Zoosyst. Evol. 101 (4) 2025, 1815–1830 | DOI 10.3897/zse.101.161344

PENSOFT

MUSEUM FÜR NATURKUNDE BERLIN

New molecular resources of *Loimia borealis* (Annelida, Terebellidae) based on genome-skimming and RNA-seq approaches, with insights into monophyly of the genus

Sheng Zeng[1], Deyuan Yang[1], Caifang He[2], Yanan Sun[3], Christopher J. Glasby[4,5], Yanjie Zhang[2]

1 *College of the Environment and Ecology, Xiamen University, Xiamen 361102, China*
2 *School of Life and Health Sciences, Hainan University, Haikou 570228, China*
3 *Laboratory of Marine Organism Taxonomy and Phylogeny, Qingdao Key Laboratory of Marine Biodiversity and Conservation, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 26000, China*
4 *Natural Sciences, Museum & Art Gallery Northern Territory, PO Box 4646, Darwin, NT 0801, Australia*
5 *Australian Museum Research Institute, 1 William Street, NSW 2010, Australia*

https://zoobank.org/4BB1250A-DB15-4F8E-BA48-D16456C3041D

Corresponding authors: Deyuan Yang (deyuanyang92@163.com); Yanjie Zhang (yanjiezhang@hainanu.edu.cn); Christopher J. Glasby (glasby93@gmail.com); Sheng Zeng (33120231152326@stu.xmu.edu.cn)

Academic editor: Greg Rouse ♦ **Received** 6 June 2025 ♦ **Accepted** 12 August 2025 ♦ **Published** 19 September 2025

Abstract

The terebellid *Loimia borealis* Wang, Sui, Kou & Li, 2020 was first described from Shouguang, China, and its molecular data, consisting of partial mitochondrial genes (cox1 and 16S) and a partial nuclear gene (18S), were reported. In this study, small-sized specimens of *L. borealis* were collected near the type location, and the species' morphological description is supplemented. Several commonly used genetic sequences, including the complete mitochondrial genome, complete nuclear high-copy genes (18S-ITS1-5.8S-ITS2-28S), and histone genes (H3, H4, H2B, and H2A), are also provided. Additionally, the transcriptome of *L. borealis* is assembled from data generated by RNA-seq. A Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment showed high completeness of the transcriptome, with 97.5% (655/672) of the metazoan conserved orthologous genes. We provide an easier pipeline to facilitate the mining of molecular markers from genome-skimming sequencing data. The monophyly of *Loimia* is supported when combined with partial mitochondrial gene markers (cox1 and 16S) and partial nuclear genes (18S, 28S, and H3), but not by individual genes. The nodes in the phylogenetic trees based on 13 protein-coding genes (13 PCGs) and 2 mitochondrial rRNA genes were well supported, with most showing maximum bootstrap support and posterior probability. The current classification of Terebelliformia and Cirratuliformia is also supported.

Key Words

Genome skimming, histone genes, *Loimia*, mitogenome, nuclear genes, pipeline, polychaete, Terebellida

Introduction

General introduction of *Loimia*

The genus *Loimia* Malmgren, 1866, belongs to the species-rich, globally distributed family Terebellidae Johnston, 1846 (de Matos Nogueira et al. 2015; Stiller et al. 2020; Read and Fauchald 2025). Most authors agree that this group is characterized among terebellids by the presence of enlarged lateral lobes on the anterior segments and adults with monopectinate uncini (all teeth arranged in a single row) (Malmgren 1866; Caullery 1944; Fauchald 1977; de Matos Nogueira et al. 2015; Wang et al. 2020; Martin et al. 2022; Hutchings et al. 2024; Hutchings et al. 2025). *Loimia* species are mostly found in tropical and subtropical areas, with no known Arctic or Antarctic

Copyright *Zeng, S. et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

records, and are predominantly associated with intertidal or coastal habitats (Lavesque et al. 2017; Hutchings et al. 2021; Hutchings et al. 2024).

Currently, 34 species have been accepted according to the World Register of Marine Species (WoRMS) (Read and Fauchald 2025). Of these recognized taxa, 14 species have been described in the past decade (de Matos Nogueira et al. 2010; de Matos Nogueira et al. 2015; Lavesque et al. 2017; Wang et al. 2020; Martin et al. 2022; Hutchings et al. 2024; Hutchings et al. 2025). However, only six species have been characterized using both morphological and molecular approaches (including partial mitochondrial genes 16S or cox1, or 16S + cox1): *Loimia borealis* Wang, Sui, Kou & Li, 2020, *Loimia macrobranchia* Wang, Sui, Kou & Li, 2020; *Loimia davidi* Martin, Capa, Martínez & Costa, 2022; *Loimia lanai* Hutchings, Daffe, Flaxman, Rouse & Lavesque, 2024; *Loimia aimehoensis* Hutchings, Daffe, Glasby & Lavesque, 2025; and *Loimia poraporaensis* Hutchings, Daffe, Glasby & Lavesque, 2025.

Difficulty identifying *Loimia* and its taxonomic status

Loimia species present significant identification challenges due to their highly similar morphological features, particularly in preserved specimens. Hutchings et al. (2024, 2025) summarized the key diagnostic characters for this genus, including (1) the shape and position of the lateral lobes, (2) the number of teeth in the uncini, and (3) the presence, number, and shape of the ventral pads. More recently, Martin et al. (2022) and Hutchings et al. (2025) demonstrated that some of these taxonomic characters change as the animal grows. Due to limited molecular data, most of these characters have not yet been evaluated across the genus using an integrated approach that combines molecular and morphological data. Quantitative morphometric studies, which have been found to be useful for delineating species and generic groups in other Terebellidae (e.g., Glasby and Hsieh 2006; Garraffoni and de Garcia Camargo 2007), are also limited.

Several phylogenetic analyses based on morphological (Glasby et al. 2004; Garraffoni and Lana 2008; de Matos Nogueira et al. 2013) or molecular (Stiller et al. 2020) data have primarily focused on high-level (family or subfamily) classification systems. Detailed discussions at the genus level remain limited, likely due to restricted taxon sampling within each genus.

Previous studies have shown that *Loimia* is closely related to *Lanice* (Garraffoni and Lana 2008; de Matos Nogueira et al. 2013; Stiller et al. 2020). Stiller et al. (2020) also placed *Loimia* within the tribe Lanicini Holthe, 1986, which includes other genera with large lateral lappets (e.g., *Axionice* Malmgren, 1866, *Lanice* Malmgren, 1866, *Lanicola* Hartmann-Schröder, 1986, *Pista* Malmgren, 1866, and *Scionella* Moore, 1903). Jirkov and Leontovich (2017) performed a cladistic analysis on genera with large lateral lobes within Terebellidae, including *Axionice* Malmgren, 1866; *Betapista* Banse, 1980; *Eupistella* Chamberlin, 1919; *Pista* Malmgren, 1866; *Loimia* Malmgren, 1866; *Lanice* Malmgren, 1866; *Lanicides* Hessle, 1917; *Lanicola* Hartmann-Schröder, 1986; *Paraxionice* Fauchald, 1972; and *Scionella* Moore, 1903. They proposed that *Loimia* and other genera (*Eupistella*, *Euscione*, *Lanice*, and *Paraxionice*) should be considered synonyms of *Axionice*. However, this has not been accepted by subsequent publications (e.g., Hutchings et al. 2021; Lavesque et al. 2021; Hutchings et al. 2024; Hutchings et al. 2025). Therefore, the taxonomic status of the genus *Loimia* remains unresolved.

Molecular resources of *Loimia*

With the development of sequencing technologies, molecular phylogeny has evolved into phylogenomics, which is based on hundreds of genes instead of a single or a few genes. Yet molecular resources for *Loimia* remain limited. They include a small number of commonly used molecular markers (cox1: 49 sequences, 16S: 37, 28S: 5, 18S: 20) for a few species (Lavesque et al. 2017; Stiller et al. 2020; Wang et al. 2020; Martin et al. 2022; Hutchings et al. 2024; Hutchings et al. 2025), one transcriptome dataset for phylogenetic analysis (Stiller et al. 2020), and one mitochondrial genome (mitogenome) (Xu et al. 2024). Although mitogenes are widely used in *Loimia* studies, the primers used to amplify cox1 and 16S appear to have low efficiency (Martin et al. 2022; Hutchings et al. 2024; Hutchings et al. 2025), suggesting the need to design new primers for the genus.

There are seven *Loimia* species in China (Wang et al. 2020). *Loimia borealis* Wang, Sui, Kou & Li, 2020 was first described from Yangkou Village, Shouguang City, Shandong Province, China (type locality), with additional material sourced from Cangkou, Qingdao. Three single-gene markers were provided: cox1, 16S, and 18S (Wang et al. 2020). On 19–20 September 2024, small-sized specimens of *Loimia borealis* Wang, Sui, Kou & Li, 2020, were sampled from a stony beach in the coastal area of Guzhenkou, Qingdao, Shandong Province, China, about 220 linear km southeast of the type locality.

In this study, we focused on three major aims: (1) supplement the molecular resources for *Loimia borealis*, including complete mitogenomes and multiple-copy nuclear genes using genome-skimming sequencing (also called low-coverage whole genome sequencing) and other nuclear genes using RNA-seq approaches; (2) supplement the morphological description of the small-sized specimens with a registered voucher collected near the type locality; and (3) provide novel insights into the monophyly of *Loimia* based on phylogenetic analyses mostly using mitogenes.

Materials and methods

Sample collection

Specimens of *Loimia borealis* Wang, Sui, Kou & Li, 2020, were collected from a stony beach at the coastal area of Guzhenkou, Institute of Oceanology, Chinese Academy of Sciences (IOCAS), Qingdao, Shandong Province, China (36.05519°N, 120.34604°E) (Table 1).

Morphological analyses

Photos of live specimens were taken with a ZEISS STEREO Discovery V20. Whole preserved specimens were observed with an OLYMPUS SZX16 stereomicroscope and photographed using a Canon EOS 5D Mark IV camera. The detailed morphology of the notochaetae and uncini was observed using an OLYMPUS CX33 with a Canon EOS 90D. For SEM, the tissue with notochaetae and uncini was dehydrated in ethanol, then replaced with hexamethyl disilazane (HMDS) solutions. After drying and sample attachment in a fume cupboard, a JSM-7100F electron microscope was used to scan the details of the notochaetae and uncini. Genera were identified following definitions in Hutchings et al. (2017). Species identification was validated by morphological comparison with the type specimen in Wang et al. (2020) and sequence alignment of cox1 and 16S.

DNA and RNA extraction and sequencing

Tissues or whole specimens preserved in 95% ethanol and RNAlater (Coolaber, Beijing, China) were used for DNA and RNA extraction, respectively. Other material was preserved in 75% ethanol for morphological examination. Voucher specimens were deposited at IOCAS (Table 1). After morphological examination, the tail or the whole specimen was used for DNA/RNA extraction. Tissues or specimens for DNA and RNA extraction were sent to Novogene Co., Ltd. in Tianjin, China, using their custom extraction methods (see Suppl. material 1). Genome skimming (GS) and RNA sequencing were conducted on the Illumina NovaSeq X Plus with a paired-end 150 bp strategy. The sequencing generated 4.62 gigabases (Gb) of raw data for GS, 14.78 Gb for the second GS, and 14.87 Gb for RNA-seq. Adapters and low-quality regions were removed using fastp v.0.23.4 (Chen et al. 2018) with default parameters. The quality of the clean data was evaluated with FastQC v.0.11.5 (Andrews 2010). The output from FastQC was subsequently aggregated and reported through MultiQC v.1.8 (Ewels et al. 2016) (Suppl. material 2: fig. S1).

Mining molecular data

Mitochondrial genomes and nuclear multiple-copy genes

GetOrganelle v.1.7.6.1 (Jin et al. 2020) was used for assembly of the cleaned data (GS), utilizing all mitogenome sequences of Terebellidae as seeds (NCBI database, 15 December 2024), with k-mer values of 17, 21, 33, 39, 45, 55, 65, 75, 85, 95, 105, 115, and 127, while other parameters were kept at default. The first assembly yielded a 15,893 bp sequence (reported as linear, average base coverage = 300.5). A second assembly was then performed with GetOrganelle, using the sequence obtained from the first assembly as the seed and keeping the remaining parameters unchanged. The circular mitogenome was obtained (15,884 bp in length, average base coverage = 322.2). The difference between the two sequences was in the control region. The circular one was used for subsequent analyses and uploaded to NCBI with the accession number PQ774336.

Annotation of the mitogenome was performed using several software programs following the methods described by Yang et al. (2024): (i) MitoZ v.3.6 (Meng et al. 2019); (ii) Mitofinder v.1.4.1 (Allio et al. 2020); (iii) MITOS2 (Donath et al. 2019); (iv) tRNA annotation with GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) incorporating ARAGORN v.1.2.38 (Laslett and Canback 2004), ARWEN v.1.2.3 (Laslett and Canbäck 2008), and tRNAscan-SE v.2.0.7 (Chan et al. 2021). All annotated files were imported into Geneious v.9.0.2 (Kearse et al. 2012), where coding regions were manually

Table 1. Specimen information of *Loimia borealis*. Abbreviations: L = length; W = width.

Museum catalog No.	Voucher No.	Size (L: mm)	Number of segments	Ventral Shields (numbers)	Thorax (L/W, mm)	Remarks
MBM287941	QD_Tere_sp3_001_240920	18	76	8	7/3.5	
MBM287942	QD_Tere_sp3_002_240920	17	54	8	9/3	
MBM287943	QD_Tere_sp3_003_240920	22	89	8	12/5	
MBM287944	QD_Tere_sp3_004_240920	27	62	8	12/5	Tail cut for DNA extraction
MBM287945	QD_Tere_sp3_005_240920	13	55	7	7/4	
	QD_Tere_sp3_008_240920	20	76	9	9/2	
	QD_Terebellidae_sp3	20	76	9	9/2	Whole specimen used for DNA and RNA extraction

checked for open reading frames (ORFs). GC content and per-base read coverage depth of the mitogenome were calculated using the visualize subcommand in MitoZ v.3.6 with default parameters.

The sequence containing nuclear ribosomal genes (18S rRNA to 28S rRNA) was assembled from the cleaned data using GetOrganelle v.1.7.6.1 (Jin et al. 2020), seeding with all 18S and 28S sequences of Terebellidae downloaded from GenBank on 15 December 2024. The resulting sequence was 9,449 bp long (PV012639), with an average base coverage of 5529.6. It was manually annotated in Geneious v.9.0.2 using primer pairs 18SA and 18SB (Medlin et al. 1988) for 18S, F63.2 and R3264.2 (Struck et al. 2006) for 28S, and ITS18SFPOLY and ITS28SRPOLY (Nygren et al. 2009) for the internal transcribed spacer (ITS) 1–5.8S-ITS2 region. The ITS1 and ITS2 were identified based on the boundaries of the 18S, 5.8S, and 28S genes.

The nuclear histone genes (H3-H2A-H2B-H4) were assembled using GetOrganelle v.1.7.6.1 with three complete histone sequences from Polychaeta in the NCBI database as seeds (*Urechis caupo* X58895, *Platynereis dumerilii* X53330, and *Chaetopterus variopedatus* AH006103). The assembled result was 4,219 bp (PV012640), with an average base coverage of 569.7. Annotation was performed using the "live annotate and predict" feature in Geneious v.9.0.2, with the three sequences as reference (80% similarity threshold).

The assembly and annotation workflow for each gene is shown in Fig. 1. Once annotations were completed, each gene or sequence was extracted for NCBI BLAST analysis. The results were downloaded as CSV files. For each sequence, the top 20 hits were sorted based on identity (see Suppl. material 3: table S1) using our custom script.

BUSCO genes

In addition to the above molecular data, near-universal single-copy BUSCO genes have been widely used across diverse taxonomic groups, such as yeast (Shen et al. 2018), insects (Du et al. 2024), invertebrates (Collins et al. 2023), Opisthokonta (Liu et al. 2024), and Mollusca (Chen et al. 2025). To evaluate the sequencing depth required to recover sufficient numbers of single-copy genes from genome assemblies using BUSCO analysis, we tested datasets with approximately 5 Gb and 20 Gb of genome size (GS), respectively. De novo genome assemblies were performed using SPAdes v.4.0.0 (Prjibelski et al. 2020) with default parameters. Genome completeness was subsequently assessed using BUSCO v.5.8.2 (Simão et al. 2015) with default settings and the metazoan_odb12 dataset.

RNA assembly

RNA assembly followed the methodology established by Martínez-Redondo et al. (2024). Detailed protocols are available in the MATEdb2 GitHub repository (https://github.com/MetazoaPhylogenomicsLab/MATEdb2). Briefly, the cleaned reads were assembled using Trinity v.2.8.5 with default parameters. Transcript quality was assessed using BUSCO v.5.8.2 with default parameters and the metazoan_odb12 dataset. Protein-coding regions (ORFs) of transcripts were identified using TransDecoder v.5.7.1 (https://github.com/TransDecoder/TransDecoder) according to the standard protocol (https://github.com/TransDecoder/TransDecoder/wiki). Subsequently, the script fetch_longest_iso.py (https://github.com/MetazoaPhylogenomicsLab/MATEdb2) was used to extract

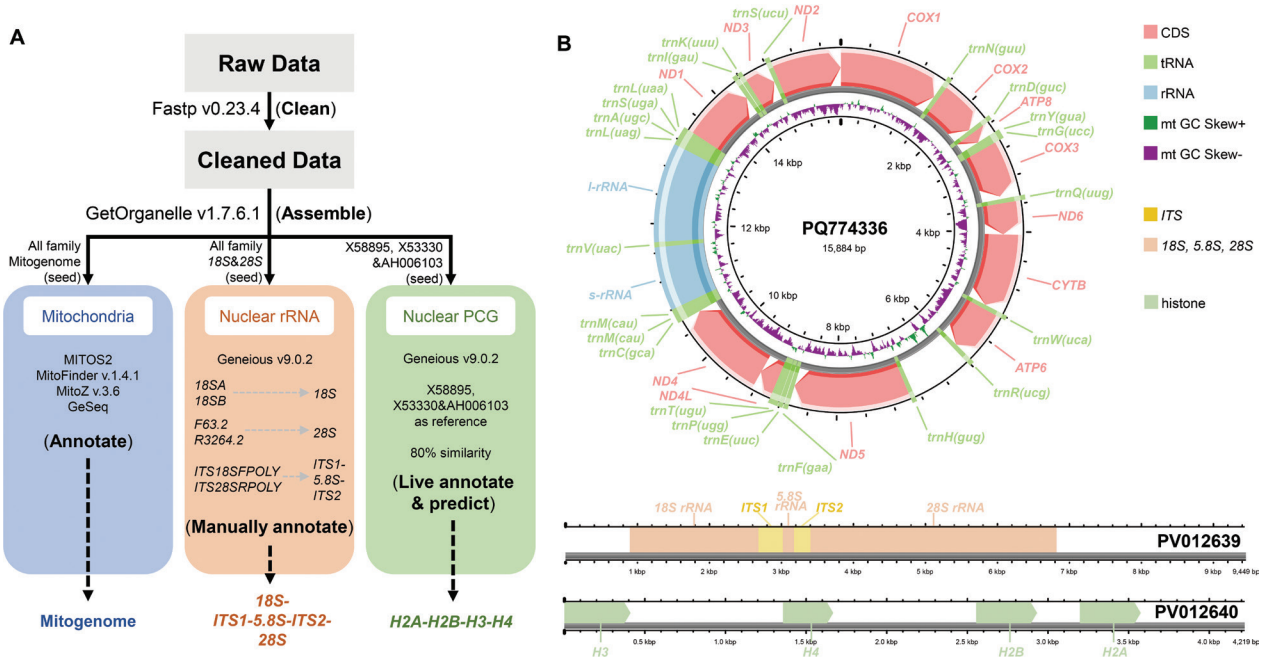

Figure 1. A. Genome skimming workflow for *de novo* assembly of the mitogenome and high-copy-number nuclear genes; **B.** Structural schematics of the assembled mitogenome, the 18S-ITS1-5.8S-ITS2-28S rRNA genes, and the H3-H2A-H2B-H4 histone gene cluster.

the longest isoform for each gene. A second BUSCO assessment was conducted to evaluate the quality of these longest isoforms.

Genome size

Genome size was estimated following Ranallo-Benavidez et al. (2020). A k-mer count histogram was generated using Jellyfish v.2.3.1 (Marçais and Kingsford 2011), then analyzed using GenomeScope 2.0 (Ranallo-Benavidez et al. 2020). More details are available at https://github.com/tbenavi1/genomescope2.0.

Phylogenetic analyses

We re-annotated all Terebellida sequences (31 sequences from the NCBI database, 25 December 2024; Suppl. material 3: table S2). Three sequences of the ampharetid *Decemunciger* (KY742027, KY774370, KY774371) were removed due to multiple stop codons or introns in the cox1, nad1, nad4, and nad5 genes. To identify potential taxonomically mislabeled sequences in the mitogenome dataset, an NCBI BLAST analysis was conducted (Suppl. material 3: table S3). The outgroup was selected according to Struck (2019). The final dataset, summarized in Suppl. material 3: table S2, consisted of 31 sequences from Terebellida and two outgroup taxa, *Owenia fusiformis* NC028712 (Sabellida: Oweniidae) and *Prionospio cirrifera* OR935936 (Spionida: Spionidae). Different datasets were used for phylogeny: (i) 13PCGs123, including all three codon positions of the 13 PCGs; (ii) 13PCGs123_2rRNAs, combining the 13PCGs dataset with two rRNA genes; (iii) 13PCGs12, excluding the third codon positions of the 13 PCGs; (iv) 13PCGs12_2rRNAs, combining the 13PCGs12 dataset with two rRNA genes; and (v) 13PCGsAA, comprising amino acid sequences translated from the 13 PCGs.

Considering the limited availability of mitogenomes for Terebellidae, DNA barcodes were also included to explore the phylogenetic position of *L. borealis*. All available terebellid genes were downloaded from NCBI (25 December 2024). The sequences of widely used molecular markers cox1, 16S, 18S, 28S, and H3 were extracted. A custom script by Yang et al. (2024) was used to organize the data (dataset in Suppl. material 3: table S4). The outgroup for this dataset was selected according to Stiller et al. (2020).

For the mitogenome, the PCGs and two rRNAs were aligned using MAFFT in normal mode. Multiple sequence alignments for the 13 PCGs were further refined using MACSE v.2 (Ranwez et al. 2018) under the "refinement strategy." TrimAl v.1.2 (Capella-Gutiérrez et al. 2009) was used to remove ambiguously aligned regions with the "automated1" setting (Suppl. material 3: table S5). Substitution models were selected with ModelFinder v.2.2.0 (Kalyaanamoorthy et al. 2017) based on the Bayesian Information Criterion (BIC) for maximum likelihood (ML) analysis and the corrected Akaike Information Criterion (AICc) for Bayesian inference (BI) analysis under a partitioned model (Suppl. material 3: table S6). ML phylogenetic reconstruction was conducted using IQ-TREE v.2.2.2 (Nguyen et al. 2015; Kalyaanamoorthy et al. 2017) with the best partition scheme and an edge-linked partition model. Node support was assessed with 200,000 ultrafast bootstraps. BI phylogenies were constructed in MrBayes v.3.2.7a (Ronquist et al. 2012), with two parallel runs and 2,000,000 generations. Convergence was evaluated by the average standard deviation of split frequencies (ASDSF), and additional generations were added if the ASDSF value exceeded 0.01. The phylogenies constructed with cox1, 16S, 18S, 28S, and H3 followed the same methods as above, and sequences are listed in Suppl. material 3: table S4.

Topological differences among the inferred phylogenies were evaluated using TreeSpace (Jombart et al. 2017) implemented in R v.4.3.1 (R Core Team 2018). Finally, ML and BI trees were visualized using iTOL v.6 (Letunic and Bork 2024).

Sequence analyses

BioKIT (Steenwyk et al. 2022) was used to compute multiple-alignment summary statistics (number of sequences, alignment lengths, number of constant sites, number of parsimony-informative sites and variable sites, and the frequency of all character states) (Suppl. material 3: table S7).

Strand asymmetries were calculated using the formulas proposed by Perna and Kocher (1995): AT-skew = (A − T)/(A + T) and GC-skew = (G − C)/(G + C). Codon usage and relative synonymous codon usage (RSCU) of the 13 PCGs were analyzed using PhyloSuite and visualized with the ggplot2 package (Wickham 2016) in R v.4.1.3 (R Core Team 2018). DnaSP v.6.0 (Rozas et al. 2017) was used to estimate nucleotide diversity (Pi) in a sliding-window analysis (window 100 bp; step 25 bp) and substitution rates (Ka/Ks), based on the mitogenes of Terebellidae listed in Suppl. material 3: table S3.

Genetic distances (p-distance and Kimura 2-parameter) between sequence pairs were calculated in MEGA X (Kumar et al. 2018). The sequence information used for calculating genetic distances is provided in Suppl. material 3: table S8. Heat maps were generated using TBtools v.2.080 (Chen et al. 2020).

Results

Morphological analyses

Description of specimen MBM287943

Tube about 5–10 cm long, fragile, made of shell fragments and small gravel; inner surface with thin, smooth membrane. Complete specimen with 62 segments, measuring 27 mm long; thorax 12 mm long. Tentacles abundant and long, reaching the end of thorax; white and lacking band-

ing (Fig. 2A–C). Body salmon or reddish in live specimens, pale when preserved (Fig. 2A–F). Anterior segments compact, inflated dorsally (Fig. 2C, E). Segments 1 and 3 each with one pair of lateral lappets; the lateral lappets of segment 3 are similar in size to those of segment 1 but slightly narrower and higher (Fig. 2C, J, K). Transverse prostomium attached to dorsal surface of upper lip. Eyespots absent. Peristomium restricted to lips, thicker at the base. Upper lip conical, wider than long, protruding anteriorly; lower lip short, button-like (Fig. 2D). Ventral shields smooth, progressively narrowing posteriorly, rectangular to trapezoidal; anterior ones up to about 5 times wider than long; present on segments 2–12, with those on segments 2–4 almost fused; ventral shields yellowish on segments 2–9 and red on segments 10–12 in live material, gradually tapering; two shields on segment 12; white in preserved specimens (Fig. 2A, C, D). Three pairs of branchiae on segments 2–4, equal in size, with short, thick stalks and dichotomous branches (Fig. 2F). Notopodia from segment 4 through segment 20; chaetae narrowly bilimbate (Suppl. material 2: fig. S2J, K). Neuropodia begin on segment 5 as low rectangular ridges slightly raised from the body surface until segment 20, where notopodia terminate, thereafter as elongate pinnules (Fig. 2I, J). Uncini pectinate, arranged in a single row until segment 10, then in double rows from segments 11–20 in back-to-back position, with 6–7 teeth, higher than long (Suppl. material 2: fig. S2A–G). Posterior segments much shorter than thoracic segments. Uncini pectinate, with a crest of 5–6 teeth in a single vertical series until the pygidium, smaller than thoracic ones (Suppl. material 2: fig. S2H, I). Notochaetae all simple capillaries, with pronounced, variable texture (Suppl. material 2: fig. S3K, O). Uncini arranged in slightly curved rows; 6–7 large teeth in a single vertical series, with numerous very small basal teeth (Suppl. material 2: fig. S3E).

Variation in other specimens

Body length ranges from 13 to 27 mm. The number of segments in complete specimens varies from 54 to 89. The number of ventral shields varies from 7 to 9, fused anteriorly. Nephridial or genital papillae on segments 6–8 can be observed in live specimens and are not easily observed after preservation.

Comments

Our specimens are smaller than those of Wang et al. (2020). In Wang et al. (2020), body length is 59.6–82.8 mm, whereas our specimens are 13–27 mm. The number of teeth in the uncini also varies. In Wang et al. (2020), uncini on segment 5 have 4 teeth, while our samples have 6 teeth. Uncini on segment 20 have 6 teeth in Wang et al. (2020), while our samples have 6–7 teeth. Nephridial or genital papillae in Wang et al. (2020) were small (c. 0.1 mm diameter as determined from their Fig. 2A, B) and present in the holotype at the posterodorsal base of notopodia of segments 6–8 but absent in other specimens; in our specimens, they were similar in size and observed in the same region in live specimens but were not large enough to be visible in Fig. 2A–C. They were not obvious after alcohol preservation but could be observed after methyl green staining. Neither Wang et al. (2020) nor our study found nephridial papillae on segments 3 and 4, but they may occur in other species of *Loimia*. We estimated papilla size in the above emendation; however, these structures are not traditionally measured because they are very small relative to the notopodium, and their shape is not particularly taxonomically useful – most studies do not image them specifically, though they may appear in close-up images of the anterior parapodia if the angle is suitable. Anterior shields are rectangular with a length about five times the width, as in our specimens.

In summary, Wang et al. (2020) found that *Loimia borealis* can be distinguished from other *Loimia* species by the absence of eyespots, large ventral lateral lobes on segment 1, well-developed lobes on segment 3 that are slightly larger than those on segment 1, 7–9 equal-sized rectangular mid-ventral shields (each with length 5 × width) from segment 2, and three pairs of equal-sized branchiae. To this, we add that the type material of *L. borealis* lacks tentacular banding, which occurs in many other *Loimia* species but not all.

Martin et al. (2022) identified size-related variation in most of the taxonomic characters they examined for *Loimia davidi* Martin, Capa, Martínez & Costa, 2022, including smaller specimens having one or two fewer teeth on the uncini. Interestingly, the smaller specimens of *L. borealis* that we examined had one or two more teeth than Wang et al.'s larger specimens. Nevertheless, there are too few data on variation in uncinal teeth relative to body size to emphasize this difference. Thus, in terms of overall similarity and diagnostic characters, we conclude that our specimens are morphologically representative of *Loimia borealis*.

Characteristics of the molecular resources

The complete mitogenome of *Loimia borealis* is 15,884 bp in length. It contains 12 intergenic intervals (totaling 81 bp) and one overlapping region (7 bp, between nad4 and nad4l), which is common in other terebellids. All 37 genes are located on the heavy (H) strand (Fig. 3A). All terebellid mitogenomes exhibit high A + T content (65.6%–70.4%) (Suppl. material 3: table S3). The A + T content of *L. borealis* is 66.6%, with A, T, G, and C constituting 28.3%, 38.3%, 13.2%, and 20.1%, respectively. The AT-skew and GC-skew of the mitogenome are negative (−0.149 and −0.206, respectively). The 13 PCGs of *L. borealis* total 11,148 bp and encode 3,716 amino acids, beginning with ATG except for nad4l (ATT). Regarding stop codons, cox1, cytb, and nad5 terminate with a single T; atp6, nad4, and nad1 end with TA; cox3 and nad2 end with TAG; and the remaining genes end with TAA.

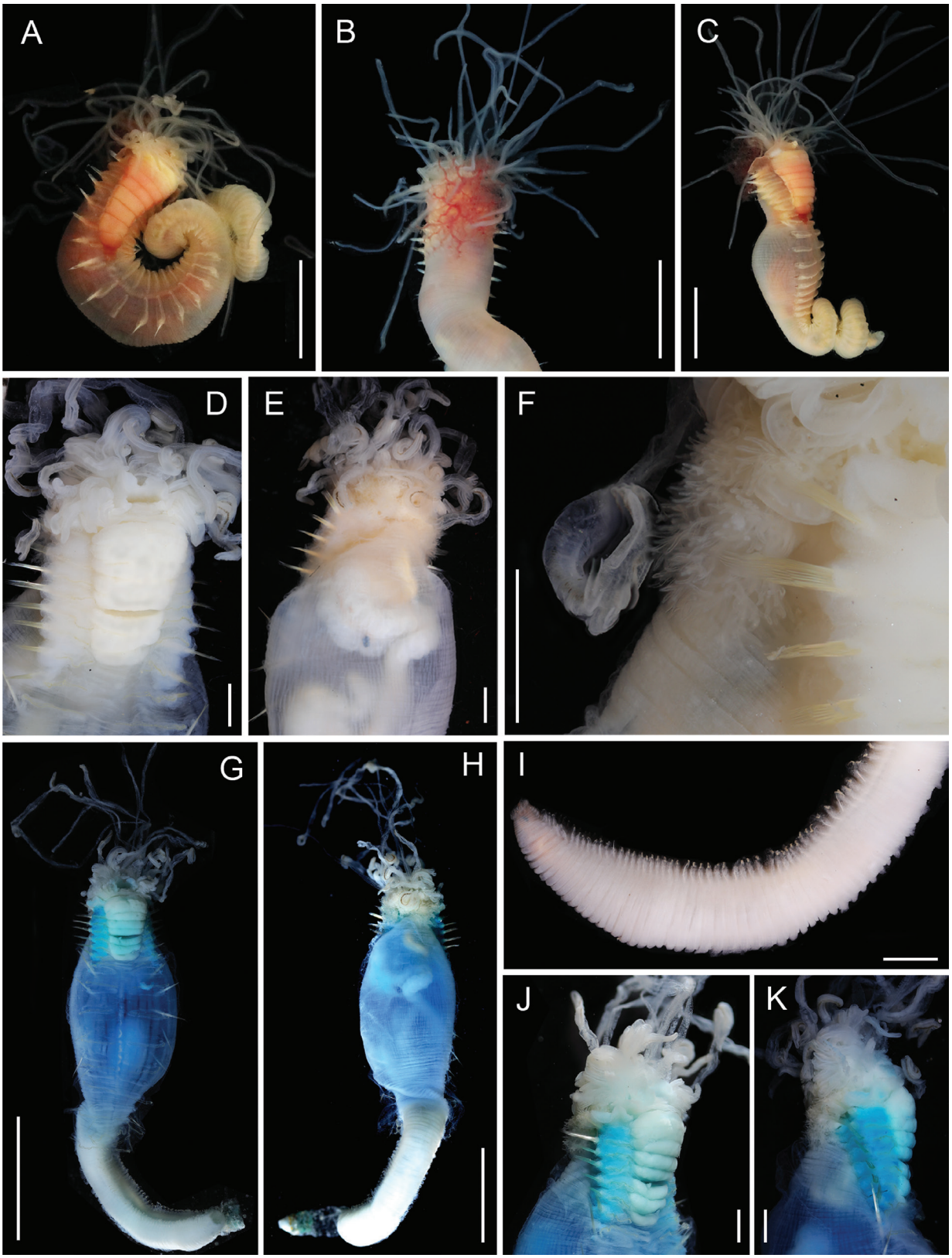

Figure 2. Details of *Loimia borealis*. **A–C.** Live specimens; **D–F, I.** Preserved in Alcohol; **G, H, J, K.** After methyl green staining. **A, G.** Ventral-lateral view; **B, H.** Dorsal view; **C, J, K.** Lateral view; **D–F.** Detail of the anterior part, ventral-lateral view; **F.** Anterior part, detail of the branchiae; **I.** Detail of the abdomen. Scale bars: 5 mm (**A–C, G, H**); 1 mm (**D–F, I–K**).

The AT- and GC-skews of the 13 PCGs are −0.177 and −0.032, respectively. The 12S rRNA (878 bp) is located between trnM and trnV, while the 16S rRNA (1,370 bp) is positioned between trnV and trnL1. A total of 23 tRNAs were identified, ranging from 64 to 71 bp (Suppl. material 3: tables S9, S10).

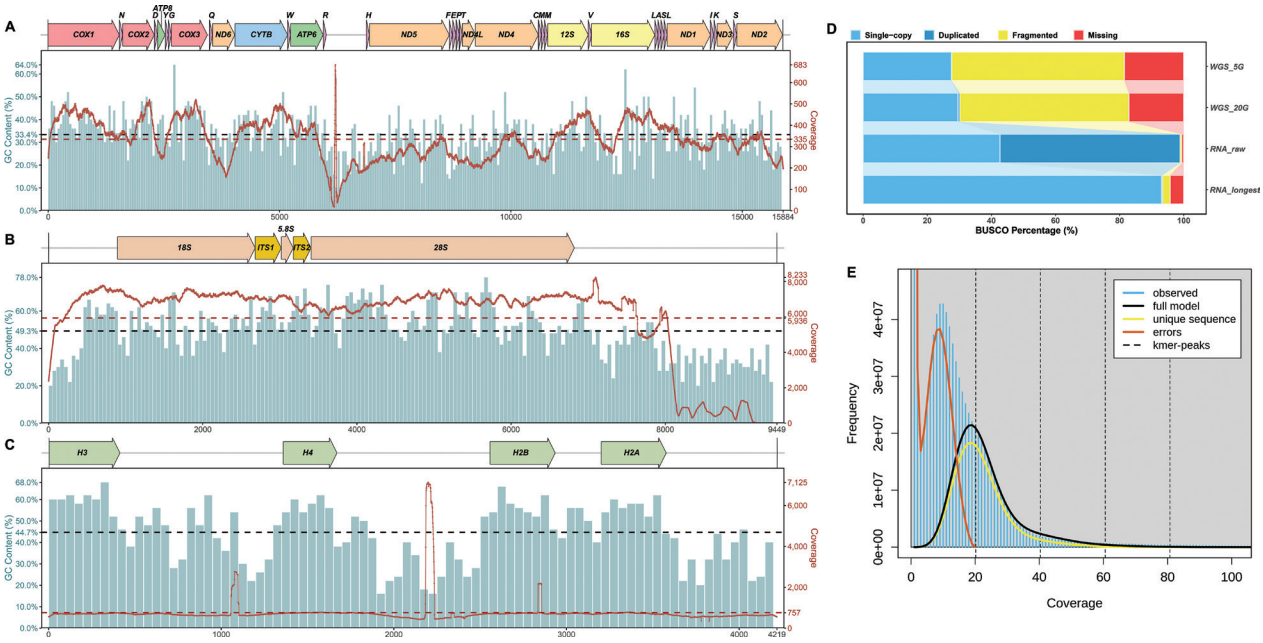

Figure 3. The molecular resources of *Loimia borealis*. The gene map of: **A.** Mitogenome, **B.** Nuclear rRNA cluster, and **C.** Histone genes. The red lines depict the distribution of coverage depth, and the green columns depict the GC content of sequences; **D.** BUSCO results for the *Loimia borealis* transcriptome assembly; **E** GenomeScope analysis of *Loimia borealis* genomic data with a k=21 model. GenomeScope2 analysis estimates a 425.31 Mb genome size with 9.37% heterozygosity (het = 0.0937), 20.2× average coverage (kcov = 20.2), and a sequencing error rate of 0.944% (err = 0.00944).

RSCU values for PCGs in the mitogenome were analyzed. The most frequently used amino acids are Leu (16.47%), Ser (9.66%), Phe (9.57%), and Ile (8.33%). The least common amino acids are Cys (0.81%), Asp (1.67%), Arg (1.75%), and Gln (1.91%). RSCU analysis indicates that the most frequently used codons are UCU (Ser), CGA (Arg), and CCU (Pro), while ACG (Thr), CGG (Arg), and GUG (Val) have the lowest frequencies (Suppl. material 3: table S11; Suppl. material 2: fig. S4).

The nuclear rRNA sequence is 9,449 bp long and contains full 18S, ITS1, 5.8S, ITS2, and 28S, with gene lengths of 1,787 bp, 338 bp, 155 bp, 231 bp, and 3,417 bp, respectively (Fig. 3B). The nuclear histone genes for both specimens are 4,219 bp, incorporating complete H3 (411 bp), H2A (378 bp), H2B (378 bp), and H4 (312 bp) (Fig. 3C).

BUSCO analysis revealed high completeness (C) of the *L. borealis* transcriptome assembly, with 97.5% (655/672) of metazoan conserved orthologous genes recovered. Of these, 59.1% (397/672) were present as duplicated copies (D), while 38.4% (258/672) were single copies (S). Following extraction of the longest isoform for each gene, the S value increased to 93.2%, comparable to the complete genome assembly of *Terebella lapidaria* (C = 98.2%, S = 96.4%; GCA_949152475.1). In contrast, GS data yielded lower C and S values with high fragmentation (F), despite approximately 40× sequencing coverage (about 20 Gb of sequencing data for an estimated genome size of ~0.5 Gb) (Fig. 3D).

GenomeScope2 k-mer analysis estimates a genome size of 425.31 Mb. The heterozygosity rate of 9.37% indicates substantial genetic diversity. The k-mer profile shows a clear bimodal distribution with the error peak (orange) centered near coverage 1–5 and the main coverage peak at approximately 20×, supporting the model's estimated coverage of 20.2×. The well-defined separation between error and genuine genomic-content peaks suggests high-quality sequencing data with a low error rate (0.944%). The model-fit value of 1.07 indicates excellent congruence between the observed k-mer distribution and the fitted statistical model (Fig. 3E).

Phylogenetic analyses

Summary statistics for multiple alignments of various datasets are available in Suppl. material 3: table S7. Heterogeneity analyses of five mitogenome datasets and five-gene datasets are shown in Suppl. material 2: fig. S5.

Gene order of the mitogenome

The mitogene arrangement in Terebellida is illustrated in Fig. 4A and Suppl. material 2: fig. S6. For the 13 PCGs and two rRNAs, Terebelliformia (15 mitogenomes) exhibits a highly conserved gene order: cox1–cox2–atp8–cox3–nad6–cytb–atp6–nad4l–nad4–nad5–12S–16S–nad1–nad3–nad2. Within Cirratuliformia, Sternaspidae (7 mitogenomes), Acrocirridae (1 mitogenome), and Cirratulidae (5 mitogenomes) display uniform gene arrangements within their respective families; however, the three mitogenomes from Flabelligeridae (each representing different genera) show distinct gene orders from one another (Fig. 4A). The mitogene arrangements (including tRNAs) generally correspond to the patterns observed for the 13

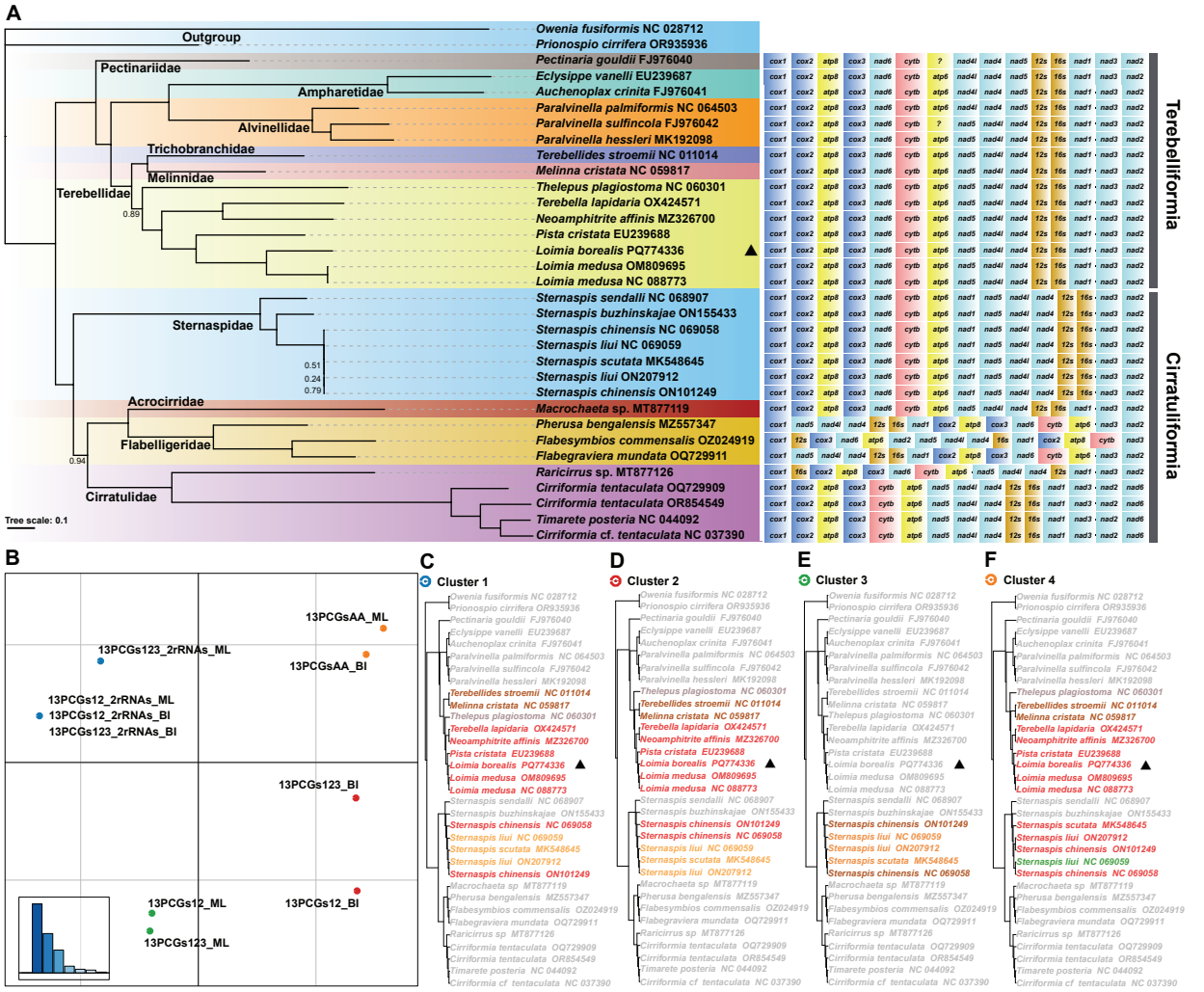

Figure 4. An analysis of phylogenies from five mitogenome datasets. **A.** Bayesian inference (BI) tree of Terebellida based on the dataset 13PCGs123_2rRNAs. The GenBank accession numbers used are listed after the species names. The scale bar (0.1) corresponds to the estimated number of substitutions per site. Numbers at nodes are statistical support values for BI posterior probabilities (PP). "Unlabeled" denotes 100% bootstrap support. Color-coded clades are different families within Terebellida. The gene order is shown to the right (13 PCGs and two rRNAs); **B.** Two-dimensional MDS plot of ten trees, colored by different clusters; **C-F.** A representative tree was selected for each cluster identified in **B**.

PCGs + 2 rRNAs (Suppl. material 2: fig. S6). In addition, atp6 is missing in *Paralvinella sulfincola* FJ976042 and *Pectinaria gouldii* FJ976040.

The position of *Loimia borealis*

In the phylogenetic trees based on mitogenome datasets, the position of *Loimia borealis* is stable, being most closely related to *Loimia medusa* (bs = 100, pp = 1) (Fig. 4). However, the phylogenetic BI tree constructed from five genes shows *L. borealis* as sister to the clade comprising *L. bandera*, *L. arborea*, and *L. medusa* (pp = 0.43) (Fig. 5A); the phylogenetic ML tree places it as sister to the clade comprising *L. bandera*, *L. arborea*, *L. juani*, and *L. medusa* (bs = 14.7) (Fig. 5B).

Single-gene analyses show variable topologies: the cox1 dataset places *L. borealis* as sister to the *L. bandera*/*L. arborea*/*L. medusa* clade with moderate support (bs = 89.8, pp = 0.87) (Suppl. material 2: fig. S7), while 16S rRNA analysis suggests an affinity with *L. tuberculata* and *Lanice* spp. (bs = 84.9, pp = 0.76) (Suppl. material 2: fig. S8). The 18S rRNA phylogeny supports a relationship with *L. bandera* and *L. arborea* (bs = 76.9, pp = 0.84) (Suppl. material 2: fig. S9), whereas both 28S rRNA and histone H3 analyses recover *L. borealis* as sister to *L. bermudensis* (28S: bs = 75.2, pp = 0.58; H3: bs = 82.2, pp = 0.54) (Suppl. material 2: figs S10, S11).

Monophyly of the genus *Loimia*

Phylogenetic trees constructed from five genes support the monophyly of *Loimia* (Fig. 5). A few sequences are most likely mislabeled. For example, *Loimia arborea* (species voucher: BC016; COI: HM473449; 28S: HM473232) did not cluster with other *Loimia* species in phylogenetic trees based on either individual genes

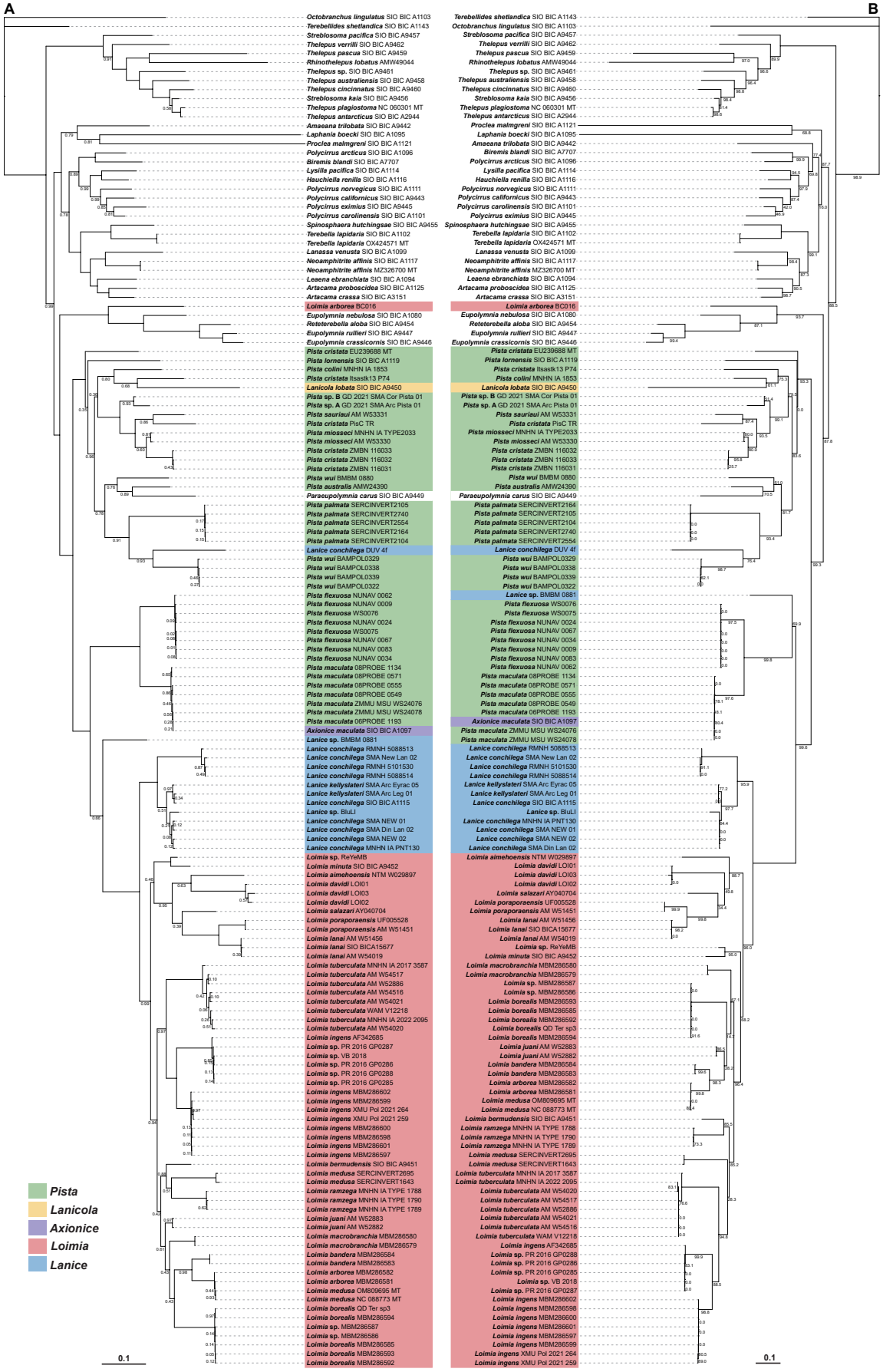

Figure 5. Bayesian inference tree (**A**) and maximum likelihood tree (**B**) inferred from five-gene dataset. Posterior probabilities (pp)/ bootstrap supports (bs) are shown on the nodes.

or the concatenated dataset. Similarly, *Lanice conchilega* (specimen voucher: DUV-4f; COI: OQ053049; 28S: OQ071312, OQ071255) clustered with *Pista* rather than with other *Lanice* specimens.

However, single-gene analyses fail to support *Loimia* monophyly. For instance, trees reconstructed from cox1 and 16S rRNA datasets reveal paraphyly, with sequences from the genus *Lanice* nested within the *Loimia* clade (Suppl. material 2: figs S7, S8).

Nucleotide diversity, evolutionary rate, and p-distance analyses

The nucleotide diversity (Pi) analysis was conducted using concatenated alignments of 13 PCGs and two rRNAs of seven terebellids. The sequence variation ratio shows variable nucleotide diversity in Terebellidae, with Pi values for the 100 bp windows ranging from 0.078 to 0.545. The cox1 (Pi = 0.207), cox2 (0.246), cox3 (0.248), and 12S (0.255) exhibit comparatively low sequence variability, whereas atp8 (0.441), nad6 (0.403), nad2 (0.390), and nad4l (0.385) show high sequence variability (Fig. 6A). A similar trend is observed in Ka/Ks, with cox1 (Ka/Ks = 0.055), cox3 (0.056), cox2 (0.086), and cytb (0.103) evolving comparatively slowly, whereas atp8 (0.531) and nad6 (0.301) are evolving comparatively fast (Fig. 6B).

The analysis of genetic divergence revealed variable levels of differentiation between *Loimia borealis* and its congeners across cox1 and 16S (Fig. 6C, Table 2). The consistent interspecific divergence exceeds 15% across both cox1 and 16S. Mean cox1 p-distances ranged from 17.49% (*L. arborea*, *L. medusa*) to 21.44% (*L. lanai*, *L. minuta*), indicating substantial interspecific variation. The 16S exhibited similar patterns of divergence, with mean p-distances ranging from 15.80% (*L. macrobranchia*) to 21.49% (*L. bandera*) (Table 2).

Discussion

The phylogenetic position of Loimia

In this study, we found no support for the proposal of Jirkov and Leontovich (2017), which was based on morphology, to consider *Betapista*, *Eupistella*, *Lanice*, *Loimia*, and *Paraxionice* as junior synonyms of *Axionice*. These authors defined a broadly circumscribed *Axionice* by the synapomorphy of very large lateral lobes on segment 1

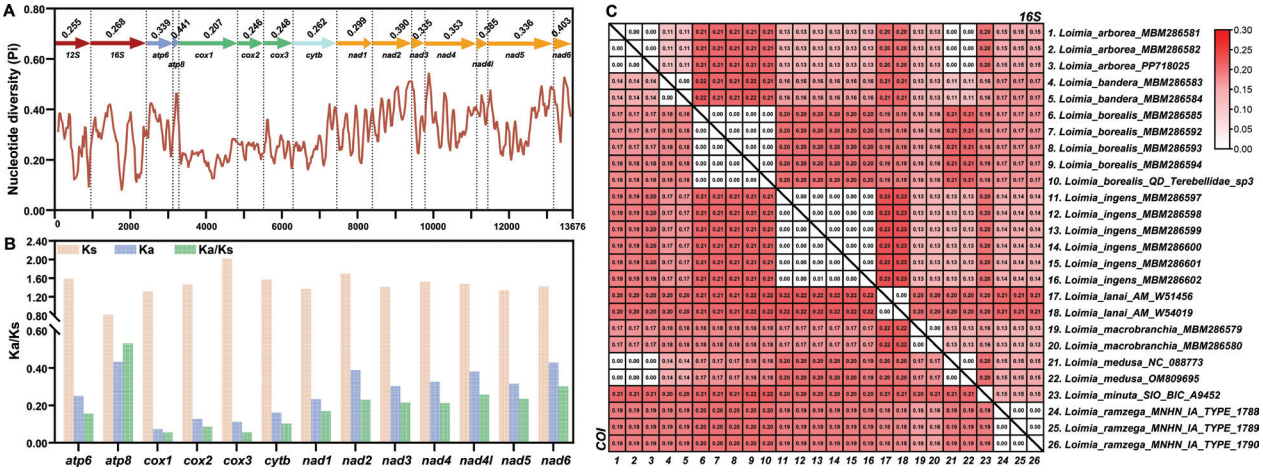

Figure 6. Nucleotide diversity analysis **A.** of two rRNAs + 13 PCGs and Ka/Ks rates **B.** of 13 PCGs based on 7 Terebellidae sequences. The Pi values for the 13 PCGs + two rRNAs are shown on the graph. The red line represents the value of nucleotide diversity (Pi) (window size = 100 bp, step size = 25 bp). The pink, purple, and green columns represent the values of Ks, Ka, and Ka/Ks, respectively; **C.** A heatmap of p-distances between *Loimia* sequences. The p-distances based on *cox1* and *16S* are shown at the bottom left and upper right, respectively.

Table 2. Summary of genetic p-distances between *Loimia borealis* and other *Loimia* species based on mitochondrial cox1 and 16S rRNA.

Species	*cox1*					*16S*				
	Count (n)	Mean	SD	Min	Max	Count (n)	Mean	SD	Min	Max
Loimia arborea	15	0.1749	0.0009	0.1741	0.1759	15	0.2115	0.0017	0.2102	0.2136
Loimia bandera	10	0.1820	0.0009	0.1813	0.1831	10	0.2149	0.0018	0.2136	0.2169
Loimia ingens	30	0.2081	0.0017	0.2065	0.2118	30	0.1992	0.0021	0.1966	0.2034
Loimia lanai	10	0.2144	0.0009	0.2136	0.2154	10	0.1932	0.0000	0.1932	0.1932
Loimia macrobranchia	10	0.1815	0.0013	0.1795	0.1831	10	0.1580	0.0018	0.1559	0.1593
Loimia medusa	10	0.1749	0.0009	0.1741	0.1759	10	0.2115	0.0018	0.2102	0.2136
Loimia minuta	5	0.2144	0.0010	0.2136	0.2154	5	0.1844	0.0019	0.1831	0.1864
Loimia ramzega	15	0.1964	0.0009	0.1957	0.1975	15	0.1708	0.0017	0.1695	0.1729

(using relative size measurements). Our phylogenetic trees inferred from partial mitogenes (mainly cox1 and 16S) show that *Loimia* and *Lanice* form a distinct clade and support the monophyly of *Loimia*. Further, the only species of *Axionice* included in our analysis, *A. maculata* (Dalyell, 1853), grouped with a species of *Pista* (*P. flexuosa*), not with *Lanice* and *Loimia*. The finding of a sister-group relationship between *Loimia* and *Lanice* is consistent with a morphological study by Garraffoni and Lana (2008), which revealed that the two genera share a single apomorphic trait: the presence of a planktotrophic aulophore larva. However, given the limited morphological support and that the present study is based only on mitogenes, further evidence for this sister-group relationship is desirable. Additional research incorporating more taxa based on mitogenomes and/or nuclear genes, together with comparative morphological analyses of *Loimia* species, is necessary to clarify the phylogenetic position of *Loimia*. Moreover, morphological variation (e.g., uncini tooth count, genital papillae) was observed between juveniles and adults, indicating that intraspecific variation may be due to developmental stage and is likely common in *Loimia*.

The power of gene order in mitogenomes

The mitochondrial gene order can provide valuable phylogenetic information (Boore 1999). However, the utility of gene order (13 PCGs and 2 rRNAs) in phylogenetic reconstruction has not been systematically studied across Annelida. Current studies are limited to small taxonomic groups. For example, Zhang et al. (2018) analyzed 16 scale worms from Aphroditiformia and found gene order to be generally conserved, except in deep-sea polynoids. Alves et al. (2020) examined 24 species of Nereididae and found that shared gene-order patterns supported monophyletic subfamilies. Sun et al. (2021) studied 10 species of *Hydroides* and observed different gene orders among species of the same genus. Hektoen et al. (2024) reported consistent gene order among 14 species of the *Prionospio* complex in Norwegian waters. More recently, Struck et al. (2023) analyzed mitochondrial gene orders of 168 annelid species and concluded that gene order provides limited phylogenetic resolution but can offer additional evidence supporting the monophyly of certain groups. In our study, the arrangement of 13 PCGs and 2 rRNAs in Terebelliformia is relatively conserved (Fig. 4A). Different clades share similar mitochondrial gene arrangements even when tRNAs are included (Suppl. material 2: Fig. S6), providing additional evidence supporting their respective clades.

A comprehensive revision of *Loimia* taxonomy is needed

The need to revise *Loimia* is widely acknowledged (Jirkov and Leontovich 2017; Lavesque et al. 2017; Martin et al. 2022; Hutchings et al. 2024). Nearly all recognized species of *Loimia* have been described solely on morphological grounds, without incorporating genetically delimited boundaries. Consequently, *Loimia* taxonomy remains at the so-called "Species Delimitation 1.0" stage (based mainly on morphology), lagging behind the "Species Delimitation 4.0" stage (based mainly on genomic data) (Karbstein et al. 2024). As we enter the Species Delimitation 4.0 era (Karbstein et al. 2024), genomic data are needed to validate morphology-based species and consolidate the taxonomic status of previously described species. This will also enable reciprocal illumination of the most phylogenetically useful morphological characters and gene families. Achieving these goals requires broad collaboration within the field.

In this study and in Yang et al. (2024), we established a streamlined bioinformatics workflow to process next-generation sequencing data, including genome skimming to obtain complete mitogenomes, complete 18S–28S rRNA gene clusters (18S, ITS1, 5.8S, ITS2, 28S), and complete histone gene sequences (H3, H4, H2A, H2B) (Fig. 1). We hope this methodology will accelerate the systematic revision of *Loimia* by integrating genomic approaches with detailed morphological studies, including morphometrics and live observations (e.g., coloration).

The necessity of curating data stored in public repositories

NCBI databases, such as the nucleotide sequence database, are invaluable resources but also potential sources of error. Careful data curation is essential, as inadequate verification can significantly increase error rates, particularly in downstream analyses, and compromise research outcomes (Yang et al. 2024; Yu et al. 2024).

In our study, we identified several mislabeled taxa. For example, *Lanice conchilega* (Pallas, 1766) (specimen voucher: DUV-4f; COI: OQ053049; 28S: OQ071312, OQ071255) and *Loimia arborea* Moore, 1903 (species voucher: BC016; COI: HM473449; 28S: HM473232). The COI sequence HM473449 has been widely used in taxonomic studies of *Loimia*, including Wang et al. (2020) and Martin et al. (2022). Notably, in Wang et al. (2020), where two taxa were used as outgroups, *L. arborea* (HM473449) did not cluster with *Loimia*. Conversely, in Martin et al. (2022), where only one taxon (*Lanice* sp.) was used as an outgroup, *L. arborea* (HM473449) clustered within *Loimia*. Our analyses reveal that uncritical use of these sequences would have resulted in errors. This finding underscores the importance of meticulous data curation when retrieving sequences from public repositories.

Author contributions

Sheng Zeng: methodology, software, formal analysis, writing – original draft, writing – review and editing; Deyuan Yang: conceptualization, funding, methodology,

software, formal analysis, writing – original draft (Abstract, Introduction, parts of Methods, Results, and Discussion), supervision; Caifang He: methodology, formal analysis (morphological section), writing – original draft (morphological section); Yanan Sun: writing – review and editing, funding; Christopher J. Glasby: writing – review and editing, supervision; Yanjie Zhang: methodology, resources, investigation, writing – review and editing, funding, supervision.

Acknowledgments

This work was financially supported by the Shandong Provincial Natural Science Foundation (Grant No. 2023HWYQ-101), the Taishan Scholars Program (Grant No. tsqn:202306288), and the National Natural Science Foundation of China (Grant No. 42306132).

References

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F (2020) MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Molecular Ecology Resources 20(4): 892–905. https://doi.org/10.1111/1755-0998.13160

Alves PR, Halanych KM, Santos CSG (2020) The phylogeny of Nereididae (Annelida) based on mitochondrial genomes. Zoologica Scripta 49(3): 366–378. https://doi.org/10.1111/zsc.12413

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [accessed March 26.2025]

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

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15): 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Caullery M (1944) Polychètes sédentaire de l'Expédition du Siboga: Ariciidae, Spionidae, Chaetopteridae, Chloraemidae, Opheliidae, Oweniidae, Sabellariidae, Sternaspidae, Amphictenidae, Ampharetidae, Terebellidae. Siboga-Expeditie, Leiden 24: 1–1.

Chan PP, Lin BY, Mak AJ, Lowe TM (2021) tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. Nucleic Acids Research 49(16): 9077–9096. https://doi.org/10.1093/nar/gkab688

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17): i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: An integrative toolkit developed for interactive analyses of big biological data. Molecular Plant 13(8): 1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chen Z, Baeza JA, Chen C, Gonzalez MT, González VL, Greve C, Kocot KM, Arbizu PM, Moles J, Schell T, Schwabe E, Sun J, Wong NLWS, Yap-Chiongco M, Sigwart JD (2025) A genome-based phylogeny for Mollusca is concordant with fossils and morphology. Science 387(6737): 1001–1007. https://doi.org/10.1126/science.ads0215

Collins G, Schneider C, Bostjancic LL, Burkhardt U, Christian A, Decker P, Ebersberger I, Hohberg K, Lecompte O, Merges D, Muelbaier H, Romahn J, Rombke J, Rutz C, Schmelz R, Schmidt A, Theissinger K, Veres R, Lehmitz R, Pfenninger M, Balint M (2023) The MetaInvert soil invertebrate genome resource provides insights into below-ground biodiversity and evolution. Communications Biology 6(1): 1241. https://doi.org/10.1038/s42003-023-05621-4

de Matos Nogueira JM, Hutchings PA, Fukuda MV (2010) Morphology of terebelliform polychaetes (Annelida: Polychaeta: Terebelliformia), with a focus on Terebellidae. Zootaxa 2460(1): 181–185. https://doi.org/10.11646/zootaxa.2460.1.1

de Matos Nogueira JM, Fitzhugh K, Hutchings P (2013) The continuing challenge of phylogenetic relationships in Terebelliformia (Annelida: Polychaeta). Invertebrate Systematics 27(2): 186–238. https://doi.org/10.1071/IS12062

de Matos Nogueira JM, Hutchings P, Carrerette O (2015) Terebellidae (Annelida, Terebelliformia) from Lizard Island, Great Barrier Reef, Australia. Zootaxa 4019(1): 484–576. https://doi.org/10.11646/zootaxa.4019.1.18

Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M (2019) Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Research 47(20): 10543–10552. https://doi.org/10.1093/nar/gkz833

Du S, Tihelka E, Yu D, Chen WJ, Bu Y, Cai C, Engel MS, Luan YX, Zhang F (2024) Revisiting the four Hexapoda classes: Protura as the sister group to all other hexapods. Proceedings of the National Academy of Sciences of the United States of America 121(39): e2408775121. https://doi.org/10.1073/pnas.2408775121

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32(19): 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fauchald K (1977) The polychaete worms. Definitions and keys to the orders, families and genera. Natural History Museum of Los Angeles County, Science Series.

Gao Z, Lu Y, Chong Y, Li M, Hong J, Wu J, Wu D, Xi D, Deng W (2024) Beef cattle genome project: Advances in genome sequencing, assembly, and functional genes discovery. International Journal of Molecular Sciences 25(13): e7147. https://doi.org/10.3390/ijms25137147

Garraffoni ARS, de Garcia Camargo M (2007) A new application of morphometrics in a study of the variation in uncinal shape present within the Terebellidae (Polychaeta): A reevaluation from digital images. Cahiers de Biologie Marine 48(3): e229. https://doi.org/10.21411/CBM.A.9BB66E70

Garraffoni AR, Lana PC (2008) Phylogenetic relationships within the Terebellidae (Polychaeta: Terebellida) based on morphological characters. Invertebrate Systematics 22(6): 605–626. https://doi.org/10.1071/IS07006

Glasby CJ, Hsieh H-L (2006) New species and new records of the *Perinereis nuntia* species group (Nereididae: Polychaeta) from Taiwan and other Indo-West Pacific shores. Zoological Studies, 553–577. https://doi.org/10.3897/zookeys.1132.87629

Glasby CJ, Hutchings PA, Hall K (2004) Assessment of monophyly and taxon affinities within the polychaete clade Terebelliformia

(Terebellida). Journal of the Marine Biological Association of the United Kingdom 84(5): 961–971. https://doi.org/10.1017/S0025315404010252h

Hektoen MM, Bakken T, Ekrem T, Radashevsky RI, Dunshea G (2024) Species delimitation and phylogenetic relationships of the *Prionospio* complex (Annelida, Spionidae) in the Northeast Atlantic. Zoologica Scripta 53(3): 358–375. ttps://doi.org/10.1111/zsc.12648

Hutchings P, Nogueira JMdM, Carrerette O (2017) Terebellidae sl: Polycirridae Malmgren, 1866, Terebellidae Johnston, 1846, Thelepodidae Hessle, 1917, Trichobranchidae Malmgren, 1866, and Telothelepodidae Nogueira, Fitzhugh & Hutchings, 2013. In: Schmidt-Rhaesa A (Ed.) Handbook of Zoology Online. De Gruyter, Berlin, Boston.

Hutchings P, Carrerette O, Nogueira JM, Hourdez S, Lavesque N (2021) The Terebelliformia-recent developments and future directions. Diversity 13(2): 1–60. https://doi.org/10.3390/d13020060

Hutchings P, Daffe G, Flaxman B, Rouse GW, Lavesque N (2024) A new species of Loimia (Annelida, Terebellidae) from Papua New Guinea, with comments on other species recorded in the region. Ocean and Coastal Research 72(suppl 1). https://doi.org/10.1590/2675-2824072.23057

Hutchings P, Daffe G, Glasby C, Lavesque N (2025) Spaghetti worms from the reef: two new species of *Loimia* (Polychaeta: Terebellidae) from Bora-Bora and Moorea (Society Islands, French Polynesia) and a range extension of *L. tuberculata* Nogueira, Hutchings & Carrerette, 2015. Zootaxa 5583(2): 328–352. https://doi.org/10.11646/zootaxa.5583.2.6

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(1): 1–31. https://doi.org/10.1186/s13059-020-02154-5

Jirkov I, Leontovich M (2017) Review of genera within the *Axionice*/*Pista* complex (Polychaeta, Terebellidae), with discussion of the taxonomic definition of other Terebellidae with large lateral lobes. Journal of the Marine Biological Association of the United Kingdom 97(5): 911–934. https://doi.org/10.1017/S0025315417000923

Jombart T, Kendall M, Almagro-Garcia J, Colijn C (2017) treespace: Statistical exploration of landscapes of phylogenetic trees. Molecular Ecology Resources 17(6): 1385–1392. https://doi.org/10.1111/1755-0998.12676

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. Nature Methods 14(6): 587–589. https://doi.org/10.1038/nmeth.4285

Karbstein K, Kösters L, Hodač L, Hofmann M, Hörandl E, Tomasello S, Wagner ND, Emerson BC, Albach DC, Scheu S, Bradler S, de Vries J, Irisarri I, Li H, Soltis P, Mäder P, Wäldchen J (2024) Species delimitation 4.0: Integrative taxonomy meets artificial intelligence. Trends in Ecology & Evolution 39(8): 771–784. https://doi.org/10.1016/j.tree.2023.11.002

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35(6): 1547–1549. https://doi.org/10.1093/molbev/msy096

Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Research 32(1): 11–16. https://doi.org/10.1093/nar/gkh152

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

Lavesque N, Bonifácio P, Londoño-Mesa MH, Le Garrec V, Grall J (2017) *Loimia ramzega* sp. nov., a new giant species of Terebellidae (Polychaeta) from French waters (Brittany, English Channel). Journal of the Marine Biological Association of the United Kingdom 97(5): 935–942. https://doi.org/10.1017/S0025315417000571

Lavesque N, Daffe G, Londoo-Mesa MH, Hutchings P (2021) Revision of the French Terebellidae sensu stricto (Annelida, Terebelliformia), with descriptions of nine new species. Zootaxa 5038(1): 1–63. https://doi.org/10.11646/zootaxa.5038.1.1

Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Research 52(W1): gkae268. https://doi.org/10.1093/nar/gkae268

Liu H, Steenwyk JL, Zhou X, Schultz DT, Kocot KM, Shen XX, Rokas A, Li Y (2024) A taxon-rich and genome-scale phylogeny of Opisthokonta. PLoS Biology 22(9): e3002794. https://doi.org/10.1371/journal.pbio.3002794

Malmgren AJ (1866) Nordiska Hafs-Annulater.[part three of three]. Öfversigt af Königlich Vetenskapsakademiens förhandlingar, Stockholm 22: 355.

Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6): 764–770. https://doi.org/10.1093/bioinformatics/btr011

Martin D, Capa M, Martínez A, Costa AC (2022) Taxonomic implications of describing a new species of *Loimia* (Annelida, Terebellidae) with two size-dependent morphotypes. European Journal of Taxonomy 833: 60–96. https://doi.org/10.5852/ejt.2022.833.1887

Martínez-Redondo GI, Vargas-Chávez C, Eleftheriadi K, Benítez-Álvarez L, Vázquez-Valls M, Fernández R (2024) MATEdb2, a collection of high-quality metazoan proteomes across the Animal Tree of Life to speed up phylogenomic studies. Genome Biology and Evolution 16(11): evae235. https://doi.org/10.1093/gbe/evae235

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71(2): 491–499. https://doi.org/10.1016/0378-1119(88)90066-2

Meng G, Li Y, Yang C, Liu S (2019) MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Research 47(11): e63. https://doi.org/10.1093/nar/gkz173

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32(1): 268–274. https://doi.org/10.1093/molbev/msu300

Nygren A, Eklöf J, Pleijel F (2009) Arctic-boreal sibling species of *Paranaitis* (Polychaeta, Phyllodocidae). Marine Biology Research 5(4): 315–327. https://doi.org/10.1080/17451000802441301

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Jour-

nal of Molecular Evolution 41(3): 353–358. https://doi.org/10.1007/BF01215182

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A (2020) Using SPAdes de novo assembler. Current Protocols in Bioinformatics 70(1): e102. https://doi.org/10.1002/cpbi.102

R Core Team (2018) A language and environment for statistical computing. In: Team RC (Ed.) R Foundation for Statistical Computing. Viena.

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nature Communications 11(1): e1432. https://doi.org/10.1038/s41467-020-14998-3

Ranwez V, Douzery EJ, Cambon C, Chantret N, Delsuc F (2018) MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Molecular Biology and Evolution 35(10): 2582–2584. https://doi.org/10.1093/molbev/msy159

Read G, Fauchald K (2025) WoRMS. World Polychaeta Database. https://www.marinespecies.org/aphia.php?p=taxdetails&id=129700 [accessed Jan 1st.2025]

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539–542. https://doi.org/10.1093/sysbio/sys029

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

Shen XX, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT, Boudouris JT, Schneider RM, Langdon QK, Ohkuma M, Endoh R, Takashima M, Manabe RI, Cadez N, Libkind D, Rosa CA, DeVirgilio J, Hulfachor AB, Groenewald M, Kurtzman CP, Hittinger CT, Rokas A (2018) Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell 175(6): 1533–1545. https://doi.org/10.1016/j.cell.2018.10.023

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics (Oxford, England) 31(19): 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Steenwyk JL, Buida TJ III, Gonçalves C, Goltz DC, Morales G, Mead ME, LaBella AL, Chavez CM, Schmitz JE, Hadjifrangiskou M, Li Y, Rokas A (2022) BioKIT: A versatile toolkit for processing and analyzing diverse types of sequence data. Genetics 221(3): iyac079. https://doi.org/10.1093/genetics/iyac079

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. https://doi.org/10.3390/biology9040073

Struck TH (2019) 2. Phylogeny. In: Purschke G, Böggemann M, Westheide W (Eds) Handbook of Zoology: Annelida (Vol. I): Annelida Basal Groups and Pleistoannelida, Sedentaria I. Walter de Gruyter GmbH & Co KG, Berlin/Boston, 37–68. https://doi.org/10.1515/9783110291582-002

Struck TH, Purschke G, Halanych KM (2006) Phylogeny of Eunicida (Annelida) and exploring data congruence using a partition addition bootstrap alteration (PABA) approach. Systematic Biology 55(1): 1–20. https://doi.org/10.1080/10635150500354910

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(4): 925–945. https://doi.org/10.1093/sysbio/syad023

Sun Y, Sun J, Yang Y, Lan Y, Ip JC, Wong WC, Kwan YH, Zhang Y, Han Z, Qiu JW, Qian PY (2021) Genomic signatures supporting the symbiosis and formation of chitinous tube in the deep-sea tubeworm *Paraescarpia echinospica*. Molecular Biology and Evolution 38(10): 4116–4134. https://doi.org/10.1093/molbev/msab203

Wang W, Sui J, Kou Q, Li X-Z (2020) Review of the genus *Loimia* Malmgren, 1866 (Annelida, Terebellidae) from China seas with recognition of two new species based on integrative taxonomy. PeerJ 8: e9491. https://doi.org/10.7717/peerj.9491

Wickham H (2016) Data Analysis. In: Wickham H (Ed.) ggplot2: Elegant Graphics for Data Analysis. Springer International Publishing, Cham, 189–201. https://doi.org/10.1007/978-3-319-24277-4_9

Xu Y, Zeng S, Meng Y, Yang D, Yang S (2024) The mitochondrial genome of *Hua aristarchorum* (Heude, 1889) (Gastropoda, Cerithioidea, Semisulcospiridae) and its phylogenetic implications. ZooKeys 1192: 237–255. https://doi.org/10.3897/zookeys.1192.116269

Yang D, Zeng S, Wang Z, Zhang Y, Yang D, Glasby CJ, Hwang JS, Cai L (2024) Molecular systematics of *Perinereis* and an investigation of the status and relationships of the cultured species *Perinereis wilsoni* Glasby & Hsieh, 2006 (Annelida, Nereididae). Zoosystematics and Evolution 100(4): 1297–1314. https://doi.org/10.3897/zse.100.127201

Yu W, Luo H, Yang J, Zhang S, Jiang H, Zhao X, Hui X, Sun D, Li L, Wei XQ, Lonardi S, Pan W (2024) Comprehensive assessment of 11 de novo HiFi assemblers on complex eukaryotic genomes and metagenomes. Genome Research 34(2): 326–340. https://doi.org/10.1101/gr.278232.123

Zhang Y, Sun J, Rouse GW, Wiklund H, Pleijel F, Watanabe HK, Chen C, Qian P-Y, Qiu J-W (2018) Phylogeny, evolution and mitochondrial gene order rearrangement in scale worms (Aphroditiformia, Annelida). Molecular Phylogenetics and Evolution 125: 220–231. https://doi.org/10.1016/j.ympev.2018.04.002

Supplementary material 1

Extraction methods

Authors: Sheng Zeng, Deyuan Yang, Caifang He, Yanan Sun, Christopher J. Glasby, Yanjie Zhang

Data type: pdf

Explanation note: 1.1 RNA extraction. 1.2 DNA extraction.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.101.161344.suppl1

Supplementary material 2

Supplementary figures

Authors: Sheng Zeng, Deyuan Yang, Caifang He, Yanan Sun, Christopher J. Glasby, Yanjie Zhang

Data type: pdf

Explanation note: **fig. S1.** The reports from FastQC. (a) Basic information on clean data, including duplicate reads (%, Dups), average GC content (%, GC), and total sequences (millions, M Seqs). (b) Sequence counts for each sample. Duplicate read counts are an estimate only. **fig. S2.** Uncini and notochaetae of *Loimia borealis*. (A) Uncini, segment 18, double rows of uncini; (B) uncini, segment 20; (C) uncini, segment 21; (D) uncini, segment 14; (E) uncini, segment 18; (F) uncini, segment 14; (G) uncini, segment 5; (H) uncini, segment 23; (I) uncini, segment 48; (J) notochaetae, segment 6; (K) notochaetae, segment 8. Scales: (A–D, H, J, and K) 50 µm; (E–G and I) 20 µm. **fig. S3.** Details of *Loimia borealis* under SEM. (A) uncini from segment 5, single rows of uncini; (B, C) uncini from segment 8, single rows of uncini; (D–E) uncini from segment 9, single rows of uncini; (F–H) uncini from posterior segment; (L–N) notochaetae from segment 18. Scales: (A–D and F–H) 10 µm; (E, K, and N) 1 µm; (I and L) 100 µm; (J and M) 10 µm. **fig. S4.** Relative synonymous codon usage of *Loimia borealis* PQ774336, *Loimia medusa* NC 088773, and *Loimia medusa* OM809695. The codon families are provided under the x-axis. **fig. S5.** Heterogeneity of sequence composition of different datasets. **fig. S6.** All genes (13 PCGs, 2 rRNAs, and tRNAs) in the order of Terebellida mitogenomes. **fig. S7.** Bayesian inference tree (left) and maximum likelihood tree (right) inferred from the cox1 dataset. **fig. S8.** Bayesian inference tree (left) and maximum likelihood tree (right) inferred from the 16S dataset. **fig. S9.** Bayesian inference tree (left) and maximum likelihood tree (right) inferred from the 18S dataset. **fig. S10.** Bayesian inference tree (left) and maximum likelihood tree (right) inferred from the 28S dataset. **fig. S11.** Bayesian inference tree (left) and maximum likelihood tree (right) inferred from the H3 dataset.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.101.161344.suppl2

Supplementary material 3

Supplementary tables

Authors: Sheng Zeng, Deyuan Yang, Caifang He, Yanan Sun, Christopher J. Glasby, Yanjie Zhang

Data type: xlsx

Explanation note: **table S1.** The BLAST results of mitochondrial gene, 18S, 28S, ITS1-5.8S-ITS2, and H3 genes (only 20 sequences are shown). **table S2.** List of 31 sequences and two outgroups used in this paper. **table S3.** The BLAST results of the Terebellida mitogenome (only 10 sequences are shown). **table S4.** cox1, 16S, 18S, 28S, ITS, and H3 gene sequence information of Terebellidae. **table S5.** Original and trimal lengths of sequences. **table S6.** Best partitioning schemes and models based on different datasets for maximum likelihood and Bayesian inference analysis. **table S7.** Summary statistics for multiple alignment of various datasets. **table S8.** The sequences used for p-distance. **table S9.** Features of the *Loimia borealis* mitogenome. **table S10.** Composition and skewness of the *Loimia borealis* mitogenome. **table S11.** Codon numbers and relative synonymous codon usage (RSCU) of 13 PCGs in the *Loimia borealis* mitogenome.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zse.101.161344.suppl3