high substitution rate, which is similar to the *nad2* gene). The *atp8*, *nad2*, and *nad4l* genes were also poorly assembled in Nereididae (Errantia) by a previous study on mitogenomic phylogeny of Nereididae ([Alves, Halanych & Santos, 2020](#)), which assembled mitochondrial genes from RNA-seq reads using Trinity. Unfortunately, the mPCGs of *Travisia forbesii* were poorly assembled probably because the RNA-seq targeted its apicomplexan parasite *Rhytidocystis* sp. (deposited as the transcriptome of *Rhytidocystis* sp. 1; Organism: *Rhytidocystis* sp. ex *Travisia forbesii*) and sequence reads derived from *T. forbesii* may be scarce.

In the present study, phylogenetic relationships in polychaetes and leeches that were not inferred by transcriptomes were well resolved probably due to a more dense taxon sampling than previous phylogenetic analyses based on transcriptomes or mitogenomes, enabled by using both mitogenome data and mitochondrial genes assembled from NGS data. Scalibregmatidae and Traviisiidae were recovered as sister lineages, corroborating the previous studies based on phylogenetic analyses based on several genes ([Paul et al., 2010](#); [Martínez, Di Domenico & Worsaae, 2013](#); [Martínez, Di Domenico & Worsaae, 2014](#); [Law, Dorgan & Rouse, 2014](#)). Although the family Haemopidae was recognized as a valid taxa in a recent study ([Kvist et al., 2023](#)), *Haemopsis* was nested within Hirudinidae in the present study, providing further support for the polyphyly of Hirudinidae suggested by phylogenetic analyses based on several nuclear and mitochondrial genes ([Phillips & Siddall, 2009](#); [Tessler et al., 2018](#)).

Aphanoneura (Aeolosomatidae + Hrabeiellidae) was recovered as the sister clade to the remaining clitellates in the present analysis, which is consistent with the results of phylogenetic analyses based on transcriptomes by [Erséus et al. \(2020\)](#). [Erséus et al. \(2020\)](#) stated a more exhaustive taxon sampling among the “polychaetous” groups (especially *Parergodrilus* and *Stygocapitella*) will be needed to conclude the sister of clitellates, as suggested by [Weigert & Bleidorn \(2016\)](#). *Parergodrilus* and *Stygocapitella*, i.e., the members of Parergodrilidae, are now shown to be closely related to Orbiniidae, not to clitellates ([Struck et al., 2015](#)). Although the phylogenetic positions of some other annelid families were not yet revealed, the present study also supports the sister relationship between Aphanoneura and clitellates.

Unfortunately, the phylogenetic positions of several lineages, which were well resolved in the phylogenetic analyses based on transcriptomes, were not resolved in the present study. For example, the phylogenetic positions of Naididae (*Olavius* spp. and *Albanidrilus*), *Randiella* sp., Lumbriculidae, and the Phreodrilidae + Capilloventridae clade were ambiguous. Although the Acanthobdellida and Hirudinida clade was sister to Branchiobdellida in [Phillips et al. \(2019\)](#) with full support, the relationship among Acanthobdellida, Branchiobdellida, and Hirudinida was not resolved in the present study. Also, the phylogenetic positions of Capitellidae, Thalamematidae, and Opheliidae, which were contentious in the recent phylogenetic studies focusing on the mitogenomes of Sedentaria ([Kobayashi, Itoh & Nakajima, 2022b](#); [Kobayashi, Itoh & Nakajima, 2022a](#)) despite well-resolved phylogenetic relationships based on transcriptomes, were not resolved in the present study. Mitogenomic phylogeny can be conducted with a more dense taxon sampling than phylogeny based on transcriptomes, it should be noted that support values in mitogenomic phylogeny of the higher classifications ([Sun et al., 2021](#); [Fig. 1 in Kobayashi, Itoh & Nakajima, 2022a](#)) are likely to be lower than phylogeny based on transcriptomes with similar taxon sampling ([Struck et al., 2015](#); [Martín-Durán et al., 2021](#)) (although it is apparent considering the difference in the total number of loci).

Although the phylogenetic analysis based on genome-wide data is a powerful method for resolving phylogenetic relationships of both higher and lower classifications, analyses based on mitogenomes or each widely sequenced mitochondrial gene (e.g., *cox1*) are still enduring needs in taxonomy and environmental assessments that do not need high

resolution but require a wide range of taxon sampling, such as DNA barcoding, eDNA, and phylogenetic analyses between closely related species. In particular, the reference sequences for species identification are in great demand for meta-analyses since the barcode coverage for marine invertebrates is still quite low; only 22% of species in the European Register of Marine Species (ERMS) list have been sequenced ([Weigand et al., 2019](#)). The gene sequences assembled in the present study include species whose mitochondrial genes are registered for the first time and they would be usable as the reference sequences for future DNA metabarcoding studies. The mitochondrial genes of some species that gave nucleotide substitutions in annealing sites of universal primers are hard or impossible to obtain gene sequences by such primers, possibly being one of the reasons for the delay in accumulating reference sequences. Although designing specific primers for lower taxa can successfully obtain gene sequences that are hard or impossible to amplify by universal primers (*e.g.*, [Sun, Kupriyanova & Qiu, 2012](#); [Williams et al., 2017](#)), universal primers with degenerated bases are still useful for some annelids with high nucleotide substitutions in mitochondrial genes and uncertain phylogenetic relationships among genera and/or species since there are no promising specific primers for such annelids yet. *e.g.*, some genera in Capitellidae ([Jeong et al., 2018](#); [Jeong, Soh & Suh, 2019](#)) and Fauveliopsidae (G Kobayashi, 2023, unpublished data). Mitochondrial gene assembly from the NGS reads of the representative species of such taxa may be effective in designing specific primers and obtaining target gene sequences of them.

CONCLUSIONS

Mitochondrial protein-coding genes were newly assembled from the reads of RNA-seq or AHE deposited in the NCBI SRA of 32 annelid species of 19 families, using a pipeline mitoRNA. The dataset for phylogenetic analyses includes mitochondrial protein-coding genes obtained from mitogenomes and transcriptomes in the database, resulting in a dense taxon sampling of polychaetes and leeches compared with previous studies using either mitogenome or transcriptome. Although the support value between families was largely low in oligochaetes, some phylogenetic relationships in polychaetes and leeches were well resolved. The assembled mitochondrial genes from publicly available NGS reads would be useful for analyses that do not require high resolutions, such as DNA barcoding, eDNA, and phylogenetic analysis among lower taxa.

ACKNOWLEDGEMENTS

I am grateful to two anonymous reviewers for their invaluable comments on the earlier version of the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by JSPS KAKENHI (grant number JP22K15174). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the author:
JSPS KAKENHI: JP22K15174.

Competing Interests

The author declares that he has no competing interests.

Author Contributions

- Genki Kobayashi conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The obtained genes are available in DDBJ/EMBL/GenBank through DNA Data Bank of Japan as TPA data:

BR001822–BR001891, YAAB01000001–YAAB01000013, YAAC01000001–YAAC01000012, YAAD01000001–YAAD01000013, YAAE01000001–YAAE01000013, YAAF01000001–YAAF01000012, YAAG01000001–YAAG01000013, YAAH01000001–YAAH01000013, YAAI01000001–YAAI01000013, YAAJ01000001–YAAJ01000013, YAAK01000001–YAAK01000012, YAAL01000001–YAAL01000013, YAAM01000001–YAAM01000011, YAAN01000001–YAAN01000013, YAAO01000001–YAAO01000012, YAAP01000001–YAAP01000013, YAAQ01000001–YAAQ01000013, YAAR01000001–YAAR01000011, YAAS01000001–YAAS01000013, YAAT01000001–YAAT01000013, YAAU01000001–YAAU01000011, YAAV01000001–YAAV01000013, YAAW01000001–YAAW01000013, YAAX01000001–YAAX01000013, YAAY01000001–YAAY01000009, YAAZ01000001–YAAZ01000004, YABA01000001–YABA01000012, PRJDB14830, SAMD00561016–SAMD00561033, SAMD00561035–SAMD00561048.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.16446#supplemental-information>.

REFERENCES

- Alves PR, Halanych KM, Santos CSG. 2020. The phylogeny of Nereididae (Annelida) based on mitochondrial genomes. *Zoologica Scripta* 49:366–378
DOI 10.1111/zsc.12413.

- Anderson FE, Williams BW, Horn KM, Erséus C, Halanych KM, Santos SR, James SW. 2017. Phylogenomic analyses of Crassiditellata support major Northern and Southern Hemisphere clades and a Pangaeian origin for earthworms. *BMC Evolutionary Biology* 17:1–18 DOI 10.1186/s12862-017-0973-4.
- Andrade SCS, Novo M, Kawauchi GY, Worsaae K, Pleijel F, Giribet G, Rouse GW. 2015. Articulating archiannelids: phylogenomics and annelid relationships, with emphasis on meiofaunal taxa. *Molecular Biology and Evolution* 32:2860–2875 DOI 10.1093/molbev/msv157.
- Bernardino AF, Li Y, Smith CR, Halanych KM. 2017. Multiple introns in a deep-sea annelid (*Decemunciger*: Ampharetidae) mitochondrial genome. *Scientific Reports* 7:4295 DOI 10.1038/s41598-017-04094-w.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973 DOI 10.1093/bioinformatics/btp348.
- Carr CM, Hardy SM, Brown TM, Macdonald TA, Hebert PDN. 2011. A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in canadian polychaetes. *PLOS ONE* 6:e22232 DOI 10.1371/journal.pone.0022232.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890 DOI 10.1093/bioinformatics/bty560.
- Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45:e18–e18 DOI 10.1093/nar/gkw955.
- Erséus C, Williams BW, Horn KM, Halanych KM, Santos SR, James SW, Creuzé M, Anderson FE. 2020. Phylogenomic analyses reveal a Palaeozoic radiation and support a freshwater origin for clitellate annelids. *Zoologica Scripta* 49(5):614–640 DOI 10.1111/zsc.12426.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3:294–299 DOI 10.1371/journal.pone.0013102.
- Forni G, Puccio G, Bourguignon T, Evans T, Mantovani B, Rota-Stabelli O, Luchetti A. 2019. Complete mitochondrial genomes from transcriptomes: assessing pros and cons of data mining for assembling new mitogenomes. *Scientific Reports* 9:14806 DOI 10.1038/s41598-019-51313-7.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29:644–652 DOI 10.1038/nbt.1883.
- Helm C, Beckers P, Bartolomaeus T, Drukewitz SH, Kourtesis I, Weigert A, Purschke G, Worsaae K, Struck TH, Bleidorn C. 2018. Convergent evolution of the ladder-like ventral nerve cord in Annelida. *Frontiers in Zoology* 15:36 DOI 10.1186/s12983-018-0280-y.

- Jeong M, Soh HY, Suh H. 2019.** Three new species of *Heteromastus* (Annelida, Capitellidae) from Korean waters, with genetic evidence based on two gene markers. *ZooKeys* 869:1–18 DOI 10.3897/zookeys.869.34380.
- Jeong M, Soh HY, Wi JH, Suh H. 2018.** A new *Notomastus* (Annelida, Capitellidae) species from Korean waters, with genetic comparison based on three gene markers. *ZooKeys* 754:141–155 DOI 10.3897/zookeys.754.23655.
- Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020.** GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21:241 DOI 10.1186/s13059-020-02154-5.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.** ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14:587–589 DOI 10.1038/nmeth.4285.
- Katoh K, Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780 DOI 10.1093/molbev/mst010.
- Kobayashi G, Itoh H, Kojima S. 2022c.** Mitogenome of a stink worm (Annelida: Traviidae) includes degenerate group II intron that is also found in five congeneric species. *Scientific Reports* 12:4449 DOI 10.1038/s41598-022-08103-5.
- Kobayashi G, Itoh H, Nakajima N. 2022a.** First mitochondrial genome of a lugworm (Annelida: Arenicolidae) and its phylogenetic position. *Journal of the Marine Biological Association of the United Kingdom* 102:635–644 DOI 10.1017/S0025315422001035.
- Kobayashi G, Itoh H, Nakajima N. 2022b.** First mitochondrial genomes of Capitellidae and Opheliidae (Annelida) and their phylogenetic placement. *Mitochondrial DNA Part B* 7:577–579 DOI 10.1080/23802359.2022.2056537.
- Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B, Halanych KM. 2011.** Phylogenomics reveals deep molluscan relationships. *Nature* 477:452–456 DOI 10.1038/nature10382.
- Kocot KM, Struck TH, Merkel J, Waits DS, Todt C, Brannock PM, Weese DA, Cannon JT, Moroz LL, Lieb B, Halanych KM. 2017.** Phylogenomics of Lophotrochozoa with consideration of systematic error. *Systematic Biology* 66:256–282 DOI 10.1093/sysbio/syw079.
- Kvist S, Earl I, Kink E, Ocegüera-Figueroa A, Trontelj P. 2023.** Phylogenetic relationships and species delimitation in *Haemopsis* (Annelida: Hirudinea: Haemopidae). *Molecular Phylogenetics and Evolution* 178:107648 DOI 10.1016/j.ympev.2022.107648.
- Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357–359 DOI 10.1038/nmeth.1923.
- Laumer CE, Fernández R, Lemer S, Combosch D, Kocot KM, Riesgo A, Andrade SCS, Sterrer W, Sørensen MV, Giribet G. 2019.** Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proceedings of the Royal Society B: Biological Sciences* 286:20190831 DOI 10.1098/rspb.2019.0831.

- Law CJ, Dorgan KM, Rouse GW. 2014.** Relating divergence in polychaete musculature to different burrowing behaviors: a study using Opheliidae (Annelida). *Journal of Morphology* 275:548–571 DOI 10.1002/jmor.20237.
- Lemer S, Kawauchi GY, Andrade SCS, González VL, Boyle JM, Giribet G. 2015.** Re-evaluating the phylogeny of Sipuncula through transcriptomics. *Molecular Phylogenetics and Evolution* 83:174–183 DOI 10.1016/j.ympev.2014.10.019.
- Ma KY, Qin J, Lin CW, Chan TY, Ng PKL, Chu KH, Tsang LM. 2019.** Phylogenomic analyses of brachyuran crabs support early divergence of primary freshwater crabs. *Molecular Phylogenetics and Evolution* 135:62–66 DOI 10.1016/j.ympev.2019.02.001.
- Martín-Durán JM, Vellutini BC, Marlétaz F, Cetrangolo V, Cvetesic N, Thiel D, Henriet S, Grau-Bové X, Carrillo-Baltodano AM, Gu W, Kerbl A, Marquez Y, Bekkouche N, Chourrout D, Gómez-Skarmeta JL, Irimia M, Lenhard B, Worsaae K, Hejnol A. 2021.** Conservative route to genome compaction in a miniature annelid. *Nature Ecology and Evolution* 5:231–242 DOI 10.1038/s41559-020-01327-6.
- Martínez A, Di Domenico M, Worsaae K. 2013.** Evolution of cave *Axiokebuita* and *Speleobregma* (Scalibregmatidae, Annelida). *Zoologica Scripta* 42:623–636 DOI 10.1111/zsc.12024.
- Martínez A, Di Domenico M, Worsaae K. 2014.** Gain of palps within a lineage of ancestrally burrowing annelids (Scalibregmatidae). *Acta Zoologica* 95:421–429 DOI 10.1111/azo.12039.
- Minh BQ, Lanfear R, Trifinopoulos J, Schrempf D, Schmidt HA. 2021.** IQ-TREE version 2.1.2: tutorials and manual phylogenomic software by maximum likelihood. Available at <http://www.iqtree.org>.
- Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2014.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274 DOI 10.1093/molbev/msu300.
- Novo M, Fernández R, Andrade SCS, Marchán DF, Cunha L, Díaz Cosín DJ. 2016.** Phylogenomic analyses of a Mediterranean earthworm family (Annelida: Hormogastridae). *Molecular Phylogenetics and Evolution* 94:473–478 DOI 10.1016/j.ympev.2015.10.026.
- Paul C, Halanych KM, Tiedemann R, Bleidorn C. 2010.** Molecules reject an opheliid affinity for *Travisia* (Annelida). *Systematics and Biodiversity* 8:507–512 DOI 10.1080/14772000.2010.517810.
- Phillips AJ, Dornburg A, Zapfe KL, Anderson FE, James SW, Erséus C, Moriarty Lemmon E, Lemmon AR, Williams BW. 2019.** Phylogenomic analysis of a putative missing link sparks reinterpretation of leech evolution. *Genome Biology and Evolution* 11:3082–3093 DOI 10.1093/gbe/evz120.
- Phillips AJ, Siddall ME. 2009.** Poly-paraphyly of Hirudinidae: many lineages of medicinal leeches. *BMC Evolutionary Biology* 9:246 DOI 10.1186/1471-2148-9-246.
- Rice P, Longden I, Bleasby A. 2000.** EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics* 16:276–277 DOI 10.1016/S0168-9525(00)02024-2.

- Richter S, Schwarz F, Hering L, Böggemann M, Bleidorn C. 2015.** The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). *Genome Biology and Evolution* 7:3443–3462 DOI [10.1093/gbe/evv224](https://doi.org/10.1093/gbe/evv224).
- Rouse G, Pleijel F, Tilic E. 2022.** *Annelida*. Oxford: Oxford University Press.
- Schmelz RM, Erséus C, Martin P, Van Haaren T, Timm T. 2021.** A proposed order-level classification in Oligochaeta (Annelida, Clitellata). *Zootaxa* 5040:589–597 DOI [10.11646/zootaxa.5040.4.9](https://doi.org/10.11646/zootaxa.5040.4.9).
- Seixas VC, Russo CA, de M, Paiva PC. 2017.** Mitochondrial genome of the Christmas tree worm *Spirobranchus giganteus* (Annelida: Serpulidae) reveals a high substitution rate among annelids. *Gene* 605:43–53 DOI [10.1016/j.gene.2016.12.024](https://doi.org/10.1016/j.gene.2016.12.024).
- Shekhovtsov SV, Shipova AA, Bulakhova NA, Berman DI. 2022.** Differentiation within the *Drawida ghilarovi* complex (Moniligastridae: Annelida) revealed by multigene transcriptomic dataset analysis. *European Journal of Soil Biology* 111:103411 DOI [10.1016/j.ejsobi.2022.103411](https://doi.org/10.1016/j.ejsobi.2022.103411).
- Shen W, Le S, Li Y, Hu F. 2016.** SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLOS ONE* 11:e0163962 DOI [10.1371/journal.pone.0163962](https://doi.org/10.1371/journal.pone.0163962).
- Stiller J, Tilic E, Rousset V, Pleijel F, Rouse GW. 2020.** Spaghetti to a tree: a robust phylogeny for Terebelliformia (Annelida) based on transcriptomes, molecular and morphological data. *Biology* 9(4):73 DOI [10.3390/biology9040073](https://doi.org/10.3390/biology9040073).
- Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012.** Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99:349–364 DOI [10.3732/ajb.1100335](https://doi.org/10.3732/ajb.1100335).
- Struck TH. 2011.** Direction of evolution within Annelida and the definition of Pleistoannelida. *Journal of Zoological Systematics and Evolutionary Research* 49:340–345 DOI [10.1111/j.1439-0469.2011.00640.x](https://doi.org/10.1111/j.1439-0469.2011.00640.x).
- Struck TH. 2019.** Phylogeny. In: Purschke G, Böggemann M, Westheide W, eds. *Handbook of Zoology/Annelida: basal groups and pleistoannelida, sedentaria I*. Berlin: De Gruyter, 37–68.
- Struck TH, Golombek A, Hoesel C, Dimitrov D, Elgetany AH. 2023.** Mitochondrial genome evolution in Annelida—a systematic study on conservative and variable gene orders and the factors influencing its evolution. *Systematic Biology* 72:925–945 DOI [10.1093/sysbio/syad023](https://doi.org/10.1093/sysbio/syad023).
- Struck TH, Golombek A, Weigert A, Franke FA, Westheide W, Purschke G, Bleidorn C, Halanych KM. 2015.** The evolution of annelids reveals two adaptive routes to the interstitial realm. *Current Biology* 25:1993–1999 DOI [10.1016/j.cub.2015.06.007](https://doi.org/10.1016/j.cub.2015.06.007).
- Struck TH, Paul C, Hill N, Hartmann S, Hösel C, Kube M, Lieb B, Meyer A, Tiedemann R, Purschke G, Bleidorn C. 2011.** Phylogenomic analyses unravel annelid evolution. *Nature* 471:95–98 DOI [10.1038/nature09864](https://doi.org/10.1038/nature09864).
- Sun Y, Daffe G, Zhang Y, Pons J, Qiu J-W, Kupriyanova EK. 2021.** Another blow to the conserved gene order in Annelida: evidence from mitochondrial genomes of

- the calcareous tubeworm genus *Hydroides*. *Molecular Phylogenetics and Evolution* **160**:107124 DOI [10.1016/j.ympev.2021.107124](https://doi.org/10.1016/j.ympev.2021.107124).
- Sun Y, Kupriyanova EK, Qiu JW. 2012.** COI barcoding of *Hydroides*: a road from impossible to difficult. *Invertebrate Systematics* **26**:539–547 DOI [10.1071/is12024](https://doi.org/10.1071/is12024).
- Taite M, Fernández-Álvarez FÁ, Braid HE, Bush SL, Bolstad K, Drewery J, Mills S, Strugnell JM, Vecchione M, Villanueva R, Voight JR, Allcock AL. 2023.** Genome skimming elucidates the evolutionary history of Octopoda. *Molecular Phylogenetics and Evolution* **182**:107729 DOI [10.1016/j.ympev.2023.107729](https://doi.org/10.1016/j.ympev.2023.107729).
- Tan MH, Gan HM, Bracken-Grissom H, Chan T-Y, Grandjean F, Austin CM. 2021.** More from less: Genome skimming for nuclear markers for animal phylogenomics, a case study using decapod crustaceans. *Journal of Crustacean Biology* **41**:ruab009 DOI [10.1093/jcbiol/ruab009](https://doi.org/10.1093/jcbiol/ruab009).
- Tessler M, De Carle D, Voiklis ML, Gresham OA, Neumann JS, Cios S, Siddall ME. 2018.** Worms that suck: Phylogenetic analysis of Hirudinea solidifies the position of Acanthobdellida and necessitates the dissolution of Rhynchobdellida. *Molecular Phylogenetics and Evolution* **127**:129–134 DOI [10.1016/j.ympev.2018.05.001](https://doi.org/10.1016/j.ympev.2018.05.001).
- Tilic E, Atkinson SD, Rouse GW. 2020a.** Mitochondrial genome of the freshwater annelid *Manayunkia occidentalis* (Sabellida: Fabriciidae). *Mitochondrial DNA Part B: Resources* **5**:3313–3315 DOI [10.1080/23802359.2020.1815604](https://doi.org/10.1080/23802359.2020.1815604).
- Tilic E, Sayyari E, Stiller J, Mirarab S, Rouse GW. 2020b.** More is needed—thousands of loci are required to elucidate the relationships of the flowers of the sea (Sabellida, Annelida). *Molecular Phylogenetics and Evolution* **151**:106892 DOI [10.1016/j.ympev.2020.106892](https://doi.org/10.1016/j.ympev.2020.106892).
- Tilic E, Stiller J, Campos E, Pleijel F, Rouse GW. 2022.** Phylogenomics resolves ambiguous relationships within Aciculata (Errantia, Annelida). *Molecular Phylogenetics and Evolution* **166**:107339 DOI [10.1016/j.ympev.2021.107339](https://doi.org/10.1016/j.ympev.2021.107339).
- Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, Geiger MF, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand AM, Willassen E, Wylter SA, Bouchez A, Borja A. 2019.** DNA barcode reference libraries for the monitoring of aquatic biota in Europe: gap-analysis and recommendations for future work. *Science of the Total Environment* **678**:499–524 DOI [10.1016/j.scitotenv.2019.04.247](https://doi.org/10.1016/j.scitotenv.2019.04.247).
- Weigert A, Bleidorn C. 2016.** Current status of annelid phylogeny. *Organisms Diversity and Evolution* **16**:345–362 DOI [10.1007/s13127-016-0265-7](https://doi.org/10.1007/s13127-016-0265-7).
- Weigert A, Helm C, Meyer M, Nickel B, Arendt D, Hausdorf B, Santos SR, Halanych KM, Purschke G, Bleidorn C, Struck TH. 2014.** Illuminating the base of the annelid tree using transcriptomics. *Molecular Biology and Evolution* **31**:1391–1401 DOI [10.1093/molbev/msu080](https://doi.org/10.1093/molbev/msu080).
- Wheeler TJ, Eddy SR. 2013.** nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**:2487–2489 DOI [10.1093/bioinformatics/btt403](https://doi.org/10.1093/bioinformatics/btt403).
- Williams LG, Karl SA, Rice S, Simon C. 2017.** Molecular identification of polydorid polychaetes (Annelida: Spionidae): is there a quick way to identify pest and alien species? *African Zoology* **52**:105–118 DOI [10.1080/15627020.2017.1313131](https://doi.org/10.1080/15627020.2017.1313131).

- Ye L, Yao T, Lu J, Jiang J, Bai C. 2021.** Mitochondrial genomes of two *Polydora* (Spionidae) species provide further evidence that mitochondrial architecture in the Sedentaria (Annelida) is not conserved. *Scientific Reports* **11**:13552 DOI [10.1038/s41598-021-92994-3](https://doi.org/10.1038/s41598-021-92994-3).
- Zhang Y, Sun J, Rouse GW, Wiklund H, Pleijel F, Watanabe HK. 2018.** Phylogeny, evolution and mitochondrial gene order rearrangement in scale worms (Aphroditiformia, Annelida). *Molecular Phylogenetics and Evolution* **125**:220–231 DOI [10.1016/j.ympev.2018.04.002](https://doi.org/10.1016/j.ympev.2018.04.002).
- Zhao H, Fan S, Aspe NM, Feng L, Zhang Y. 2022.** Characterization of 15 earthworm mitogenomes from northeast china and Its phylogenetic implication (Oligochaeta: Lumbricidae, Moniligastridae). *Diversity* **14**:714 DOI [10.3390/d14090714](https://doi.org/10.3390/d14090714).