



Data Article

The complete mitochondrial genome data of *Pholas orientalis* (Gmelin, 1791) from Malaysia

Hao Xuan Ho^a, Teek Foh Chong^a, Wei Lun Ng^{b,c,*}, Shiou Yih Lee^{a,**}

^aFaculty of Health and Life Sciences, INTI International University, Nilai, Negeri Sembilan, Malaysia

^bChina-ASEAN College of Marine Sciences, Xiamen University Malaysia, Sepang, Selangor, Malaysia

^cCollege of Ocean and Earth Sciences, Xiamen University, Xiamen, Fujian, China

ARTICLE INFO

Article history:

Received 26 November 2023

Revised 20 May 2024

Accepted 28 May 2024

Available online 2 June 2024

Dataset link: [Pholas orientalis mitochondrion, complete genome \(Original data\)](#)

Keywords:

Angelwing clam

Genetic resources

Marine organism

Mitogenome

Next-generation sequencing

Phylogenetic analysis

ABSTRACT

Pholas orientalis (angelwing clam) is a mollusc species found in the coastal areas of Southeast Asia. Despite its economic significance, genetic information on the species is lacking. In this study, a *P. orientalis* specimen was collected from Kedah, Malaysia, and its complete mitochondrial genome was assembled using whole-genome sequencing data generated on an DNBSEQ-G400 platform. The circular mitochondrial genome of *P. orientalis* is 18,995 bp in size and contains 12 protein-coding genes (PCGs), 22 tRNAs, two rRNAs, and three control regions (D-loops). All genes are located on the heavy strand. The mitogenome has a base composition of 25.4 % A, 41.5 % T, 22.1 % G, and 11 % C, exhibiting a bias towards AT content (66.9 %). The mitochondrial genomes of *P. orientalis* and 11 other Pholadoidea species were included in a phylogenetic analysis, which indicated that *P. orientalis* is closely related to *Xyloredo nooi*. The data reported in this study represents the first time that a *Pholas* mitochondrial genome has been reported. Such data will contribute to the better understanding of genetic relationships between *P. orientalis* and

* Corresponding author at: China-ASEAN College of Marine Sciences, Xiamen University Malaysia, Sepang, Selangor, Malaysia; College of Ocean and Earth Sciences, Xiamen University, Xiamen, Fujian, China.

** Corresponding author at: Faculty of Health and Life Sciences, INTI International University, Nilai, Negeri Sembilan, Malaysia.

E-mail addresses: weilun.ng@xmu.edu.my (W.L. Ng), shiouyih.lee@newinti.edu.my (S.Y. Lee).

its relatives, leading to informed conservation and sustainable utilization of the species.

© 2024 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>)

Specifications Table

Subject	Omics: Genomics
Specific subject area	Angelwing clam, Mitogenomics
Data format	Raw, Analyzed
Type of data	Table, Figure, GenBank, FASTA
Data collection	A <i>Pholas orientalis</i> specimen was collected from Kedah, Malaysia. Genomic DNA was extracted from the whole specimen using a DNA extraction kit and whole-genome sequencing was performed on a DNBSEQ-G400 platform generating 2 Gbp of PE150 reads. The assembly of mitochondrial genome was performed using GetOrganelle coupled with Geneious Prime, followed by annotation using MITOS Web Server and GeSeq and visualization using OrganellarGenomeDRAW. For phylogenetic analysis, multiple sequence alignment was performed using ClustalW followed by construction of a maximum likelihood tree using IQ-tree.
Data source location	Institution: Faculty of Health and Life Sciences, INTI International University City/Town/Region: Kuala Kedah, Kedah Country: Malaysia Latitude and longitude (and GPS coordinates, if possible) for collected samples/data: Latitude: 6.089021 N, Longitude: 100.278474E
Data accessibility	Repository name: NCBI Sequence Read Archive (SRA) Data identification number: SRR24236449 Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/SRR24236449 Repository name: NCBI BioProject Data identification number: PRJNA958121 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA958121 Repository name: NCBI BioSample Data identification number: SAMN34277396 Direct URL to data: https://www.ncbi.nlm.nih.gov/biosample/SAMN34277396 Repository name: NCBI GenBank Data identification number: OQ858578 Direct URL to data: https://www.ncbi.nlm.nih.gov/nuccore/OQ858578

1. Value of the Data

- This data describes the complete mitochondrial genome sequence of *P. orientalis* originating from Malaysia, which will be useful for future species identification and characterization, molecular taxonomy, species conservation, DNA barcoding and phylogenetic analysis.
- This data offers genetic information that will serve as a reference for data comparison and enables researchers to resolve taxonomic issues.
- This data provides protein-coding sequences that are valuable in phylogenetic reconstruction, as they improve molecular resolution and enhance statistical confidence compared to partial gene sequences.

2. Background

Pholas orientalis, commonly known as the angelwing clam, is a mollusc species that belongs to Pholadidae of Pholadoidea. It is a deep-burrowing clam that is predominantly found in the coastal areas of Southeast Asia [1]. *Pholas orientalis* is highly regarded as a delicacy in Asia [1–5]. Studies on the life cycle and breeding programs of the species have been conducted, and

the clam is being harvested sustainably along the Strait of Malacca, Malaysia [4]. Despite the economic significance, the genetic information of *P. orientalis* is lacking. To date, the phylogenetic relationships within Pholadoidea have been investigated using the 18S and 28S rRNA gene sequences, but the limited molecular resolution has led to an unresolved relationship among *Pholas* species and their relatives [6]. A subsequent attempt to reconstruct the phylogenetic tree using complete mitochondrial genome (mitogenome) data was carried out by involving most of the species of Pholadoidea, but members of Pholadidae were not included in the study [7]. Thus, the relationship between Pholadidae and its sister family is still unknown at the mitogenome level. In this study, we sequenced and assembled the first complete mitogenome sequence of *P. orientalis* to determine its genomic structure and phylogenetic relationships within Pholadoidea.

3. Data Description

The complete mitogenome sequence of *P. orientalis* was assembled with 2 Gbp whole-genome sequence data, resulting in a circular genome of 18,995 bp in length and a sequencing depth of 54x. It contains 12 protein-coding genes, 22 tRNA genes, two rRNA (12S rRNA and 16S rRNA) genes, and three D-loops (Table 1, Fig. 1). All the genes were annotated on the heavy strand. The

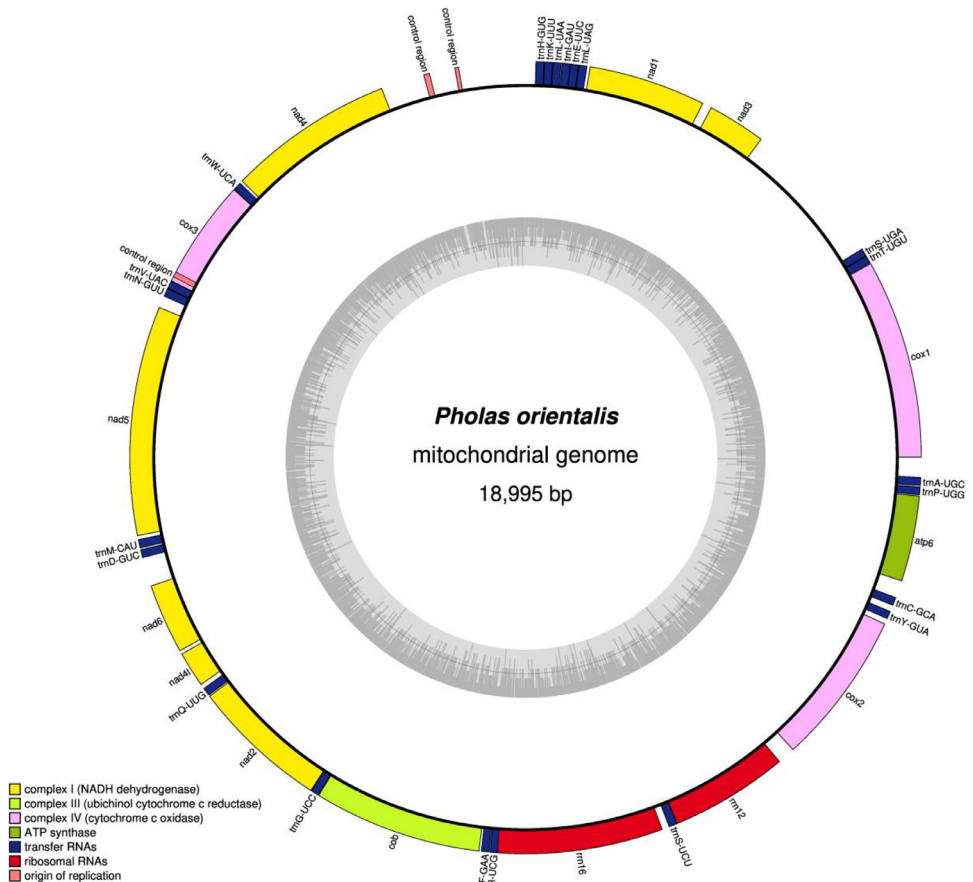


Fig. 1. The complete mitochondrial genome map of *Pholas orientalis*.

Table 1Mitogenome features of *P. orientalis*. Direction indicates the forward strand (+).

Gene (codon)	Position		Length (bp)	Direction	Start/stop codon
	Start	Stop			
cox1	1	1548	1548	+	ATT/TAA
trnT-UGU	1552	1615	64	+	
trnS-UGA	1616	1678	63	+	
nad3	2819	3274	456	+	GTG/TAG
nad1	3343	4254	912	+	ATG/TAA
trnL-UAG	4275	4339	65	+	
trnE-UUC	4341	4404	64	+	
trnI-GAU	4404	4469	66	+	
trnL-UAA	4473	4538	66	+	
trnK-UUU	4540	4603	64	+	
trnH-GUG	4605	4671	67	+	
control region	5261	5294	34	+	
control region	5489	5537	49	+	
nad4	5872	7164	1293	+	ATG/TAA
trnW-UCA	7183	7248	66	+	
cox3	7250	8101	852	+	ATG/TAG
control region	8021	8067	47	+	
trnV-UAC	8101	8164	64	+	
trnN-GUU	8166	8230	65	+	
nad5	8321	10,099	1779	+	TTC/TAG
trnM-CAU	10,126	10,195	70	+	
trnD-GUC	10,205	10,271	67	+	
nad6	10,493	11,038	546	+	ATT/TAA
nad4l	11,064	11,342	279	+	TTC/TAA
trnQ-UUG	11,376	11,443	68	+	
nad2	11,451	12,515	1065	+	ATA/TAA
trnG-UCC	12,518	12,581	64	+	
cob	12,582	13,892	1311	+	TTC/TAG
trnF-GAA	13,906	13,968	63	+	
trnR-UCG	13,972	14,034	63	+	
rnr16	14,011	15,314	1304	+	
trnS-UCU	15,368	15,435	68	+	
rnr12	15,434	16,354	921	+	
cox2	16,488	17,687	1200	+	ATA/TAA
trnY-GUA	17,728	17,791	64	+	
trnC-GCA	17,839	17,905	67	+	
atp6	18,037	18,702	666	+	ATG/TAA
trnP-UGG	18,708	18,772	65	+	
trnA-UGC	18,779	18,845	67	+	

order of the 12 protein-coding genes was *COX1*, *NAD3*, *NAD1*, *NAD4*, *COX3*, *NAD5*, *NAD6*, *NAD4L*, *NAD2*, *COB*, *COX2*, and *ATP6*. When compared to another species in Pholadoidea, *Xyloredo nooi*, an extensive rearrangement was detected. This indicates that the gene order within Pholadoidea is not conservative at the mitogenome level. In addition, four overlaps were observed at the boundaries between the genes *COX3* and *trnV* (*UAC*), *trnG* (*UCC*) and *COB*, *trnR* (*UCG*) and *rnr16*, as well as *trnS* (*UCU*) and *rnr12*. These overlaps are largely promoted by the compactness of the mitogenome [8]. The mitogenome has an overall base composition of 25.4 % A, 41.5 % T, 22.1% G, and 11 % C, demonstrating a bias towards AT bases (66.9 %). All published mitogenomes of Pholadoidea so far exhibited the same AT-skew [7]. The overall GC-content was 33.1 %, which is lower than most published mitogenome sequences of Pholadoidea, but higher when compared to species such as *Xyloredo* spp. [7].

Based on current sampling, the maximum likelihood (ML) trees reconstructed using sequences derived from the 12 protein-coding genes revealed that Pholadidae was first to diverge in Pholadoidea before Xylophagidae and Terebinthidae when using the first and third datasets (Figs. 2 and 3); Pholadidae was found to be sister to Xylophagidae when using the second

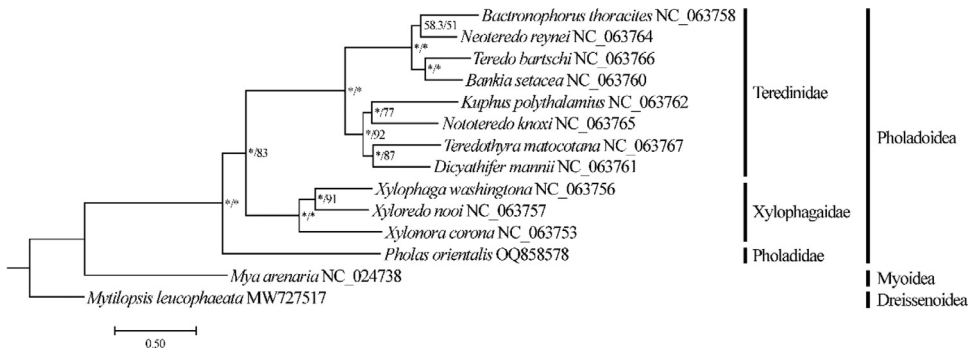


Fig. 2. The maximum likelihood phylogenetic tree based on nucleotide sequences of 12 concatenated mitochondrial protein-coding genes of *P. orientalis* and 11 other Pholadoidea species. The numbers at each node represent SH-aLRT support (%) / ultrafast bootstrap support, in which strong branch support (SH-aLRT ≥ 80 %; UFboot ≥ 95 %) is indicated with an asterisk (*).

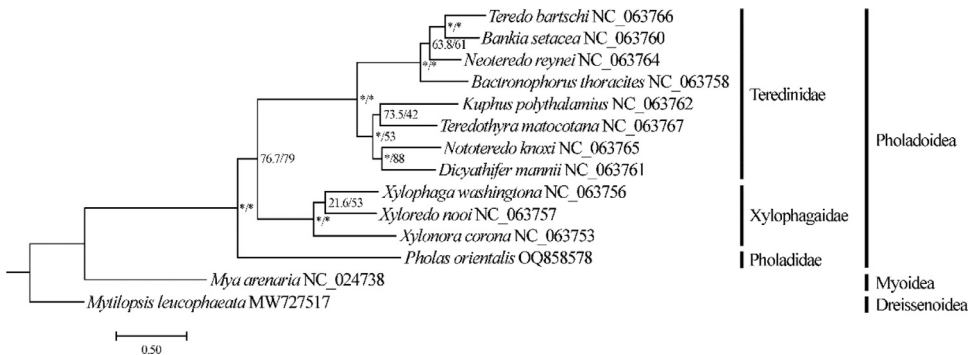


Fig. 3. The maximum likelihood phylogenetic tree based on nucleotide sequences (reverse-translated) of 12 concatenated mitochondrial protein-coding genes of *P. orientalis* and 11 other Pholadoidea species. The numbers at each node represent SH-aLRT support (%) / ultrafast bootstrap support, in which strong branch support (SH-aLRT ≥ 80 %; UFboot ≥ 95 %) is indicated with an asterisk (*).

dataset (Fig. 4). Among the three ML trees constructed using the three different datasets, a stronger backbone at the family level was identified in the ML tree using the first dataset when compared to the two other datasets; the divergence of Pholadidae was well-resolved and the branch node between Teredinidae and Xylophagaidae was reliable when referring to the SH-aLRT support value, but not with the ultrafast bootstrap value (i.e. 83 %). However, the ML trees based on the 12 taxa reaffirms the monophyly of Pholadoidea, which is similar to the previous study that is based on the nuclear 18S and 28S rDNA sequences [7]. However, it is noteworthy that the amino acid-based tree that is derived from the second dataset (Fig. 4) displayed distinct topologies, contrasting with the nucleotide-based trees derived from the first and third datasets (Figs. 2 and 3). Eventually, the findings from the nucleotide-based trees is congruent with the morphological classification and genetic analyses that support the sister relationship between Teredinidae and Xylophagaidae [6,7,9].

Although none of the ML trees were fully resolved in this study, based on the overall branch support, the nucleotide-based phylogenetic tree that is based on the first dataset exhibited the best confidence for the evolutionary relationships within Pholadoidea. Nevertheless, our work demonstrated the utility of mitogenome sequences in the phylogenetic analysis of Pholadoidea and provided a foundation for future phylogenetic and taxonomic research of these ecologically unique bivalves.

[16] and were trimmed using TrimAl v.1.3 [17] prior to concatenation. Based on the Bayesian information criterion, the most optimum substitution models suggested by ModelFinder [18] for the first and third datasets would be the transversion model (TVM) with empirical base frequencies (+F) coupled with invariable site (+I) plus discrete Gamma model (+G) (=TVM+F+I+G). For the second dataset, the recommended substitution model was the mitochondrial metazoa model (mtZOA)+F+I+G (=mtZOA+F+I+G). ML trees were reconstructed using IQ-TREE [19], in which the ultrafast bootstrap approximation approach (UFboot) [20] and Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) [21] were applied on each branch node with 1000 replicates.

Limitations

Not applicable.

Ethics Statement

The experiment complied with the ARRIVE guidelines and were carried out in accordance with the U.K. Animals (Scientific Procedures) Act, 1986 and associated guidelines; EU Directive 2010/63/EU for animal experiments; or the National Institutes of Health guide for the care and use of laboratory animals (NIH Publications No. 8023, revised 1978).

Data Availability

[Pholas orientalis mitochondrion, complete genome \(Original data\)](#) (NCBI GenBank).

CRediT Author Statement

Hao Xuan Ho: Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization; **Teek Foh Chong:** Resources, Data curation, Funding acquisition; **Wei Lun Ng:** Methodology, Validation, Investigation, Resources, Writing – review & editing, Supervision; **Shiou Yih Lee:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – review & editing, Supervision, Project administration.

Acknowledgments

The authors thank Dr. Chao Jiang of the National Resource Center for Chinese Materia Medica, Chinese Academy of Chinese Medical Science for the help provided in mitogenome assembly. This study was supported by the INTI IU Research Seeding Grant under Grant [number FHLS-01-14-2022].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J.D. Ronquillo, R.S. McKinley, Developmental stages and potential mariculture for coastal rehabilitation of endangered Pacific angelwing clam, *Pholas orientalis*, *Aquaculture* 256 (1–4) (2006) 180–191, doi:[10.1016/j.aquaculture.2006.01.030](https://doi.org/10.1016/j.aquaculture.2006.01.030).
- [2] M.S.M. Golez, L.V. Laureta, M.J.S. Apines-Amar, C.M. Caipang, Growth and survival of endangered angelwing clam, *Pholas orientalis* fed different algal diets, *Adv. Environ. Sci.* 3 (3) (2011) 237–242.
- [3] L.V. Laureta, L.A.G. Piñosa, S.M. Golez, Restoration of the oriental angelwing (*Pholas orientalis*) resources through broodstock transplantation in Western Visayas, Philippines, *Extreme Life Biospeol. Astrobiol.* 6 (2) (2014) 108–120.
- [4] H.M. Yusop, M.F.S. Ramli, The study of *Pholas orientalis*'s life cycle and its relationship with the ecological parameters in states of Kedah, Perak and Selangor, Malaysia, *J. Agric. Vet. Sci.* 10 (2017) 15–20, doi:[10.9790/2380-1006011520](https://doi.org/10.9790/2380-1006011520).
- [5] H.M. Manullang, B.H. Siswoyo, U. Hasan, R. Machrizal, The population dynamics of *Pholas orientalis* (Gmelin, 1791) in the area of Kelang Beach, Serdang Bedagai, Aquac. Aquar. Conser. Legis. 14 (1) (2021) 21–27 <http://bioflux.com.ro/docs/2021.21-27.pdf>.
- [6] D.L. Distel, M. Amin, A. Burgoyne, E. Linton, G. Mamangkey, W. Morrill, W. Nove, N. Wood, J. Yang, Molecular phylogeny of Pholadoidea Lamarck, 1809 supports a single origin for xylotrophy (wood feeding) and xylotrophic bacterial endosymbiosis in Bivalvia, *Mol. Phylogenet. Evol.* 61 (2) (2011) 245–254, doi:[10.1016/j.ympev.2011.05.019](https://doi.org/10.1016/j.ympev.2011.05.019).
- [7] Y. Li, M.A. Altamia, J.R. Shipway, M.R. Brugler, A.F. Bernardino, Z. Lin, S. Oliveira, P. Sumida, C.R. Smith, A. Trindade-Silva, K.M. Halanych, D.L. Distel, Contrasting modes of mitochondrial genome evolution in sister taxa of wood-eating marine bivalves (Teredinidae and Xylophagaidae), *Genome Biol. Evol.* 14 (6) (2022) evac089, doi:[10.1093/gbe/evac089](https://doi.org/10.1093/gbe/evac089).
- [8] F. Ghiselli, A. Gomes-Dos-Santos, C.M. Adema, M. Lopes-Lima, J. Sharbrough, J.L. Boore, Molluscan mitochondrial genomes break the rules, *Philos. Trans. R. Soc. B: Biol. Sci.* 376 (1825) (2021) 20200159, doi:[10.1098/rstb.2020.0159](https://doi.org/10.1098/rstb.2020.0159).
- [9] S. Monari, Phylogeny and biogeography of pholadid bivalve *Barnea* (*Anchomasa*) with considerations on the phylogeny of Pholadoidea, *Acta Palaeontol. Pol.* 54 (2) (2009) 315–335, doi:[10.4202/app.2008.0068](https://doi.org/10.4202/app.2008.0068).
- [10] J.M. Poutiers, Bivalve (Acephala, Lamellibranchia, Pelecypoda), in: K.E. Carpenter, V.H. Niem (Eds.), *FAO Species Identification Guide for Fishery Purposes The Living Marine Resources of the Western Central Pacific Volume 1 Seaweeds, corals, Bivalves and Gastropods*, Food and Agriculture Organization of the United Nations, Rome, 1998, pp. 354–355.
- [11] J.J. Jin, W.B. Yu, J.B. Yang, Y. Song, C.W. dePamphilis, T.S. Yi, D.Z. Li, GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes, *Genome Biol.* 21 (1) (2020) 241, doi:[10.1186/s13059-020-02154-5](https://doi.org/10.1186/s13059-020-02154-5).
- [12] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics* 28 (12) (2012) 1647–1649, doi:[10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199).
- [13] A. Donath, F. Jühling, M. Al-Arab, S.H. Bernhart, F. Reinhardt, P.F. Stadler, M. Middendorf, M. Bernt, Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes, *Nucleic Acids Res.* 47 (20) (2019) 10543–10552, doi:[10.1093/nar/gkz833](https://doi.org/10.1093/nar/gkz833).
- [14] M. Tillich, P. Lehwork, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, GeSeq - versatile and accurate annotation of organelle genomes, *Nucleic Acids Res.* 45 (2017) W6–W11 (W1), doi:[10.1093/nar/gkx391](https://doi.org/10.1093/nar/gkx391).
- [15] S. Greiner, P. Lehwork, R. Bock, OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes, *Nucleic Acids Res.* 47 (2019) W59–W64 (W1), doi:[10.1093/nar/gkz238](https://doi.org/10.1093/nar/gkz238).
- [16] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, *Bioinformatics* 23 (21) (2007) 2947–2948, doi:[10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404).
- [17] S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, *Bioinformatics* 25 (15) (2009) 1972–1973, doi:[10.1093/bioinformatics/btp348](https://doi.org/10.1093/bioinformatics/btp348).
- [18] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermini, ModelFinder: fast model selection for accurate phylogenetic estimates, *Nat. Methods* 14 (6) (2017) 587–589, doi:[10.1038/nmeth.4285](https://doi.org/10.1038/nmeth.4285).
- [19] L.T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, *Mol. Biol. Evol.* 32 (1) (2015) 268–274, doi:[10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300).
- [20] D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh, UFBBoot2: improving the ultrafast bootstrap approximation, *Mol. Biol. Evol.* 35 (2) (2018) 518–522, doi:[10.1093/molbev/msx281](https://doi.org/10.1093/molbev/msx281).
- [21] M. Anisimova, M. Gil, F. Dufayard, C. Dessimoz, O. Gascuel, Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes, *Syst. Biol.* 60 (5) (2011) 685–699, doi:[10.1093/sysbio/syr041](https://doi.org/10.1093/sysbio/syr041).