

First mitochondrial genomes of Capitellidae and Opheliidae (Annelida) and their phylogenetic placement

Genki Kobayashi^a , Hajime Itoh^b and Nobuyoshi Nakajima^b 

^aSeto Marine Biological Laboratory, Field Science Education and Research Center, Kyoto University, Nishimuro, Japan; ^bNational Institute for Environmental Studies, Tsukuba, Japan

ABSTRACT

The complete mitochondrial genomes of *Notomastus* sp. (15,776 bp) (Annelida: Capitellidae) and *Armandia* sp. (18,538 bp) (Annelida: Opheliidae) were assembled for the first time. A group II intron (303 bp) was found in *cox1* of *Notomastus* sp. A phylogenetic analysis revealed that *Notomastus* sp. and *Armandia* sp. were monophyletic, and this clade was clustered with echiurans, although the possibility of the effect of long-branch attraction should be considered.

ARTICLE HISTORY

Received 23 December 2021
Accepted 11 March 2022

KEYWORDS



Capitellida; Opheliida;
polychaete; mitoge-
nome; Sedentaria

Annelida is an ecologically diverse group with over 20,000 described species (Capa and Hutchings 2021). Sedentaria includes derivative annelid species that inhabit sediment and possess a cylindrical body with reduced parapodia. Although mitogenomes are often used for inferring the phylogenetic relationships of closely related annelids (Sun et al. 2021), several annelid families are yet undetermined. In this study, we determined the mitogenomes of two families of Sedentaria, namely Capitellidae and Opheliidae, for the first time, examined the features of mitogenomes, and reconstructed their phylogenetic relationships.

The specimens were collected from the intertidal zone in Wakayama, Japan (33°41'N, 135°21'E–135°22'E) and deposited at the Rishiri Town Museum (voucher numbers: RTMan168 (*Notomastus* sp.) and RTMan169 (*Armandia* sp.); contact person Shinri Tomioka, rishiritownmuseum@town.rishiri.hokkaido.jp). DNA extraction and long polymerase chain reaction (PCR) procedures were performed according to the methods used by Kobayashi et al. (2021). Primer sets were designed for each species to amplify nearly complete mitogenomes of *Notomastus* sp. and *Armandia* sp. based on their partial 16S sequences, which were determined using 16SarL/16SbrH (Palumbi 1996) (*Notomastus* sp.) or 16Sann-f2/16Sann-r2 (Kobayashi and Kojima 2021) (*Armandia* sp.). 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. Mitogenomes were assembled with GetOrganelle v1.7.1a (Jin et al. 2020) using the 16S sequences of each species as seed sequences. Assembled contigs and the 16S gene sequence were manually concatenated and the complete

mitogenome of *Notomastus* sp. was obtained. The second assembly was conducted for *Armandia* sp. using the *nad5* gene sequence, which was detected with the MITOS2 web server (Donath et al. 2019) from the contig of the first assembly, as a seed sequence, since only a partial mitogenome was obtained from the first assembly. The partial 16S sequence of *Armandia* sp., which was not determined by the above method, was confirmed by Sanger sequencing using 16Sar/16Sb-Arm (16Sb-Arm: 5'-CGYCGGTCTRAACTCAGCTC-3'; this study). Then, the complete mitogenome of *Armandia* sp. was obtained by concatenating the 16S sequence and the contigs obtained in the second assembly. The PCGs and tRNAs were identified using the MITOS2 web server. The tRNAs were also detected using ARWEN (Laslett and Canback 2008) implemented in ARAGORN (Laslett and Canback 2004). The cloverleaf secondary structures of tRNAs were validated using ARWEN. The annotated mitogenome sequences were deposited in GenBank through DNA Data Bank of Japan with accession numbers LC661358 (*Notomastus* sp.) and LC661359 (*Armandia* sp.).

Maximum likelihood phylogeny was reconstructed using amino acid sequences of 13 PCGs according to Kobayashi et al.'s (2021) procedure using the following softwares: IQ-TREE v1.6.12 (Nguyen et al. 2015) for phylogenetic analysis, ModelFinder (Kalyaanamoorthy et al. 2017) for selecting the best-fit substitution models for each of 13 PCGs, SeqKit (Shen et al. 2016) for translating nucleotide sequences, MAFFT v7 (Katoh and Standley 2013) for sequence alignment, and FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) for illustrating the phylogenetic tree. A dataset consisting of 41

CONTACT Genki Kobayashi  genkikobayashi5884@gmail.com  Seto Marine Biological Laboratory, Field Science Education and Research Center, Kyoto University, 459 Shirahama, Nishimuro, Wakayama 649-2211, Japan

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

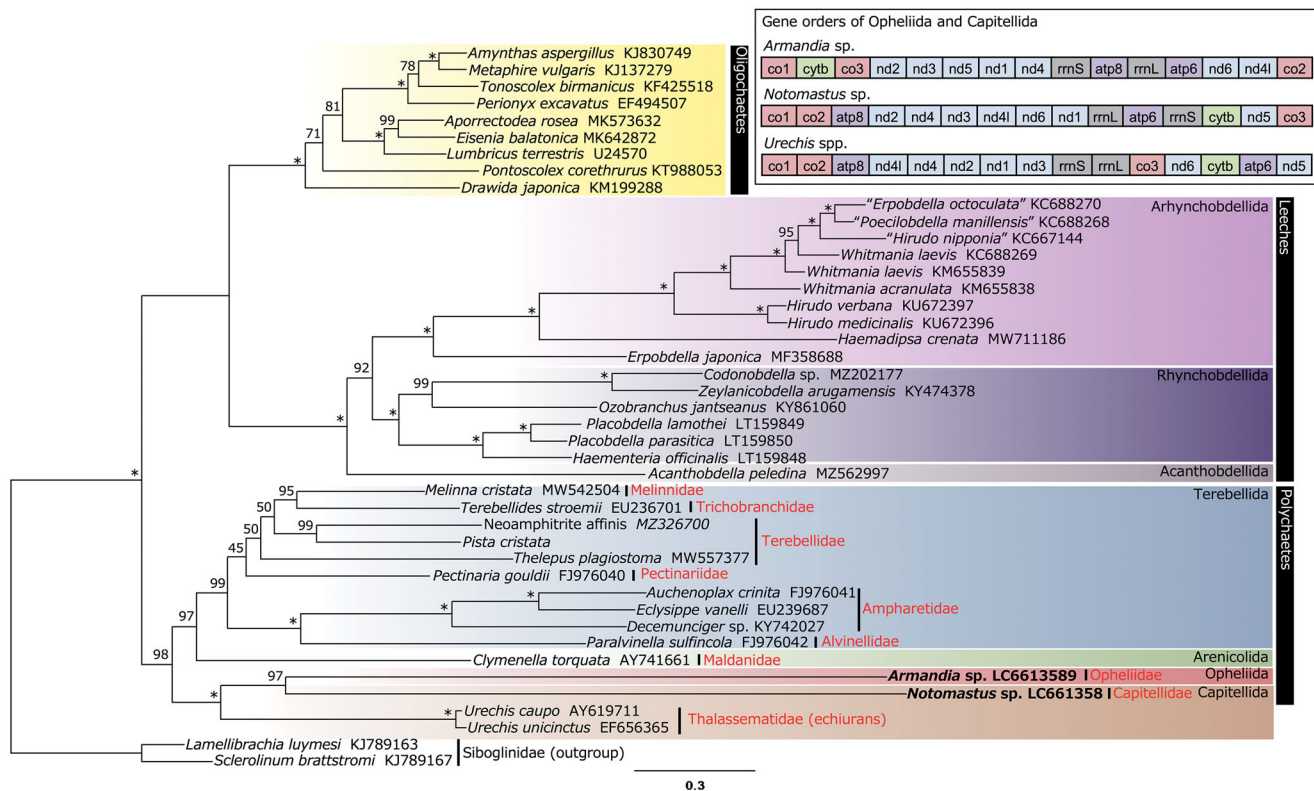


Figure 1. Maximum likelihood phylogeny of a subset of Sedentaria based on the amino acid sequences of 13 mitochondrial genome PCGs (4185 characters). The best-fit substitution models for each of the PCGs were selected by ModelFinder using the Bayesian Information Criterion as follows: mtART + I + G4 (*nad1*), mtInV + F + G4 (*atp6*), mtInV + F + I + G4 (*nad2*, *nad4*, and *nad5*), mtMet + G4 (*nad6*), mtMet + I + G4 (*nad3* and *nad4l*), mtZOA + R4 (*cox1*); mtZOA + R5 (*cytb*), mtZOA + I + G4 (*cox2*), and mtZOA + G4 (*cox3*). Asterisks indicate Bootstrap value = 100%. Gene orders of the PCGs and rRNAs of Capitellida and Ophiellida are shown above the tree. The orders of polychaetes and the family of echiurans follow the classification of Struck (2019) and Goto et al. (2020), respectively.

mitogenome sequences of Sedentaria and 2 outgroups (Siboglinidae) was obtained from GenBank.

The complete mitogenomes of *Notomastus* sp. and *Armandia* sp. consisted of 15,776 bp (AT content = 57.8%) and 18,538 bp (AT = 60.2%), respectively. Thirteen PCGs and two rRNAs were annotated for both species. *Notomastus* sp. possesses 24 tRNAs and *Armandia* sp. possesses 23 tRNAs, including a tRNA corresponding to the stop codon (anticodon: TTA). Both the species have tRNAs coded on the negative strand (*trnS2* and *trnD* in *Notomastus* sp. and *trnStop* in *Armandia* sp.), which is unusual in Annelida (see Daffe et al. 2021). An intron (303 bp) was found in *cox1* of *Notomastus* sp. based on the alignment of the dataset. This intron was regarded as a group II intron since it included a sequence (starting with GTGCG and ending with AG) similar to the motif of group II intron (starting with GUGYG and ending with AY) (Bonen and Vogel 2001). This intron was inserted in a unique position (1115–1417) and was shorter than other known annelid group II introns in the *cox1* region (Vallès et al. 2008; Richter et al. 2015; Bernardino et al. 2017; Kobayashi et al. 2022). The gene orders of PCGs (Figure 1) were unique for each species among annelids (see Sun et al. 2021). The resultant tree revealed that *Notomastus* sp. and *Armandia* sp. were monophyletic, and this clade was clustered with the echiuran species of *Urechis* (Thalassematidae) (Figure 1). This result was inconsistent with that of previous studies that have shown a sister relationship between

echiurans and capitellids (e.g. Struck et al. 2007). Although the close relationship between *Notomastus* sp. and *Armandia* sp. in this study might be affected by long-branch attraction, TreeShrink v1.3.9 (Mai and Mirarab 2018) did not identify any abnormally long branches (the option *-q* was set to 0.05 or 0.10³).

Authors' contributions

GK conceived the study. All authors were involved in the analysis and interpretation of the data; the drafting of the paper, revising it critically for intellectual content; and the final approval of the version to be published; and that all authors agree to be accountable for all aspects of the work.

Acknowledgments

We thank Ryutaro Goto, Luna Yamamori, Mariko Kawamura, Kohki Yamamoto, and Takahiro Sugiyama (Seto Marine Biological Laboratory, Kyoto University) for their help in field sampling; Shinri Tomioka (Rishiri Town Museum) for her help in the identification of *Notomastus* sp. and deposition of specimens; and Naoto Jimi (Sugashima Marine Biological Laboratory, Nagoya University) for providing information on ophiellid gene sequences.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Project for the Promotion of Environmental Genomics Studies by NIES [1620AQ007].

ORCID

Genki Kobayashi  <http://orcid.org/0000-0002-5938-0824>
Nobuyoshi Nakajima  <http://orcid.org/0000-0002-0040-9557>

Data availability statement

The sequencing data that support the findings of this study are openly available in GenBank of NCBI under the accession no. LC661358 (*Notomastus* sp.) and LC661359 (*Armandia* sp.). The associated BioProject, SRA, and BioSample numbers are PRJDB12658, DRA013119, and SAMD00428326 (*Notomastus* sp.) and PRJDB12659, DRA013120, SAMD00428327 (*Armandia* sp.), respectively.

References

- Bernardino AF, Li Y, Smith CR, Halanych KM. 2017. Multiple introns in a deep-sea annelid (Decemunciger: Ampharetidae) mitochondrial genome. *Sci Rep.* 7(1):1–8.
- Bonen L, Vogel J. 2001. The ins and outs of group II introns. *Trends Genet.* 17(6):322–331.
- Capa M, Hutchings P. 2021. Annelid diversity: historical overview and future perspectives. *Diversity.* 13(3):129.
- Daffe G, Sun Y, Ahyong ST, Kupriyanova EK. 2021. Mitochondrial genome of *Sabella spallanzanii* (Gmelin, 1791) (Sabellida: Sabellidae). *Mitochondrial DNA B Resour.* 6(2):499–501.
- 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 Res.* 47(20):10543–10552.
- Goto R, Monnington J, Sciberras M, Hirabayashi I, Rouse GW. 2020. Phylogeny of Echiura updated, with a revised taxonomy to reflect their placement in Annelida as sister group to Capitellidae. *Invert Systematics.* 34(1):101–111.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology.* 21: 241.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14(6):587–589.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kobayashi G, Kojima S. 2021. *Travisia sanrikuensis*, a new species of Traviidae (Annelida) from the lower bathyal zone of the northwestern Pacific. *SpecDiv.* 26(2):131–136.
- Kobayashi G, Itoh H, Fukuda H, Kojima S. 2021. The complete mitochondrial genome of the sand bubbler crab *Scopimera globosa* and its phylogenetic position. *Genomics.* 113(1):831–839.
- Kobayashi G, Itoh H, Kojima S. 2022. Mitogenome of a stink worm (Annelida: Traviidae) includes degenerate group II intron that is also found in five congeneric species. *Scientific Reports.* 12: 4449.
- Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32(1): 11–16.
- Laslett D, Canback B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics.* 24(2): 172–175.
- Mai U, Mirarab S. 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics.* 19(5):272.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Palumbi SR. 1996. Nucleic acids II: the polymerase chain reaction. In: Hillis DM, Mortiz C, Mable BK, editors. *Molecular systematics.* Sunderland (MA): Sinauer & Associates Inc.; p. 205–247.
- Richter S, Schwarz F, Hering L, Böggemann M, Bleidorn C. 2015. The utility of genome skimming for phylogenomic analyses as demonstrated for glycerid relationships (Annelida, Glyceridae). *Genome Biol Evol.* 7(12):3443–3462.
- Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One.* 11(10):e0163962.
- Struck TH. 2019. Phylogeny. In: Purschke G, Böggemann M, Westheide W, editors. *Handbook of zoology/Annelida: basal groups and Pleistoannelida, Sedentaria I.* Berlin (Germany): De Gruyter; p. 37–68.
- Struck TH, Schult N, Kusen T, Hickman E, Bleidorn C, McHugh D, Halanych KM. 2007. Annelid phylogeny and the status of Sipuncula and Echiura. *BMC Evol Biol.* 7:57.
- Sun Y, Daffe G, Zhang Y, Pons J, Qiu J-W, Kupriyanova EK. 2021. Another blow to the conserved gene order in Annelida: evidence from mitochondrial genomes of the calcareous tubeworm genus *Hydroides*. *Mol Phylogenet Evol.* 160:107124.
- Vallès Y, Halanych KM, Boore JL. 2008. Group II introns break new boundaries: presence in a bilaterian's genome. *PLoS One.* 3(1):e1488.