

## Research Article

**Cite this article:** Kobayashi G, Itoh H, Nakajima N (2022). 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. <https://doi.org/10.1017/S0025315422001035>

Received: 28 April 2022

Revised: 19 July 2022

Accepted: 31 October 2022

First published online: 6 January 2023

### Key words:

Arenicolida; intertidal zone; Maldanomorpha; mitogenome; polychaetes

### Author for correspondence:

Genki Kobayashi,

E-mail: [genkikobayashi5884@gmail.com](mailto:genkikobayashi5884@gmail.com)

# First mitochondrial genome of a lugworm (Annelida: Arenicolidae) and its phylogenetic position

Genki Kobayashi<sup>1</sup> , Hajime Itoh<sup>2</sup> and Nobuyoshi Nakajima<sup>2</sup> 

<sup>1</sup>Ishinomaki Senshu University, 1 Shinmito Minamisakai, Ishinomaki, Miyagi 986-8580, Japan and <sup>2</sup>National Institute for Environmental Studies, 16-2, Onogawa, Tsukuba, Ibaraki 305-8506, Japan

## Abstract

The annelid mitochondrial genomes (mitogenomes) have been well documented, and phylogenetic analyses based on the mitogenomes provide insightful implications for annelid evolution. However, the mitogenomes of some families remain unknown. Herein, we determined the complete mitogenome of the lugworm *Abarenicola claparedi oceanica* (15,524 bp), representing the first mitogenome from the family Arenicolidae. The gene order of this species is the same as the various lineages in Sedentaria. The maximum likelihood phylogenetic analyses were performed based on six different datasets, including 43 ingroups (oligochaetes, hirudineans, echiurans and closely related polychaetes) and two outgroups (Siboglinidae), namely, aligned and trimmed datasets consisting of the nucleotide sequences of protein-coding genes (PCGs) and rRNAs, and amino acid sequences of PCGs. Phylogenetic analyses based on the nucleotide sequences yielded trees with better support values than those based on the amino acid sequences. Arenicolidae is clustered with Maldanidae in all analyses. Analyses based on nucleotide sequences confirm the monophyly of Terebellidae, which was paraphyletic in recent mitogenomic phylogenetic studies. We also performed the phylogenetic analysis based on the RY-coding of the nucleotide sequences of PCGs only to yield phylogeny with generally low support values. Additional mitogenome sequences of related ingroup species would be needed to comprehensively understand the phylogenetic relationship, which was not present in this study.

## Introduction

Mitochondrial genomes (mitogenomes) are often used for phylogenetic analyses focusing on the phylogenetic relationship of the shallow nodes from inter-familial to intra-species levels (Hirase *et al.*, 2016; Zhang *et al.*, 2018; Tan *et al.*, 2019; Alves *et al.*, 2020; Kobayashi *et al.*, 2021; Ogoh *et al.*, 2021; Sun *et al.*, 2021). The animal mitochondrial DNA is generally circular, comprising 15–20 kilobase pairs (kbp) (Boore, 1999), and is relatively conserved as compared with the quite diverse plant mitogenomes ranging from 6 kbp to  $\geq 11$  Mbp (Mower, 2020). The mitogenomes of animals are usually comprised of 13 protein-coding genes (PCGs; *cox1–3*, *atp6*, *atp8*, *cytb*, *nad1–6* and *nad4l*), 2 rRNAs and 22 tRNAs (Boore, 1999), and this conservation ensures the homology of genes when performing the phylogenetic analysis based on the mitogenomes.

The phylum Annelida consists of over 20,000 described species, such as polychaetes, echiurans, sipunculans, hirudineans and oligochaetes (Capa & Hutchings, 2021). Annelids include some early-branching lineages, Errantia (mobile annelids with well-developed parapodia) and Sedentaria (sessile annelids with less-developed parapodia) (Struck, 2019; Rouse *et al.*, 2022). Sedentaria comprises ecologically and morphologically diverse members from the high-altitude mountains to the hadal zone, such as echiurans, vestimentiferans and clitellates (hirudineans and oligochaetes) (Manca *et al.*, 1998; Paterson *et al.*, 2009). The species of Sedentaria are interesting research subjects due to their evolutionary history. However, several families are not yet included in phylogenomics, and their phylogenetic position remains insufficiently resolved.

A family in Sedentaria, Arenicolidae includes four genera (*Branchiomaldane*, *Abarenicola*, *Arenicolides* and *Arenicola*) and 24 described species (Darbyshire, 2020). Arenicolids are known as ‘lugworms’ and are well-known polychaetes for their eye-catching faecal mounds in the intertidal to shallow sublittoral zones. Generally, arenicolids are deposit feeders, inhabiting silty sand by making burrows (Jumars *et al.*, 2015) that can process sediments up to 280 times their dry body weight per day (Taghon, 1988). The phylogenetic position of Arenicolidae is well known. The sister relationship between Arenicolidae and Maldanidae has been supported by molecular phylogeny (Bleidorn *et al.*, 2005; Struck *et al.*, 2007) and morphology (Bartolomaeus & Meyer, 1997). Arenicolidae is included in the clade of polychaetes (e.g. Terebellida, Caprellidae, Opheliidae and Scalibregmatidae), echiurans and clitellates (hirudineans and oligochaetes) (Struck *et al.*, 2015; Helm *et al.*, 2018; Martín-Durán *et al.*, 2021). Although a robust phylogenetic framework is established, the phylogenetic positions of some families in this clade remain unclear. The mitogenomic phylogeny might



provide implications for the phylogenetic relationships in this clade, although the position of Traviidae and Terebellidae are unknown in the previous analyses based on mitogenomes (Zhong *et al.*, 2011; Nam *et al.*, 2021; Kobayashi *et al.*, 2022a, 2022b). The intra-familial relationship of Arenicolidae was also not entirely understood by phylogenetic analyses based on 16S or 18S rRNA sequences (Bleidorn *et al.*, 2005; Darbyshire, 2017). The mitogenome sequences of arenicolids would therefore help elucidate the intra-familial relationships of Arenicolidae in the future and the inter-familial relationships of related families. However, mitogenomes are not yet available for Arenicolidae.

In this study, we determined and examined the first mitochondrial genome of Arenicolidae. The phylogenetic trees were reconstructed based on the nucleotide and amino acid sequences of the part of Sedentaria, including the mitogenomes of recently published families, which resulted in the most inclusive taxon sampling of ingroup families to elucidate the phylogenetic relationship of *Abarenicola claparedi oceanica* and closely related families.

## Materials and methods

A specimen of *Abarenicola oceanica* was collected from the shallow subtidal zone (~1 m) in Rishiri Island, Hokkaido (45°06'57"N 141°17'18"E) (Kobayashi *et al.*, 2018a). The specimen was fixed and preserved in 70% ethanol and was deposited at the Rishiri Town Museum (RTMANL001). The total DNA was extracted from the body wall tissue using a DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany).

Long PCR was performed for the mitogenome following a method developed by Wu *et al.* (2009). The partial sequence of the 16S rRNA gene sequence was determined using the method described by Kobayashi *et al.* (2018b), using a universal 16S primer set (16Sar: 5'-CGCCTGTTTATCAAAAACAT-3'/16Sbr: CCGGTCTGAAGTCAGATCACGT-3'; Palumbi, 1996). A primer set for long PCR (MalMito16SF: 5'-GAARAAGWTTGGCACCTCGATGTTGGCTTAG-3'/MalMito16SR: 5'-ATTATATGCTACCTTAGCACGGTCAGGRTACCGC-3') was designed based on the dataset consisting of the 16S rRNA gene sequence of arenicolids and maldanids. The PCR mixture for long PCR comprised 2.2 µl of sterilized water, 6.25 µl of KOD One PCR Master Mix (TOYOBO, Osaka, Japan), 1.0 µl of 10 µM forward and reverse primers and 3.3 µl of template DNA solution diluted 10 times with sterilized water. PCR amplification was performed as follows: 60 s at 94°C for an initial denaturation, 35 cycles of 10 s at 98°C and 3 min at 68°C. PCR product of >15 kb in size was checked by electrophoresis in 1% agarose gel at 100 V for 40 min and used as a sample for next-generation sequencing.

A paired-end sequencing (2 × 300 bp) of the mitogenome amplicons was performed using an Illumina MiSeq System (Illumina) at the National Institute for Environmental Studies, Japan. The raw reads were registered at the NCBI repository (DRA013990, PRJDB13458 and SAMD00467651). A mitogenome sequence was assembled by GetOrganelle v1.7.1a (Jin *et al.*, 2020) with the default setting except for -R set to 20 using the 16S rRNA gene sequence as a seed sequence. Assembled contigs and the 16S rRNA gene sequence were manually concatenated, and the complete mitogenome of *A. claparedi oceanica* was obtained. The PCGs and tRNAs were identified using the MITOS2 web server, and then the start and end positions of PCGs were manually checked. The annotated mitogenome sequence was deposited in GenBank through the DNA Data Bank of Japan with the accession number LC707921. Compositional skews were calculated as follows: AT skew = (A - T)/(A + T), GC skew = (G - C)/(G + C). A graphical map of the mitogenome was drawn using CGView (Stothard & Wishart, 2004) and edited using Inkscape 0.91 (<http://www.inkscape.org>).

The maximum likelihood (ML) phylogeny was reconstructed based on the following subsets: (1) nucleotide sequences of 13 PCGs (nucP), (2) nucleotide sequences of 13 PCGs and 2 rRNAs (nucPR), (3) nucleotide sequences of 13 PCGs with ambiguous positions trimmed (nucPt), (4) nucleotide sequences of 13 PCGs and 2 rRNAs with ambiguous positions trimmed (nucPRt), (5) amino acid sequences of 13 PCGs (aaP) and (6) amino acid sequences of 13 PCGs with ambiguous positions trimmed (aaPt). In addition, phylogenetic analysis based on the RY-coding of the nucleotide sequences of PCGs was performed to reduce base compositional biases. This is because RY-coding was judged an effective strategy in inferring the phylogenetic relationship within Terebellida using mitogenomes by Zhong *et al.* (2011). The same group was included in this study. On the other hand, the RY-coding considerably reduces information; therefore, the phylogenetic analysis based on the RY-coding was considered only for discussion on Terebellida. The nucleotide sequences were RY-coded with a Perl program RYcode (<https://github.com/ebraun68/RYcode>).

Each dataset consisted of 43 ingroups (oligochaetes, hirudineans, echiurans and closely related polychaetes) and two outgroups (Siboglinidae). The nucleotide sequences of mitogenomes of some Sedentaria, such as spionids and serpulids, are quite different from those of other Sedentaria (e.g. Sun *et al.*, 2021), and they are not suitable for phylogenetic analyses. We therefore focused on the part of Sedentaria, which is phylogenetically related to Arenicolidae, and chose siboglinids as outgroups, referring to the current understanding of the phylogenetic relationships of Sedentaria (Struck, 2019; Rouse *et al.*, 2022). The sequences were obtained from GenBank except for the newly determined sequence of *A. claparedi oceanica* (Table 1). The ML analysis was conducted with IQ-TREE v1.6.12 (Nguyen *et al.*, 2014) using 1000 ultrafast bootstrap replicates. 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 (Kalyanamoorthy *et al.*, 2017), except for JC2 for the RY-coding dataset, implemented in IQ-TREE (Supplementary data). The amino acid sequences were translated using invertebrate mitochondrial code with SeqKit (Shen *et al.*, 2016). The PCGs were aligned with MAFFT v7 (Katoh & Standley, 2013) using the default option for PCGs. Q-INS-I option, which considers the secondary structure, was used for rRNAs in the MAFFT web server (<https://mafft.cbrc.jp/alignment/server/index.html>). FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) was used to illustrate phylogenetic trees.

The gene orders of the ingroup were mapped on the cladogram of nucPt to compare gene orders, including PCGs, rRNAs and tRNAs. Species whose mitogenomes were partially determined were excluded from mapping. The *trnR* of *Perionyx excavatus* (EF494507) was not annotated and detected with ARAGORN (base 5856 to 5915) (Laslett & Canback, 2004).

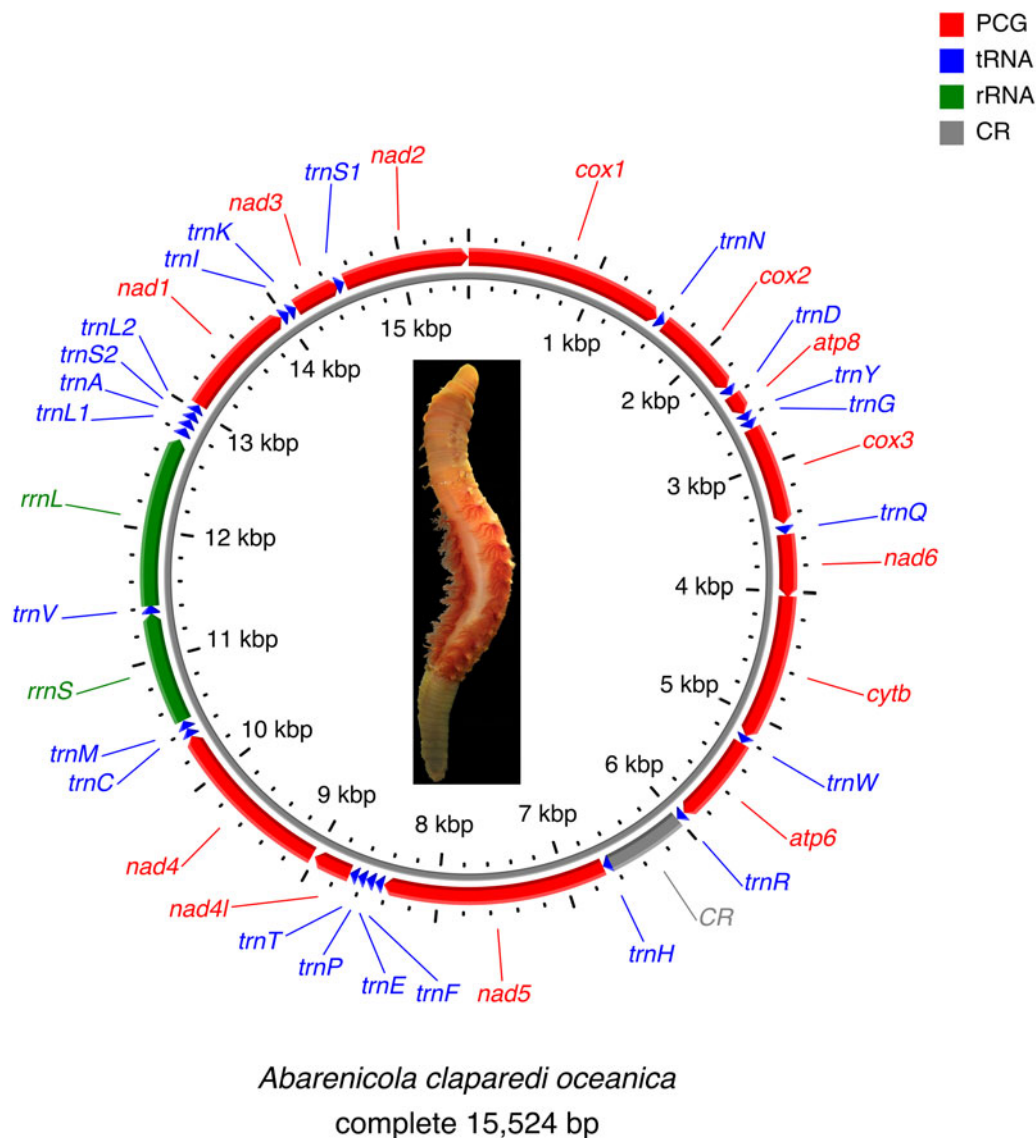
## Results

The mitogenome of *Abarenicola claparedi oceanica* was circular and consisted of 15,524 bp. The mitogenome included 13 PCGs (*atp6* and *atp8*, *cox1-cox3*, *cytb*, *nad1-nad6* and *nad4l*), 22 tRNA genes (one for each of the amino acids except for *trnL* and *trnS*) and 2 rRNA genes (*rrnS* or 12S rRNA and *rrnL* or 16S rRNA) (Figure 1 and Table 2). All determined genes were encoded on the '+' strand (Figure 1). Non-coding region (621 bp) between *trnR* and *trnH* with high A + T content (79.6%) was a putative control region. GC skew of all genes was negative, indicating that C outnumbered G, while AT skew was largely negative or almost zero except for *cox2*, *rrnL* and *rrnS* (Table 3). The gene order, including the tRNAs of *A. claparedi*

**Table 1.** Mitogenome sequences used in this study. Bold indicates the sequence identified in this study

Group	Classification <sup>a</sup>	Family	Species <sup>b</sup>	GenBank accession no.		
Oligochaetes	Moniligastrida	Moniligastridae	<i>Drawida japonica</i>	KM199288		
			Crassiclitellata	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 <sup>c</sup>	<i>Pontoscolex corethrurus</i>			KT988053	
		Hirudineans	Acanthobdellida	Acanthobdellidae	<i>Acanthobdella peledina</i>	MZ562997
Arhynchobdellida	Haemadipsidae				<i>Haemadipsa crenata</i>	MW711186
			Hirudiniformes	Hirudinidae	<i>Hirudo medicinalis</i>	KU672396
					<i>'Hirudo nipponia'</i>	KC667144
			<i>Hirudo verbana</i>	KU672397		
			<i>'Hirudinaria manillensis'</i>	KC688268		
			<i>Whitmania acranulata</i>	KM655838		
			<i>Whitmania laevis</i>	KC688269		
			<i>Whitmania laevis</i>	KM655839		
Erpobdelliformes	Erpobdellidae		<i>Erpobdella japonica</i>	MF358688		
		<i>'Erpobdella octoculata'</i>	KC688270			
Rhynchobdellida	Glossiphoniiformes	Glossiphoniidae	<i>Haementeria officinalis</i>	LT159848		
			<i>Placobdella lamothei</i>	LT159849		
			<i>Placobdella parasitica</i>	LT159850		
			Oceanobdelliformes	Piscicolidae	<i>Codonobdella</i> sp.	MZ202177
					<i>Zeylanicobdella arugamensis</i>	KY474378
					<i>Ozobranchus jantseanus</i>	KY861060
Polychaetes	Terebellida	Travisiidae	<i>Travisia sanrikuensis</i>	LC677172		
			Ampharetidae	<i>Auchenoplax crinita</i>	FJ976041	
		<i>Decemunciger</i> sp.		KY742027		
		<i>Eclysippe vanelli</i>		EU239687		
		<i>Paralvinella sulfincola</i>		FJ976042		
		<i>Melinna cristata</i>		MW542504		
		<i>Terebellides stroemii</i>		EU236701		
		<i>Neoamphitrite affinis</i>		MZ326700		
		<i>Pista cristata</i>		EU239688		
		<i>Thelepus plagiotoma</i>	MW557377			
		Pectinariidae	<i>Pectinaria gouldii</i>	FJ976040		
		<b>Arenicolida</b>	<b>Arenicolidae</b>	<b><i>Abarenicola claparedi oceanica</i></b>	<b>LC707921</b>	
		Capitellida	Cepitellidae	<i>Maldanidae</i>	<i>Clymenella torquata</i>	AY741661
<i>Notomastus</i> sp.	LC661358					
<i>Opheliidae</i>	<i>Armandia</i> sp.			LC661359		
Echiurans	Thalassematidae <sup>d</sup>	<i>Urechis caupo</i>	AY619711			
		<i>Urechis unicinctus</i>	EF656365			
Outgroup	-	Siboglinidae	<i>Lamellibrachia luyesi</i>	KJ789163		
			<i>Sclerolinum brattstromi</i>	KJ789167		

<sup>a</sup>The classifications are after Schmelz *et al.* (2021) for oligochaetes, Tessler *et al.* (2018) for hirudineans and Struck (2019) for polychaetes. <sup>b</sup>Quotation marks indicate that species were possibly erroneously identified and should be species of *Whitmania* (Ye *et al.*, 2015). <sup>c</sup>James (2012). <sup>d</sup>Goto *et al.* (2020).



**Fig. 1.** Gene map of the mitochondrial genome of *Abarenicola claparedi oceanica*. The photograph shows *A. claparedi oceanica*. Red: protein-coding genes, blue: transfer RNAs, green: ribosomal RNAs, grey: control region.

*oceanica*, was the same as that observed in various lineages in Sedentaria (Bleidorn *et al.*, 2009; Ocegüera-Figueroa *et al.*, 2016) (Figures 2 and 3).

The analyses based on each dataset (nucleotide sequences of PCGs, nucleotide sequences of PCGs and rRNAs and amino acids) yielded different tree topologies. In contrast, the exclusion of ambiguously aligned positions had little effect on the topology (Figures 2 and S1–S5). In all phylogenetic analyses, *A. claparedi oceanica* was clustered with *Clymenella torquata* (Maldanidae) (bootstrap values, BS = 100% in nucleotide datasets and 85%/95% in aaP/aaPt). The monophyly of the Terebellida, Arenicolida and Traviisiidae clade was supported in all analyses, although the phylogenetic relationships in this clade differed between analyses. *Travisia sanrikuensis* was clustered with Arenicolida in nucP (BS = 84%), nucPt (BS = 82%) and aaP/aaPt (BS < 50%), whereas this species was sister to the Arenicolida and Terebellida clade in nucPR (BS = 79%) and nucPRt (BS = 97%) and was clustered with Capitellidae and Opheliidae clade in the phylogeny based on the RY-coding (BS = 94%) (Figures 4 and S6). Pectinariidae was sister to the other Terebellida species in all analyses except for the analysis based on the RY-coding. Terebellidae was not

monophyletic in the analyses based on the amino acid sequences. However, the support values for the position of terebellid species were low, *Thelepus plagiostoma* was clustered with the Ampharetidae and Alvinellidae clade (BS = 71% in aaP and 69% in aaPt) and the other two terebellids (*Neoamphitrite affinis* and *Pista cristata*) were sister to the Melinnidae and Trichobranchidae clade (BS = 70% in aaP and 68% in aaPt) (Figure 4C). Monophyletic Terebellidae has clustered with the Melinnidae and Trichobranchidae clade. The clade, including these three families, was clustered with Ampharetidae and Alvinellidae in the analyses based on nucleotide sequences with generally high support values (BS = 76–97%). A phylogenetic analysis based on the RY-coding resulted in a different topology from the other analyses (Figure 5 and S6). The BS values were generally low in Terebellida except for the Ampharetidae and Alvinellidae clade (BS = 100%) and the *N. affinis* and *P. cristata* clade (BS = 98%).

The phylogenetic relationships of other clades (echiurans + Capitellidae + Opheliidae, hirudineans and oligochaetes) remained broadly stable between analyses. Thalassematidae (echiurans) was sister to the Capitellidae and Opheliidae clade (BS = 96–100%).

**Table 2.** Summary of the mitogenome of *Abarenicola claparedi oceanica*

Gene	Start	Stop	Length	Start/stop codons
<i>cox1</i>	1	1557	1557	ATG/TAA
<i>trnN</i> (GTT)	1562	1630	69	
<i>cox2</i>	1631	2323	693	ATG/TAG
<i>trnD</i> (GTC)	2323	2391	69	
<i>atp8</i>	2393	2563	171	ATG/TAA
<i>trnY</i> (GTA)	2562	2626	65	
<i>trnG</i> (TCC)	2626	2693	68	
<i>cox3</i>	2694	3473	780	ATG/TAA
<i>trnQ</i> (TTG)	3489	3556	68	
<i>nad6</i>	3557	4039	483	ATG/TAA
<i>cytb</i>	4032	5171	1140	ATG/TAG
<i>trnW</i> (TCA)	5170	5235	66	
<i>atp6</i>	5236	5937	702	ATG/TAA
<i>trnR</i> (TCG)	5941	6007	67	
CR	6008	6626	619	
<i>trnH</i> (GTG)	6627	6691	65	
<i>nad5</i>	6692	8425	1734	ATG/TAA
<i>trnF</i> (GAA)	8430	8496	67	
<i>trnE</i> (TTC)	8499	8568	70	
<i>trnP</i> (TGG)	8570	8636	67	
<i>trnT</i> (TGT)	8637	8702	66	
<i>nad4l</i>	8703	9002	300	ATG/TAA
<i>nad4</i>	8996	10,354	1358	ATG/TAG
<i>trnC</i> (GCA)	10,356	10,421	66	
<i>trnM</i> (CAT)	10,422	10,488	67	
<i>rrnS</i>	10,490	11,361	872	
<i>trnV</i> (TAC)	11,357	11,428	72	
<i>rrnL</i>	11,417	12,739	1323	
<i>trnL1</i> (TAG)	12,772	12,839	68	
<i>trnA</i> (TGC)	12,845	12,911	67	
<i>trnS2</i> (TGA)	12,912	12,978	67	
<i>trnL2</i> (TAA)	12,980	13,043	64	
<i>nad1</i>	13,044	13,977	934	ATG/T(AA)
<i>trnI</i> (GAT)	13,978	14,047	70	
<i>trnK</i> (TTT)	14,048	14,114	67	
<i>nad3</i>	14,115	14,486	372	ATG/TAA
<i>trnS1</i> (TCT)	14,467	14,536	70	
<i>nad2</i>	14,537	15,524	987	ATG/T(AA)

The position of Thalamematidae was not resolved in the phylogenetic analysis based on the RY-coding, whereas Capitellidae and Opheliidae were monophyletic with full support (Figure S6). In the hirudinean clade, Acanthobdellidae was sister to the clade consisting of monophyletic Rhynchobdellida (Glossiphoniidae was clustered with the Piscicolidae and Ozobranchiidae clade) (BS = 79–100%) and monophyletic Arhynchobdellida (Erpobdellidae and Hirudinidae) (BS = 100%). In oligochaetes, Rhinodrilidae was clustered with Moniligastridae and Crassicitellata (Megascolecidae and Lumbricidae) clade (BS = 70–100%).

## Discussion

This is the first study to determine the mitogenome sequence of Arenicolidae using *Abarenicola claparedi oceanica* and perform phylogenetic analyses using closely related families registered in GenBank. Using previously published mitogenome data (Bolbat *et al.*, 2021; Nam *et al.*, 2021; Kobayashi *et al.*, 2022a, 2022b), mitogenomic phylogenetic trees were reconstructed with the most inclusive taxon sampling of the part of the Sedentaria families (oligochaetes, hirudineans, echiurans and some polychaetes). Analyses based on the nucleotide sequences confirmed the monophyly of Terebellidae, which was paraphyletic in analyses based on the amino acid sequences and previous mitogenomic phylogenies.

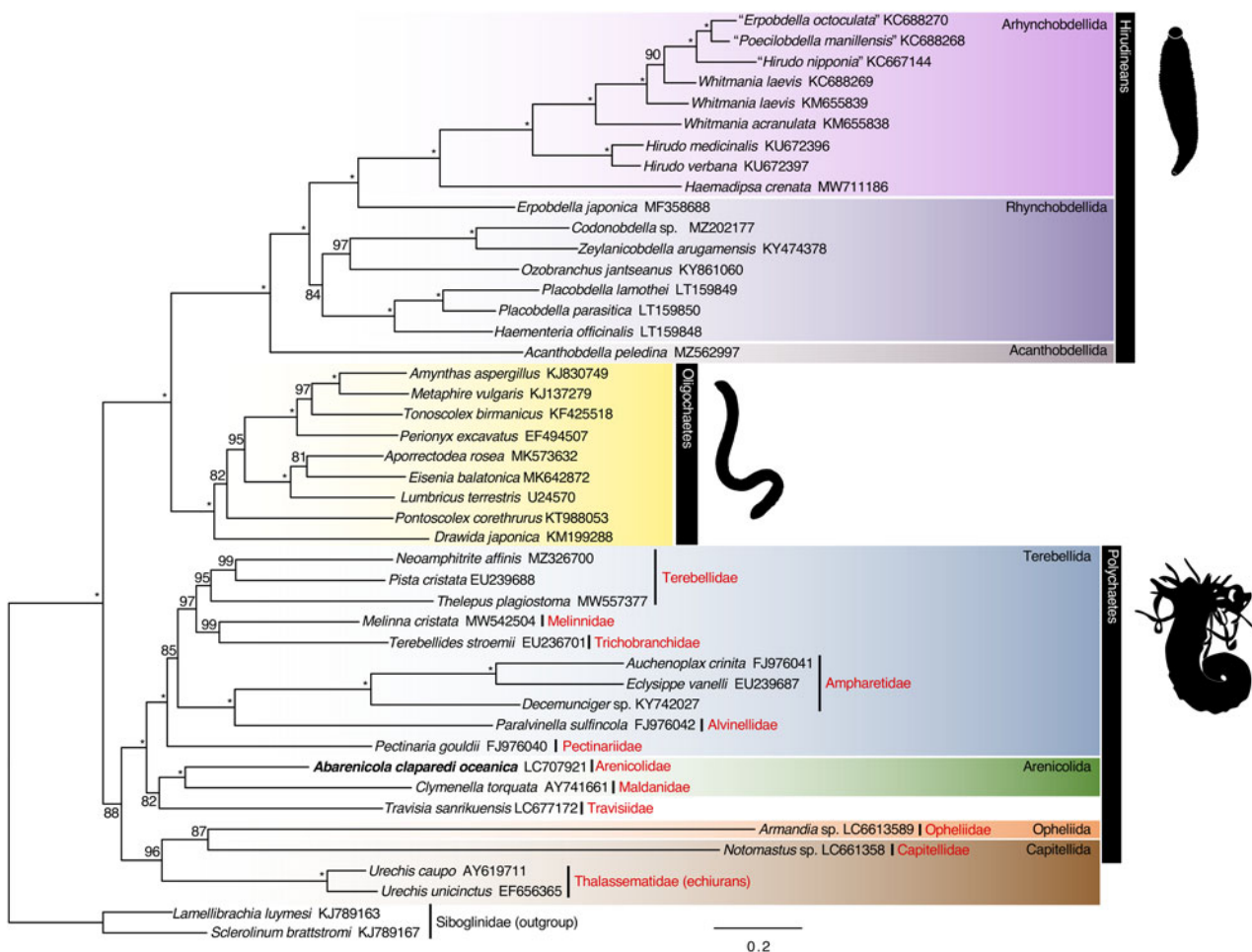
The mitogenome of *Abarenicola claparedi oceanica* is similar to the mitogenome of several lineages of Sedentaria with respect to the genes and gene order, including the tRNAs. Although this gene order is conserved in the lineages of Sedentaria, such as Megascolecidae and Moniligastridae (oligochaetes), Erpobdellidae and Ozobranchiidae (hirudineans) and Siboglinidae (Bleidorn *et al.*, 2009; Ocegüera-Figueroa *et al.*, 2016), the present study is the first in the Terebellida + Arenicolida + Traviisiidae clade. This finding supports that this gene order was conserved in the ancestral lineage of Sedentaria because a common origin and the subsequent changes of the gene order in each lineage are more parsimonious than obtaining the same gene order in multiple lineages independently. The gene order of *Clymenella torquata* (Maldanidae), which was clustered with *A. claparedi oceanica* in the phylogenetic analysis in this study, differed in *trnK* located between the *trnR* and *trnH* (Jennings and Halanych, 2005). The gene order of *C. torquata* may therefore be uniquely obtained in this species or the ancestor of Maldanidae.

The sister relationship of Arenicolidae and Maldanidae established in this study has been well understood (Bartolomaeus & Meyer, 1997; Bleidorn *et al.*, 2005; Struck *et al.*, 2007). The phylogenetic relationship of Terebellida, namely, the Ampharetidae + Alvinellidae clade, the Trichobranchiidae + Terebellidae + Melinnidae clade and Pectinariidae, showed similar results to the transcriptome analyses (Stiller *et al.*, 2020) except for Terebellidae, which was sister to the Trichobranchiidae and Melinnidae clade in this study. However, Trichobranchiidae was sister to the Terebellidae and Melinnidae clade in Stiller *et al.* (2020). Although the close relationship between Terebellidae and Trichobranchiidae recovered in mitogenomic phylogeny was suggested to be due to compositional bias in the mitogenomes of the Ampharetidae and Alvinellidae (Zhong *et al.*, 2011), phylogeny based on transcriptomes also supported that Trichobranchiidae is more close to Terebellidae than Ampharetidae or Alvinellidae (Stiller *et al.*, 2020). Zhong *et al.* (2011) stated that the compositional bias in the mitogenome of Ampharetidae and Alvinellidae influences the phylogenetic relationships of Terebellidae and Trichobranchiidae. They conclude that no approach completely ameliorated the influence of bias, although the phylogenetic analysis based on the RY-coding was suggested as the most effective strategy in their analyses. Unfortunately, the support values of the phylogenetic analysis based on the RY-coding were generally low in the present study. This method did not sufficiently resolve the phylogenetic relationship in Terebellida. The paraphyletic status of Terebellidae in mitogenomic phylogeny was suggested by the phylogeny based on the nucleotide sequences of PCGs in Nam *et al.* (2021) and by the phylogeny based on the amino acid sequences in Kobayashi *et al.* (2022b). In the present analyses, although Terebellidae was paraphyletic in analyses based on amino acid sequences, it was monophyletic with high support values in analyses based on nucleotide sequences.

A previous study on the mitogenomic phylogeny indicated a close relationship between Traviisiidae and the Terebellida and

**Table 3.** Nucleotide composition (%) of 13 protein-coding genes, ribosomal RNAs, control region and the skewness of *Abarenicola claparedi oceanica*

	Length	A	C	G	T	A + T	AT-skew	GC-skew
Mitogenome	15,524	31.9	22.4	13.5	32.2	64.1	0.00	-0.25
<i>atp6</i>	702	27.2	24.8	12.8	35.2	62.4	-0.13	-0.32
<i>atp8</i>	171	33.3	26.9	7.0	32.7	66.1	0.01	-0.59
<i>cox1</i>	1557	28.8	22.8	15.9	32.5	61.3	-0.06	-0.18
<i>cox2</i>	693	32.5	24.4	15.2	28.0	60.5	0.07	-0.23
<i>cox3</i>	780	26.7	24.6	17.3	31.4	58.1	-0.08	-0.17
<i>cytb</i>	1140	30.4	21.8	14.1	33.7	64.1	-0.05	-0.21
<i>nad1</i>	934	30.2	23.3	12.8	33.6	63.8	-0.05	-0.29
<i>nad2</i>	987	31.8	23.4	9.7	35.1	66.9	-0.05	-0.41
<i>nad3</i>	372	26.1	22.3	13.7	38.0	64.0	-0.19	-0.24
<i>nad4</i>	1358	32.7	23.2	10.9	33.3	65.9	-0.01	-0.36
<i>nad4l</i>	300	32.3	21	13.7	33.0	65.3	-0.01	-0.21
<i>nad5</i>	1734	30.3	24.8	11.2	33.7	64.0	-0.05	-0.38
<i>nad6</i>	483	29.0	22.6	9.5	38.9	68.0	-0.15	-0.41
<i>rrnL</i>	1323	37.1	21.2	14.7	27.0	64.0	0.16	-0.18
<i>rrnS</i>	872	35.4	21.4	17.2	26.0	61.4	0.15	-0.11
CR	619	42.3	12.4	7.9	37.3	79.6	0.06	-0.22



**Fig. 2.** Maximum likelihood phylogeny of a subset of Sedentaria based on the dataset, including the nucleotide sequences of 13 protein-coding genes (PCGs) after ambiguous positions were excluded (10,938 characters; nucPt). Numbers above the branches represent the maximum likelihood bootstrap values (BS). Asterisks indicate the branches with BS = 100%. *Abarenicola claparedi oceanica*, whose nucleotide sequence is newly obtained, is shown in bold. Illustrations: *Pontoscolex corethrurus* (oligochaetes) and an unassigned hirudinean, obtained from phylopic.org, and a schematic illustration of Terebellidae (polychaetes).

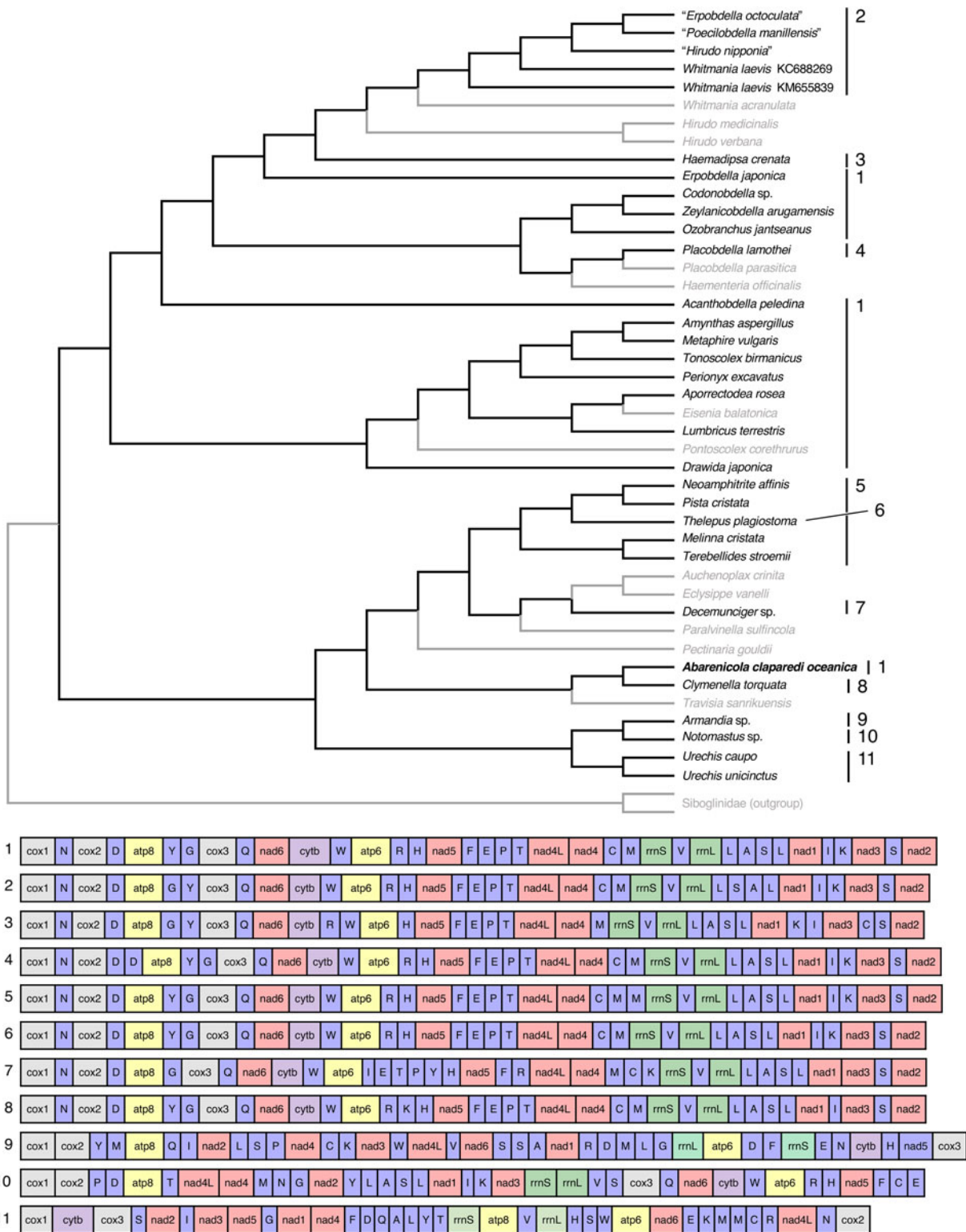
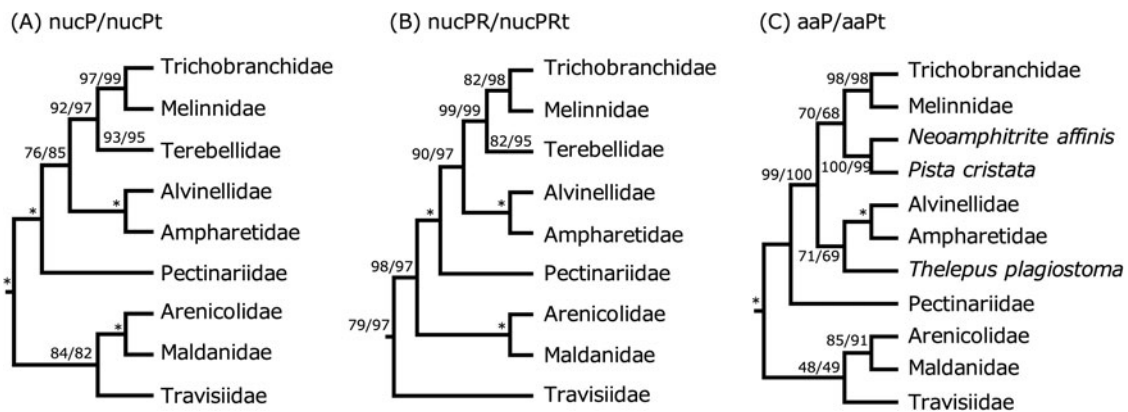


Fig. 3. Comparison of the gene orders of the ingroup based on the phylogenetic tree shown in Figure 2. The species with partial mitogenomes are not related (species names and branches shown in grey). *Abarenicola claparedi oceanica*, whose nucleotide sequence is newly obtained, is shown in bold.

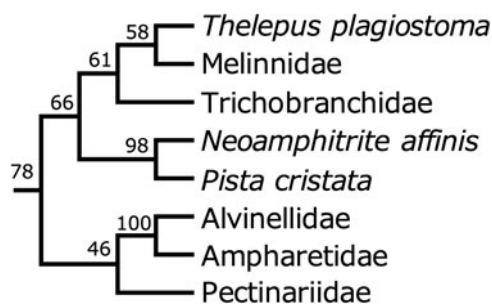
Arenicolida clade (Kobayashi *et al.*, 2022a). However, the phylogenetic position was not sufficiently clarified as the position of Traviisiidae was poorly supported. The precise placement of Traviisiidae was still ambiguous in the present analyses, which are more inclusive than the previous mitogenomic phylogenetic studies due to the addition of at least four polychaete families (Arenicolidae, Capitellidae, Melinnidae and Opheliidae) to the dataset of Kobayashi *et al.* (2022b). Scalibregmatidae was represented as the sister family of Traviisiidae based on the molecular phylogenetic

analyses (Persson & Pleijel, 2005; Paul *et al.*, 2010; Law *et al.*, 2014; see Blake & Maciolek, 2020 for review on taxonomic status of Traviisiidae), and Scalibregmatidae showed a sister relationship to the Arenicolida and Terebellida clade in the phylogenetic analysis based on transcriptomes (Helm *et al.*, 2018; Martín-Durán *et al.*, 2021). These results suggest that Traviisiidae is one of the early-branching lineages in the Terebellida and Arenicolida clade.

The results of this study support the monophyly of both Arhynchobdellida and Rhynchobdellida. The phylogenetic



**Fig. 4.** Schematic cladograms of the phylogenetic relationship between the families in the Terebellida + Arenicolida + Traviidae clade as suggested by the maximum likelihood phylogenetic analyses in this study. (A) The nucleotide sequence of protein-coding genes (nucP) and after trimming ambiguously aligned positions (nucPt), (B) nucleotide sequences of PCGs and ribosomal RNAs (nucPR) and after trimming (nucPRT) and (C) amino acid sequences of PCGs (aaP) and after trimming (aaPt). Numbers above the branches represent the maximum likelihood bootstrap values (BS). The order of BS corresponds to each heading. Figures 2 and S1–S5 show the whole phylogenetic trees. BS on the branch of Terebellidae in (A) and (B) indicates BS for three terebellid species. Species names show the paraphyletic Terebellidae in (C). Asterisks indicate the branches with BS=100%.



**Fig. 5.** Schematic cladogram of the phylogenetic relationship of Terebellida suggested by the maximum likelihood phylogeny based on the RY-coding of the nucleotide sequences of protein-coding genes (PCGs). Species names represent the paraphyletic Terebellidae. The numbers above the branches represent the maximum likelihood bootstrap values (BS).

relationship of hirudineans was contentious as the results were not congruent (Martin, 2001; Rousset et al., 2008; Tessler et al., 2018). For example, the Rhynchobdellida (proboscis-bearing leeches) was paraphyletic; Glossiphoniidae (Rhynchobdellida) was clustered with Arhynchobdellida (leeches without proboscis), not with Rhynchobdellida (Ozobranchidae and Piscicolidae) (Tessler et al., 2018). Moreover, the mitogenomic studies with limited taxon sampling of annelids showed the paraphyletic status of Rhynchobdellida (Wang et al., 2018; Sosa-Jiménez et al., 2020). These results were consistent with the paraphyly of Rhynchobdellida. Contrarily, the phylogenetic analysis based on the mitogenomes (Bolbat et al., 2020, 2021), anchored hybrid enrichment (Phillips et al., 2019) and transcriptomes (Erséus et al., 2020) with more taxon sampling revealed the sister relationship between Glossiphoniidae and Oceanobdelliformes (only Piscicolidae in Erséus et al. (2020)) and thus the monophyly of Rhynchobdellida. However, the number of the Arhynchobdellida families is still limited in the phylogenomic studies therefore further taxon sampling would be required to conclude the monophyly of hirudinean orders.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0025315422001035>

**Acknowledgements.** We are grateful to Hirokazu Abe (Ishinomaki Senshu University) and Shinri Tomioka (Rishiri Town Museum) for their help in collecting the specimen; Tomoyuki Nakano and Ryutaro Goto (Kyoto University) for their help in using the facilities at the Seto Marine Biological Laboratory,

the Field Science Education and Research Center, Kyoto University; and two anonymous reviewers for their invaluable comments on the earlier draft.

**Financial support.** This work was supported by the Project for the Promotion of Environmental Genomics Studies by NIES (grant number 1620AQ007) and by JSPS KAKENHI (grant number JP22K15174).

**Data availability.** The mitogenome sequence of *Abarenicola claparedi oceanica* is available in GenBank under the accession no. LC707921. The raw reads were registered at the NCBI repository (DRA013990, PRJDB13458 and SAMD00467651).

## References

- Alves PR, Halanych KM and Santos CSG (2020) The phylogeny of Nereididae (Annelida) based on mitochondrial genomes. *Zoologica Scripta* **49**, 366–378.
- Bartolomaeus T and Meyer K (1997) Development and phylogenetic significance of hooked setae in Arenicolidae (Polychaeta, Annelida). *Invertebrate Biology* **116**, 227.
- Blake JA and Maciolek NJ (2020) Traviidae Hartmann-Schröder, 1971, new family status. In Purschke G, Böggemann M and Westheide W (eds), *Handbook of Zoology/Annelida: Basal Groups and Pleistoannelida, Sedentaria II*. Berlin: De Gruyter, pp. 302–311.
- Bleidorn C, Podsiadlowski L, Zhong M, Eckhaut I, Hartmann S, Halanych KM and Tiedemann R (2009) On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evolutionary Biology* **9**, 1–11.
- Bleidorn C, Vogt L and Bartolomaeus T (2005) Molecular phylogeny of lugworms (Annelida, Arenicolidae) inferred from three genes. *Molecular Phylogenetics and Evolution* **34**, 673–679.
- Bolbat A, Matveenko E, Dzyuba E and Kaygorodova I (2021) The first mitochondrial genome of *Codonobdella* sp. (Hirudinea, Piscicolidae), an endemic leech species from Lake Baikal, Russia and reassembly of the *Piscicola geometra* data from SRA. *Mitochondrial DNA Part B* **6**, 3112–3113.
- Bolbat A, Vasiliev G and Kaygorodova I (2020) The first mitochondrial genome of the relic *Acanthobdella peledina* (Annelida, Acanthobdellida). *Mitochondrial DNA Part B: Resources* **5**, 3300–3301.
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research* **27**, 1767–1780.
- Capa M and Hutchings P (2021) Annelid diversity: historical overview and future perspectives. *Diversity* **13**, 129.
- Darbyshire T (2017) A re-evaluation of the *Abarenicola assimilis* group with a new species from the Falkland Islands and key to species. *Journal of the Marine Biological Association of the United Kingdom* **97**, 897–910.
- Darbyshire T (2020) Arenicolidae Johnston, 1835. *Handbook of Zoology/Annelida: Pleistoannelida, Sedentaria III and Errantia I*. Berlin: De Gruyter, pp. 163–185.



- Erséus C, Williams BW, Horn KM, Halanych KM, Santos SR, James SW, Creuzé M and Anderson FE (2020) Phylogenomic analyses reveal a Palaeozoic radiation and support a freshwater origin for clitellate annelids. *Zoologica Scripta* **49**, 614–640.
- Goto R, Monnington J, Sciberras M, Hirabayashi I and Rouse GW (2020) Phylogeny of Echiura updated, with a revised taxonomy to reflect their placement in Annelida as sister group to Capitellidae. *Invertebrate Systematics* **34**, 101–111.
- Helm C, Beckers P, Bartolomaeus T, Drukewitz SH, Kourtesis I, Weigert A, Purschke G, Worsaae K, Struck TH and Bleidorn C (2018) Convergent evolution of the ladder-like ventral nerve cord in Annelida. *Frontiers in Zoology* **15**, 36.
- Hirase S, Takeshima H, Nishida M and Iwasaki W (2016) Parallel mitogenome sequencing alleviates random rooting effect in phylogeography. *Genome Biology and Evolution* **8**, 1267–1278.
- James SW (2012) Re-erection of Rhinodrillidae Benham, 1890, a senior synonym of Pontoscolecidae James, 2012 (Annelida: Clitellata). *Zootaxa* **68**, 67–68.
- Jennings RM and Halanych KM (2005) Mitochondrial genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): evidence for conserved gene order in Annelida. *Molecular Biology and Evolution* **22**, 210–222.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS and Li DZ (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**, 241.
- Jumars PA, Dorgan KM and Lindsay SM (2015) Diet of worms emended: an update of polychaete feeding guilds. *Annual Review of Marine Science* **7**, 497–520.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A and Jermin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–589.
- Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.
- Kobayashi G, Abe H, Itoh H, Tomioka S and Kojima S (2018a) New records of two arenicolid species (Annelida: Arenicolidae) from Rishiri Island, northern Japan with a brief review of previous records of Japanese arenicolids. *Rishiri Studies* **37**, 95–100. [In Japanese with English abstract.]
- Kobayashi G, Goto R, Takano T and Kojima S (2018b) Molecular phylogeny of Maldanidae (Annelida): multiple losses of tube-capping plates and evolutionary shifts in habitat depth. *Molecular Phylogenetics and Evolution* **127**, 332–344.
- Kobayashi G, Itoh H, Fukuda H and Kojima S (2021) The complete mitochondrial genome of the sand bubbler crab *Scopimera globosa* and its phylogenetic position. *Genomics* **113**, 831–839.
- Kobayashi G, Itoh H and Kojima S (2022a) Mitogenome of a stink worm (Annelida: Traviidae) includes degenerate group II intron that is also found in five congeneric species. *Scientific Reports* **12**, 4449.
- Kobayashi G, Itoh H and Nakajima N (2022b) First mitochondrial genomes of Capitellidae and Opheliidae (Annelida) and their phylogenetic placement. *Mitochondrial DNA Part B* **7**, 577–579.
- Laslett D and Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* **32**, 11–16.
- Law CJ, Dorgan KM and Rouse GW (2014) Relating divergence in polychaete musculature to different burrowing behaviors: a study using Opheliidae (Annelida). *Journal of Morphology* **275**, 548–571.
- Manca M, Ruggiu D, Panzani P, Asioli A, Mura G and Nocentini AM (1998) Report on a collection of aquatic organisms from high mountain lakes in Khumbu Valley (Nepalese Himalayas). *Memorie Dell'Istituto Italiano Di Idrobiologia* **57**, 77–98.
- Martin P (2001) On the origin of the Hirudinea and the demise of the Oligochaeta. *Proceedings of the Royal Society B: Biological Sciences* **268**, 1089–1098.
- Martín-Durán JM, Vellutini BC, Marlétaz F, Cetrangolo V, Cvetic N, Thiel D, Henriët 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 and Hejnol A (2021) Conservative route to genome compaction in a miniaturized annelid. *Nature Ecology and Evolution* **5**, 231–242.
- Mower JP (2020) Variation in protein gene and intron content among land plant mitogenomes. *Mitochondrion* **53**, 203–213.
- Nam S-E, Park HS, Kim SA, Kim B-M and Rhee J-S (2021) The complete mitochondrial genome of the terebellid polychaete *Thelepus plagiostoma* (Terebellida; Terebellidae). *Mitochondrial DNA Part B* **6**, 3114–3116.
- Nguyen L-T, Schmidt HA, von Haeseler A and Minh BQ (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–274.
- Oceguera-Figueroa A, Manzano-Marín A, Kvist S, Moya A, Siddall ME and Latorre A (2016) Comparative mitogenomics of leeches (Annelida: Clitellata): genome conservation and placobdella-specific trnD gene duplication. *PLoS ONE* **11**, e0155441.
- Ogoh K, Futahashi R and Ohmiya Y (2021) Intraspecific nucleotide polymorphisms in seven complete sequences of mitochondrial DNA of the luminous ostracod, *Vargula hilgendorffii* (Crustacea, Ostracoda). *Gene Reports* **23**, 101074.
- Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In Hillis DM, Mortiz C and Mable BK (eds), *Molecular Systematics*. Sunderland, MA: Sinauer & Associates, pp. 205–247.
- Paterson GLJ, Glover AG, Barrio Froján CRS, Whitaker A, Budaeva N, Chimonides J and Doner S (2009) A census of abyssal polychaetes. *Deep-Sea Research Part II: Topical Studies in Oceanography* **56**, 1739–1746.
- Paul C, Halanych KM, Tiedemann R and Bleidorn C (2010) Molecules reject an opheliid affinity for *Travisia* (Annelida). *Systematics and Biodiversity* **8**, 507–512.
- Persson J and Pleijel F (2005) On the phylogenetic relationships of *Axiokibbutia*, *Travisia* and *Scalibregmatidae* (Polychaeta). *Zootaxa* **14**, 1–14.
- Phillips AJ, Dornburg A, Zapfe KL, Anderson FE, James SW, Erséus C, Moriarty Lemmon E, Lemmon AR and Williams BW (2019) Phylogenomic analysis of a putative missing link sparks reinterpretation of leech evolution. *Genome Biology and Evolution* **11**, 3082–3093.
- Rouse GW, Pleijel F and Tilic E (2022) *Annelida*. Oxford: Oxford University Press.
- Rousset V, Plaisance L, Erséus C, Siddall ME and Rouse GW (2008) Evolution of habitat preference in Clitellata (Annelida). *Biological Journal of the Linnean Society* **95**, 447–464.
- Schmelz RM, Erséus C, Martin P, Van Haaren T and Timm T (2021) A proposed order-level classification in Oligochaeta (Annelida, Clitellata). *Zootaxa* **5040**, 589–597.
- Shen W, Le S, Li Y and Hu F (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962.
- Sosa-Jiménez VM, Torres-Carrera G, Manzano-Marín A, Kvist S and Oceguera-Figueroa A (2020) Mitogenome of the blood feeding leech *Haementeria acucueyetzin* (Hirudina: Glossiphoniidae) from Tabasco, Mexico. *Mitochondrial DNA Part B: Resources* **5**, 3328–3330.
- Stiller J, Tilic E, Rousset V, Pleijel F and Rouse GW (2020) Spaghetti to a tree: a robust phylogeny for Terebelliformia (Annelida) based on transcriptomes, molecular and morphological data. *Biology* **9**, 73.
- Stothard P and Wishart DS (2004) Circular genome visualization and exploration using CGView. *Bioinformatics (Oxford, England)* **21**, 537–539.
- Struck TH (2019) Phylogeny. In Purschke G, Böggemann M and Westheide W (eds), *Handbook of Zoology/Annelida: Basal Groups and Pleistoannelida, Sdentaria I*. Berlin: De Gruyter, pp. 37–68.
- Struck TH, Golombek A, Weigert A, Franke FA, Westheide W, Purschke G, Bleidorn C and Halanych KM (2015) The evolution of annelids reveals two adaptive routes to the interstitial realm. *Current Biology* **25**, 1993–1999.
- Struck TH, Schult N, Kusen T, Hickman E, Bleidorn C, McHugh D and Halanych KM (2007) Annelid phylogeny and the status of Sipuncula and Echiura. *BMC Evolutionary Biology* **7**, 57.
- Sun Y, Daffe G, Zhang Y, Pons J, Qiu J-W and 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.
- Taghon GL (1988) The benefits and costs of deposit feeding in the polychaete *Abarenicola pacifica*. *Limnology and Oceanography* **33**, 1166–1175.
- Tan MH, Gan HM, Lee YP, Bracken-Grissom H, Chan TY, Miller AD and Austin CM (2019) Comparative mitogenomics of the Decapoda reveals evolutionary heterogeneity in architecture and composition. *Scientific Reports* **9**, 1–16.
- Tessler M, de Carle D, Voiklis ML, Gresham OA, Neumann JS, Cios S and 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.

- Wang Y, Huang M, Wang R and Fu L** (2018) Complete mitochondrial genome of the fish leech *Zeylanicobdella arugamensis*. *Mitochondrial DNA Part B: Resources* **3**, 659–660.
- Wu Z, Shen X, Sun M, Ren J, Wang Y, Huang Y and Liu B** (2009) Phylogenetic analyses of complete mitochondrial genome of *Urechis unicinctus* (Echiura) support that echiurans are derived annelids. *Molecular Phylogenetics and Evolution* **52**, 558–562.
- Ye F, Liu T, Zhu W and You P** (2015) Complete mitochondrial genome of *Whitmania laevis* (Annelida, Hirudinea) and comparative analyses within *Whitmania* mitochondrial genomes. *Belgian Journal of Zoology* **145**, 114–128.
- Zhang Y, Sun J, Rouse GW, Wiklund H, Pleijel F and Watanabe HK** (2018) Phylogeny, evolution and mitochondrial gene order rearrangement in scale worms (Aphroditiformia, Annelida). *Molecular Phylogenetics and Evolution* **125**, 220–231.
- Zhong M, Hansen B, Nesnidal M, Golombek A, Halanych KM and Struck TH** (2011) Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. *BMC Evolutionary Biology* **11**, 369.