scientific reports

Check for updates

OPEN

Relative performance of three phylogenetic methods based on complete mitochondrial genomes of barnacle

Seongjun Bae

Mitochondrial Genome analysis is essential for understanding phylogenetic relationships. However, few studies have compared the performance of phylogenetic approaches for marine invertebrates, which have a complex evolutionary history. This study compared three phylogenetic trees based on 34 complete mitochondrial genomes, including *Amphibalanus eburneus*, *Fistulobalanus kondakovi*, and *Megabalanus rosa*, in terms of (1) gene order, (2) concatenated protein-coding genes, and (3) universal cytochrome c oxidase subunit I (*COX1*) marker regions. Each phylogenetic tree exhibited significant topological differences (Robinson–Foulds distance of 0.55–0.92). The protein-coding genes (78.8%) performed significantly better in terms of monophyletic preservation than the *COX1* marker region (61.3%) and gene order (50.0%). Gene order analysis identified two genomic regions (I and II) as hotspots (regions with concentrated rearrangement activity) with significantly elevated breakpoint densities (319 and 100 breakpoints, respectively; $p < 0.001$), indicating concentrated genome rearrangement activity. Although all three methods consistently preserved some families, they strongly suggested that taxonomic re-evaluation is necessary for Balanidae. In conclusion, gene order provides insights into genome evolution patterns, concatenated protein-coding genes are the most suitable for phylogenetic studies, and *COX1* markers are useful for rapid species identification rather than phylogenic classification. This comparative analysis provides important insights into the effects of method selection on mitochondrial phylogeny, especially when addressing complex phylogenetic problems in marine invertebrates.

Keywords Mitochondrial phylogenomics, Comparative phylogenetics, Gene order, Molecular systematics

Molecular systematics has greatly improved our understanding of evolutionary relationships, and mitochondrial genomes have become essential for resolving taxonomic issues in various lineages[1,2]. Complete mitochondrial genomes provide information on various aspects, including the nucleotide sequences of protein-coding genes (PCGs) and the physical arrangement of ribosomal RNA (rRNA) and transfer RNA (tRNA)[3].

The various types of mitochondrial data have given rise to different phylogenetic approaches, each with its own advantages and limitations. Gene order-based studies have utilized the rarity of genome rearrangements in various marine invertebrate taxa (e.g., echinoderms[4]mollusks[5]ascidians[6]and barnacles[7,8] to elucidate their relationships. Combining the analysis of PCGs can provide relatively high phylogenetic resolution when utilizing diverse taxa with diverse evolutionary rates[9–11]. Single-marker approaches use standardized regions, such as cytochrome c oxidase subunit I (*COX1*), enabling relatively rapid and accurate identification of molecular species based on extensive reference databases[12,13].

Although each method has its own advantages, few studies have systematically compared their performances using the same dataset for marine invertebrates. Some studies have compared phylogenetic trees based on the 13 PCGs and gene orders in Nemertea[14] or the 13 PCGs and *COX1* marker regions in ascidians[15]. Direct comparisons between methods are essential for determining the most effective method for mitochondrial phylogenomics and identifying the potential causes of phylogenetic inconsistencies. Marine arthropods are useful models for comparative methodological studies because of their long evolutionary histories and diverse mitochondrial genome structures[16–18].

Department of Ecology and Conservation, National Marine Biodiversity Institute of Korea, Seocheon 33662, Republic of Korea. email: silverto@naver.com

Barnacles (Cirripedia) have particularly diverse mitochondrial gene arrangement patterns[7,19] and are relatively well characterized in terms of complete mitochondrial genomes[20–34]. These characteristics make barnacles ideal for comparing mitochondrial genomic information from various perspectives. Various phylogenetic methods have been applied to barnacle mitochondrial genome data, yielding diverse results[8]. Gene order analysis has revealed unique rearrangement patterns across barnacle phylogenies[7]. Phylogenetic tree analysis based on the 13 PCGs revealed potential conflicts with the traditional classification system[35]. These results emphasize the need for a systematic comparison using consistent datasets and diverse approaches.

This study describes the complete mitochondrial genomes of three barnacle species (*Amphibalanus eburneus*, *Fistulobalanus kondakovi*, and *Megabalanus rosa*) and compares phylogenetic trees constructed from 34 complete mitochondrial genomes with regards to (1) gene order, (2) concatenated PCGs, and (3) universal *COX1* marker regions. To assess the relative reliability of each approach, their performance was quantitatively evaluated based on the Robinson–Foulds (RF) distance metric and monophyletic preservation rate. This comparative analysis provides insights into the methodological considerations that extend beyond barnacle phylogeny to broader issues in mitochondrial phylogenetics. The study highlights the relevance of mitochondrial genome data in evolutionary studies and emphasizes the importance of appropriate methods for addressing complex phylogenetic problems.

Materials and methods

Sample collection and complete mitochondrial genome

From February to March 2023, *A. eburneus*, *F. kondakovi*, and *M. rosa* were collected from the coast of the Republic of Korea. *Amphibalanus eburneus* was collected in Masan (35°12'00.05" N, 128°34'41.23" E), *F. kondakovi* in Buan (35°35'09.46" N, 126°36'22.07" E), and *M. rosa* in Pohang (35°52'50.82" N, 129°31'41.70" E; Table S1). Complete mitochondrial genome analysis was performed for all three species. Genomic DNA was extracted from the samples using a DNeasy Blood & Tissue DNA Kit (Qiagen, Hilden, Germany) and a genomic library was constructed using a QIAseq FX Single Cell DNA Library Kit (Qiagen) with paired-end reads. Next-generation sequencing (NGS) was performed using a NovaSeq X Series 10B Reagent Kit (300 cycles; Illumina, San Diego, CA, USA) on a NovaSeq 6000 system (Illumina). Sequencing yielded 48,694,468 paired-end raw reads for *A. eburneus*, 48,537,090 for *F. kondakovi*, and 45,801,380 for *M. rosa*. After quality control with Trim_Galore v0.6.1 (https://github.com/felixkrueger/TrimGalore) to remove adapter sequences and low-quality data, 48,194,603 clean reads were obtained for *A. eburneus*, 47,940,315 for *F. kondakovi*, and 44,767,988 for *M. rosa*.

Trimmed read data were used for mitochondrial genome assembly using de novo assembly combined with reference-based mapping. Initial assembly was performed using MitoZ v3.5[36] with the parameters "genetic_code 5" and "clade Arthropoda". *Amphibalanus eburneus*, *F. kondakovi*, and *M. rosa* were assembled using *A. amphitrite* (KF588709) as references, resulting in circular mitochondrial Genomes of 15,348 bp. After quality correction using Polypolish v0.5.0[37], no sequence errors were found in the assembly results, and the average read depth was 223.0× for *A. eburneus*, 246.1× for *F. kondakovi*, and 496.0× for *M. rosa*. The assembled complete mitochondrial Genomes of the three species contained 13 PCGs, 22 tRNAs, and 2 rRNAs. Circular maps were generated using CGView server[38].

Dataset compilation and phylogenetic analysis

Different phylogenetic methods were compared using a dataset of 34 complete mitochondrial Genomes. This included the three newly sequenced species in the present study and 31 previously published complete mitochondrial genomes obtained from the National Center for Biotechnology Information (NCBI) GenBank (accessed December 2024). The dataset included 31 species from the order Balanomorpha (ingroup) and three species from the order Pollicipedomorpha (outgroup; Table 1).

Three phylogenetic approaches were used in this study. For gene order-based phylogenetic analysis, Maximum Likelihood for Gene-Order[39] (MLGO) analysis was used to construct a maximum likelihood tree based on the arrangement and orientation of all mitochondrial genes, considering both gene position and strand orientation (+ or −) to capture the complete gene order information. Branch Support was assessed using 1,000 bootstrap replicates. Two datasets were used for the nucleotide sequence-based phylogenetic analysis. One consisted of the concatenated nucleotide sequences of the 13 PCGs, and the other consisted solely of the *COX1* marker region (LCO1490/HCO2198; 658 bp), as described by Folmer et al.[12]. For both datasets, sequence alignment was performed using CLUSTAL Omega[40] in Geneious Prime software (version 2025.0; Biomatters, Auckland, New Zealand). Phylogenetic trees were constructed using the maximum likelihood approach in raxmlGUI 2.0[41]. The best-fitting nucleotide Substitution model was determined using the GTR model for both datasets. Node Support was assessed using 1,000 bootstrap replicates for both phylogenetic analyses.

Comparative assessment of phylogenetic methods

The performance of the three phylogenetic approaches was evaluated using two complementary methods in R v4.0.2[42] (http://www.R-project.org/). To assess topological similarities between the phylogenetic trees, RF distances were calculated using "phangorn"[43] package. The RF distance quantifies the dissimilarity between two trees based on the number of bipartitions (evolutionary splits) present in one but absent in the other. Raw RF distances were calculated using the "RF.dist" function and normalized through division with the maximum possible RF distance ($2n$–6, where n is the number of taxa), resulting in values ranging from 0 (identical topologies) to 1 (maximally different topologies). This normalization allowed the direct comparison of tree similarities, independent of the number of taxa. The resulting distance matrix was visualized as a heatmap using "ggplot2"[44] package.

For the monophyletic preservation assessment, the "ape"[45] package was used to determine whether established taxonomic groups formed monophyletic clades in each tree. Phylogenetic trees were imported in

Group	Order	Family	Species	Accession no.	Total length (bp)	Reference
In group	Balanomorpha	Balanidae	*Acasta cyathus*	MN615272	15,251	Unpublished
			Acasta sulcata	KJ754818	15,326	Tsang et al., 2017
			Amphibalanus amphitrite	KF588709	15,683	Shen et al., 2015
			Amphibalanus eburneus	**PV231789**	**15,348**	**This study**
			Amphibalanus improvisus	MZ049958	15,336	Bae et al., 2021
			Amphibalanus reticulatus	OK662945	15,062	Unpublished
			Armatobalanus allium	KJ754817	15,063	Tsang et al., 2017
			Balanus balanus	KM660676	15,955	Unpublished
			Balanus trigonus	MW646099	15,560	Liu et al., 2021
			Fistulobalanus albicostatus	MK617531	15,665	Song et al., 2020
			Fistulobalanus kondakovi	**PV231790**	**16,447**	**This study**
			Megabalanus ajax	KF501046	15,510	Shen et al., 2016
			Megabalanus coccopoma	OK631889	15,098	Zhang et al., 2023
			Megabalanus rosa	**PV231791**	**15,133**	**This study**
			Megabalanus tintinnabulum	MN481499	15,107	Feng et al., 2019
			Perforatus perforatus	OR327030	15,536	Jeong et al., 2024
			Semibalanus balanoides	MG010647	15,119	Nunez et al., 2021
			Semibalanus cariosus	MT528637	15,118	Nunez et al., 2021
			Striatobalanus amaryllis	KF493890	15,064	Tsang et al., 2015
			Striatobalanus tenuis	OL979158	15,067	Mao et al., 2024
		Catophragmidae	*Catomerus polymerus*	MH791045	15,446	Chan et al., 2018
		Chelonibiidae	*Chelonibia testudinaria*	KJ754819	14,906	Tsang et al., 2017
		Chionelasmatidae	*Eochionelasmus ohtai*	MF939636	15,585	Kim et al., 2018
		Chthamalidae	*Chthamalus challengeri*	KY865097	15,358	Chen et al., 2019
			Chthamalus malayensis	MW076458	15,230	Mao et al., 2021
		Tetraclitidae	*Tesseropora rosea*	KY865099	15,330	Cai et al., 2018
			Tetraclita japonica	MH119183	15,192	Ge et al., 2019
			Tetraclita rufotincta	KY865100	15,002	Song et al., 2017
			Tetraclita serrata	KJ434948	15,200	Shen et al., 2015
			Tetraclita squamosa squamosa	MT232759	15,191	Feng et al., 2020
			Tetraclitella divisa	KJ754822	14,973	Tsang et al., 2017
Out group	Pollicipedomorpha	Pollicipedidae	*Pollicipes polymerus*	AY456188	15,634	Lavrov et al., 2004
			Pollicipes pollicipes	OQ525896	15,148	Unpublished
			Capitulum mitella	AY514042	14,915	Lim and Hwang, 2006

Table 1. Details of complete mitochondrial genomes. Species analyzed in this study are highlighted in bold.

Newick's format, and taxon labels were standardized to ensure consistent comparisons. For each taxonomic group (18 Genera and 9 families), monophyly was assessed using the "is.monophyletic" function, which tests whether all group members share an exclusive common ancestor. The percentage of preserved monophyletic taxonomic groups was calculated for each method to quantify their relative performance in preserving the established taxonomic relationships and elucidate how well each aligns with current classification systems.

Gene order breakpoints were quantified based on the concept of Blanchette et al.[46] using pairwise comparisons of gene adjacency across clades. Breakpoint densities were calculated for predefined genomic regions, and enrichment ratios were obtained by comparing densities with entire genome expectations. Statistical significance was evaluated via permutation testing (10,000 replicates), with effect sizes estimated using Cohen's d and deviations from uniformity tested via chi-square statistics. Breakpoint analysis was also performed using the "ape" package.

Results

Complete mitochondrial genomes

The three mitochondrial genomes showed similar nucleotide compositions. *Amphibalanus eburneus* contained 37.2% A, 15.9% C, 11.4% G, and 35.5% T. *Fistulobalanus kondakovi* had 38.2% A, 17.2% C, 10.8% G, and 33.8% T. *Megabalanus rosa* had 31.6% A, 17.5% C, 13.8% G, and 37.0% T. All three species had AT contents of 72.7%, 72.0%, and 68.6%, respectively (Fig. 1).

The start and stop codon patterns differed among the species. ATG was the most common start codon (eight in *A. eburneus*, seven in *F. kondakovi*, and eight in *M. rosa*), followed by ATT (two in *A. eburneus*, two in *F. kondakovi*, and three in *M. rosa*) and GTG codons (two in *A. eburneus*, two in *F. kondakovi*, and one in *M. rosa*). Some start codons occurred rarely, such as ATA and ATC in *F. kondakovi* and TTG in *M. rosa*. Notably, the ACA

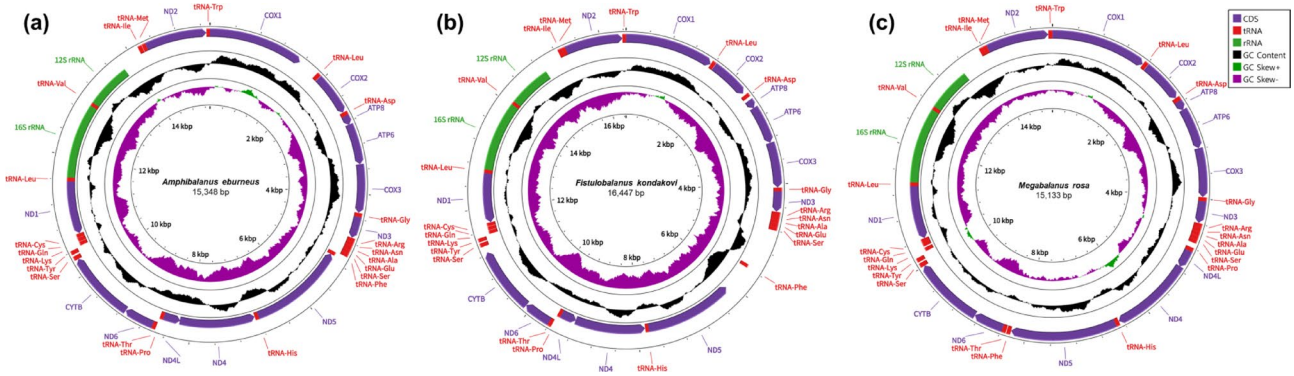

Fig. 1. Circular maps of complete mitochondrial genomes. (**a**) *Amphibalanus eburneus* (15,348 bp), (**b**) *Fistulobalanus kondakovi* (16,447 bp), and (**c**) *Megabalanus rosa* (15,133 bp). Protein-coding genes, rRNAs, and tRNAs are shown in purple, red, and green, respectively, with the inner circle displaying GC content.

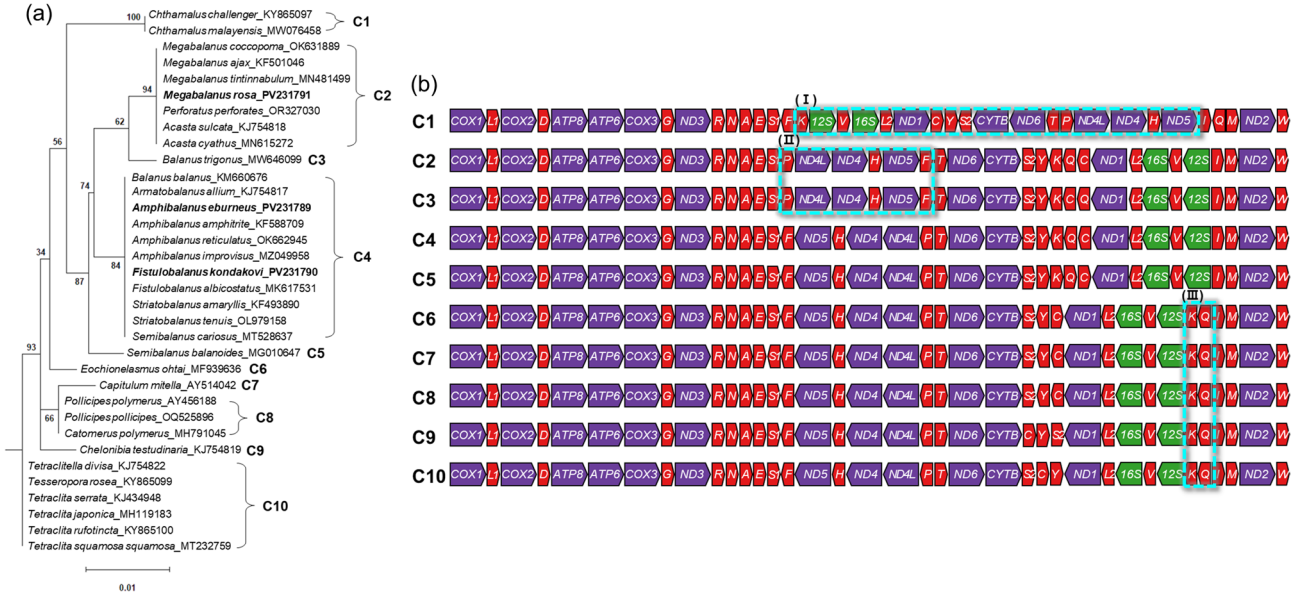

Fig. 2. Gene order-based phylogenetic analysis. (**a**) Phylogenetic tree revealing 10 distinct clades (C1–C10). (**b**) Comparative gene order patterns across clades with three major rearrangement regions (I, II, and III; highlighted with cyan dotted boxes). Protein-coding genes, rRNAs, and tRNAs are shown in purple, red, and green, respectively. Gene arrows indicate transcriptional orientation. Species analyzed in this study are highlighted in bold.

of *A. eburneus COX1* is a nonstandard start codon rarely observed in the mitochondrial genome. TAA was the most common stop codon (eight in *A. eburneus*, ten in *F. kondakovi*, and eight in *M. rosa*); only *M. rosa* exhibited a TAG stop codon. The incomplete stop codon "T—" was present in all three species (five in *A. eburneus*, three in *F. kondakovi*, and four in *M. rosa*). The transcriptional orientation of the PCGs varied among the species. In *A. eburneus* and *F. kondakovi*, most Genes were in the 5′–3′ (+) orientation, while *ND1*, *ND4*, *ND4L*, and *ND5* were in the 3′–5′ (−) orientation. In *M. rosa*, all PCGs except *ND1* were in the 5′–3′ direction (Table S2).

Gene order-based phylogenetic analysis

The gene order-based phylogenetic tree revealed an unexpected pattern in which the outgroup taxa (Pollicipedomorpha) did not form a distinct clade from the ingroup taxa (Balanomorpha; Fig. 2). The tree exhibited 10 distinct clades that were only partially aligned with conventional taxonomic groupings. The three species of outgroup were included in clades seven and eight. *Pollicipes polymerus* and *Pollicipes pollicipes* formed C8 together with *Catomerus polymerus*, and *Capitulum mitella* formed C7. These results contradict the traditional taxonomic classification in which Balanomorpha and Pollicipedomorpha are considered distinct orders.

In the case of Balanomorpha, which was set as an ingroup, the two Chthamalidae species formed independent clusters. The 25 species of the Balanidae family were divided into four clades. The seven species of the genera *Megabalanus*, *Perforatus*, and *Acasta* were grouped into C2, and 11 species of the genera *Balanus*, *Armatobalanus*,

Amphibalanus, *Fistulobalanus*, *Striatobalanus*, and *Semibalanus* were grouped into C4. Notably, *Balanus trigonus* and *Semibalanus balanoides* were separated into the distinct clades C3 and C5, respectively. *Eochionelasmus ohtai* (Chionelasmatidae) and *Chelonibia testudinaria* (Chelonibiidae) diverged at C6 and C9, respectively, and three species of the genera *Pollicipes* and *Catomerus* were grouped into C8. All six species included in Tetraclitidae were grouped as C10 (Fig. 2a).

The arrangement from C1 to C10 was inconsistent with some major differences. Noticeable changes in the gene order tree were observed in regions I, II, and III (cyan dotted lines in Fig. 2b). Regions I and II showed exceptionally high breakpoint concentrations of 319 and 100, respectively, higher than the entire genome (87), confirming their status as hotspots (regions with concentrated rearrangement activity). Region III exhibited minimal breakpoint activity, with only two breakpoints, but demonstrated broad distribution across five clades from C6 to C10. Statistical analysis confirmed significant rearrangement patterns in all regions ($p < 0.001$, Table S3).

In region I, C1 showed significant variation across tRNA-CDS-rRNA. Compared to C6–C10, the overall order of trnK, rrnS, trnV, rrnL, trnL2, *ND1*, trnC, trnY, trnS2, *CYTB*, *ND6*, trnT, trnP, *ND4L*, *ND4*, trnH, and *ND5* was inverted. In region II, the order of C2 and C3 was also reversed compared to that of C4–C10, with trnP, *ND4L*, *ND4n*, trnH, *ND5*, and trnF appearing in the reverse order. C6–C9 in region III differed from C2–C5 in positions trnK and trnQ. In C2–C5, they were located between trnY and trnC, whereas in C6–C9 they were located between rrnS and trnI. Excluding regions I, II, and III, trnC and trnQ had a different order in C3 compared to C2 and C4–C5, and C4 and C5 had opposite orientations of trnQ and trnC. For C6–C10, the orientation of trnS1 and the orientations and positions of trnS2, trnY, and trnC differed (Fig. 2b).

Phylogenetic analysis based on 13 protein-coding genes

The phylogenetic tree constructed using the concatenated sequences of the 13 PCGs showed a different topology than the gene order-based tree. The outgroup (Pollicipedomorpha) formed a monophyletic group that was clearly separated (>95% bootstrap support) from the ingroup (Balanomorpha), consistent with the traditional taxonomic classification. Within Balanomorpha, most families except Balanidae were consistent with a previously established phylogenetic classification system (Fig. 3).

Balanidae does not form a monophyletic group due to the nested position of Chthamalidae, forming two distinct clades. The first clade included species from the genera *Semibalanus*, *Balanus*, *Striatobalanus*, *Fistulobalanus*, *Amphibalanus*, and *Armatobalanus*. The second clade comprised species from the genera

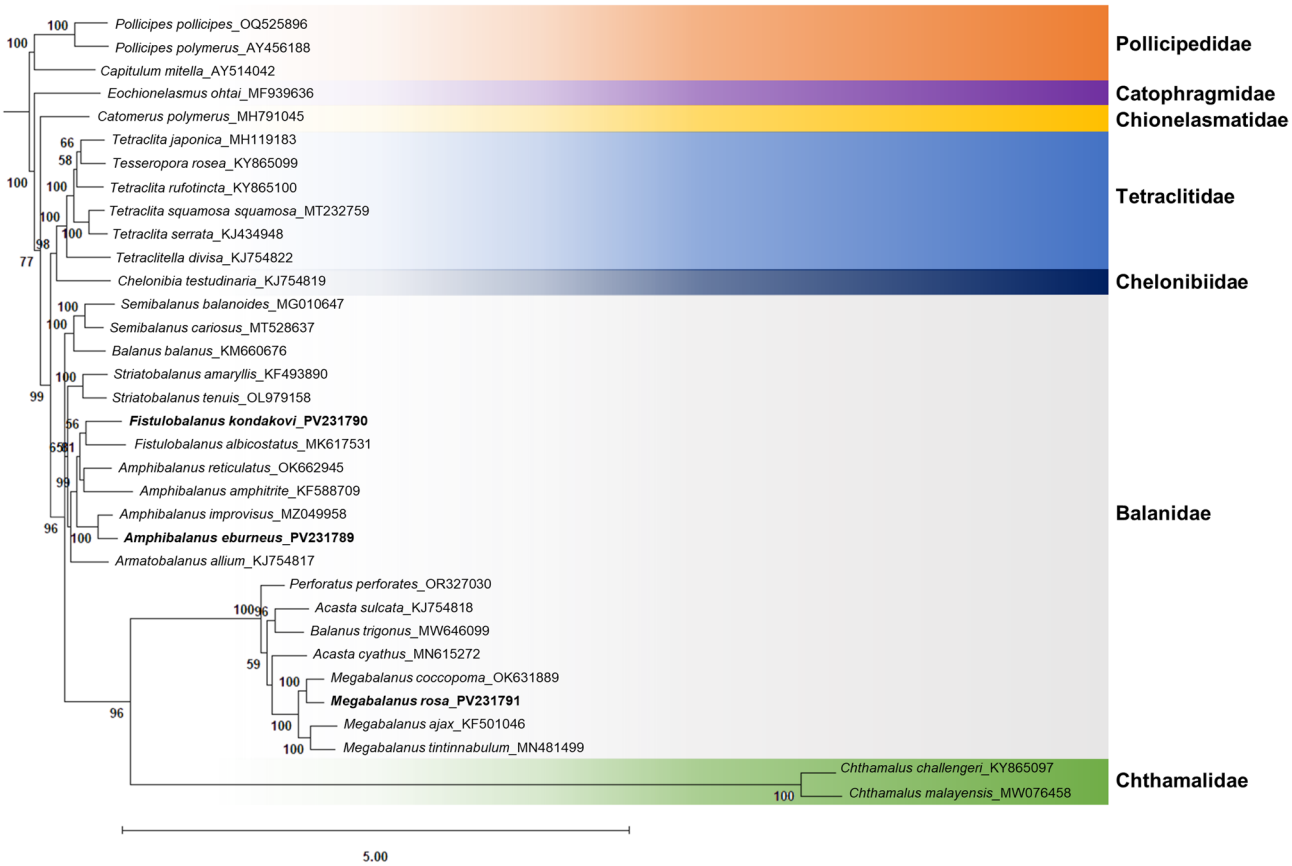

Fig. 3. Maximum likelihood phylogenetic tree based on 13 protein-coding genes. Bootstrap values ≥ 50% are shown at nodes. Species analyzed in this study are highlighted in bold. Scale bar represents substitutions per site.

Perforatus, *Acasta*, *Balanus*, and *Megabalanus*, with *Chthamalus* nested within this clade. The three species whose mitochondrial genomes are newly reported in this study, formed identical clades within their respective genera. Notably, despite belonging to the same genus *Balanus*, *B. balanus* and *B. trigonus* were located in the first and second clades, respectively. The families Catophragmidae, Chionelasmatidae, Tetraclitidae, Chelonibiidae, and Chthamalidae each exhibited robust monophyly and were distinctly divergent from both Balanidae clades. *Chthamalus challengeri* and *Chthamalus malayensis* (Chthamalidae) formed a clade (100% bootstrap support) nested within Balanidae (Fig. 3).

Phylogenetic analysis based on COX1 marker region

The phylogenetic tree based on the *COX1* universal primers (658 bp, LCO/HCO) differed from the 13 PCG phylogenetic tree in that the ingroups and outgroups were not clearly separated, and *E. ohtai* (Balanidae) formed a clade with the outgroup (*C. mitella*). The ingroups did not form a single clade in some families, and some mixing occurred between taxonomic groups. Balanidae was dispersed across two clades, and Chthamalidae was also fragmented. Within Balanidae, only *Megabalanus* formed a relatively consistent group (72–100% bootstrap support), including *M. ajax*, *M. tintinnabulum*, *M. rosa*, and *M. coccopoma*. The target species of this study, *A. eburneus* and *M. rosa*, formed clades with *A. improvisus* and *M. coccopoma*, respectively, while *F. kondakovi* formed an independent clade (Fig. 4).

Chthamalidae did not form a single clade, with *C. challengeri* positioned at the base of the phylogenetic tree near *C. polymerus* (Catophragmidae), whereas *C. malayensis* formed a group with *Acasta* (Balanidae). Tetraclitidae formed a relatively consistent group (83–91% bootstrap support) in the lower part of the tree, but *Tetraclitella divisa* was positioned outside the main clade and clustered with *Chelonibia testudinaria* (Chelonibiidae). Overall, the *COX1* marker region-based phylogenetic tree had lower resolution, and there were numerous cases in which multiple taxonomic groups belonging to the same family did not form a monophyletic group (Fig. 4).

Comparative assessment of phylogenetic methods

The three phylogenetic methods showed significant differences in terms of RF distances and monophyletic preservation. The RF distance matrix revealed significant topological differences among the approaches. The gene order tree showed the highest difference compared to the other methods (0.82 and 0.92 versus the 13 PCG and *COX1* marker region tree, respectively). In contrast, the 13 PCG and *COX1* marker region trees showed relatively close topological similarities (RF distance of 0.55; Fig. 5a). The 13 PCG tree performed well in terms of monophyletic preservation, preserving 78.8% of the expected single lineages based on taxonomic classification. The *COX1* marker region tree showed moderate performance, preserving 61.3% of the expected monophyletic groups. The Gene order trees showed the lowest performance, preserving only 50.0% of the expected monophyletic groups (Fig. 5b). The *COX1* marker region and gene order tree had large RF distances (0.92), but were relatively similar in terms of monophyletic preservation (50.0% and 61.3%, respectively). Conversely, the 13 PCG and *COX1* marker region trees had low RF distances (0.55), but differed relatively widely in terms of monophyletic preservation (78.8% and 61.3%, respectively).

Discussion

Complete mitochondrial genome characteristics

The complete mitochondrial genome characteristics of *(A) eburneus*, *F. kondakovi*, and *M. rosa* were largely consistent with those of other barnacle species. The full genome size (15,348, 16,447, and 15,133 bp, respectively) was within the minimum–maximum range (14,906–17,374 bp) for previously reported barnacle genomes[7]. All three species showed the typical 37 genes (13 PCGs, 22 tRNAs, and 2 rRNAs) found in barnacles, with a nucleic acid composition of 68.6–72.7% AT. These results are consistent with the patterns observed in other barnacle studies, such as *M. tintinnabulum* (67.3%)[26]*P. perforatus* (71.1%)[23]and *(B) improvisus* (72.8%)[30].

The start and stop codons of the PCG had unique features, especially the start codon of *M. rosa*, which differed from that in previous studies. Most start codons were the standard ATG, whereas *ATP8* and *ND4L* had the alternative invertebrate start codons ATT and GTG (NCBI Gene code 5)[47,48]consistent with other barnacle studies on five species of *Tetraclita*[25,32–34,49] and three species of *Acasta*[7]. The exceptional alternative start codons ACA and TTG occurred in the *COX1* genes of *A. eburneus* and *M. rosa*, respectively. The ACA codon may occur in some *COX1* genes of the Cirripedia subclass, as evidenced from its detection in *Capitulum mitella*[50].

Although methodological verification (depth of the *COX1* start codon: 414–423×; Fig. S1c) was performed during assembly, this is the first mitochondrial genome study to identify TTG as the start codon in barnacles, which has otherwise been reported in marine invertebrates such as nematodes and roundworms[51]; thus, further research is needed. Most stop codons were TAA and the incomplete "T—". The exclusive use of TAG in *CYTB* of *M. rosa* is consistent with previous studies on numerous species including four *Megabalanus* species[8,26,29]. Therefore, TAG is considered an important termination codon in the mitochondrial *CYTB* of barnacles.

Patterns and limitations of gene order phylogenetics

Gene order-based phylogenetic analysis revealed unexpected patterns that sharply contrasted with those of the traditional classification system. The outgroup (Pollicipedomorpha) and ingroup (Balanomorpha) did not form a distinct clade; instead, Tetraclitidae formed an independent clade. These results highlight the limitations of gene order-based phylogenetics. Despite frequent gene order in barnacle mitochondrial genomes[8,52]the phylogenetic tree topology showed the lowest monophyletic preservation performance (50.0%). This suggests that mitochondrial gene order alone may not accurately reflect the evolutionary relationships of barnacles.

The discrepancy between the phylogenetic tree based on gene order and traditional taxonomy indicates complex evolutionary patterns in mitochondrial gene order. This has been confirmed in barnacle species

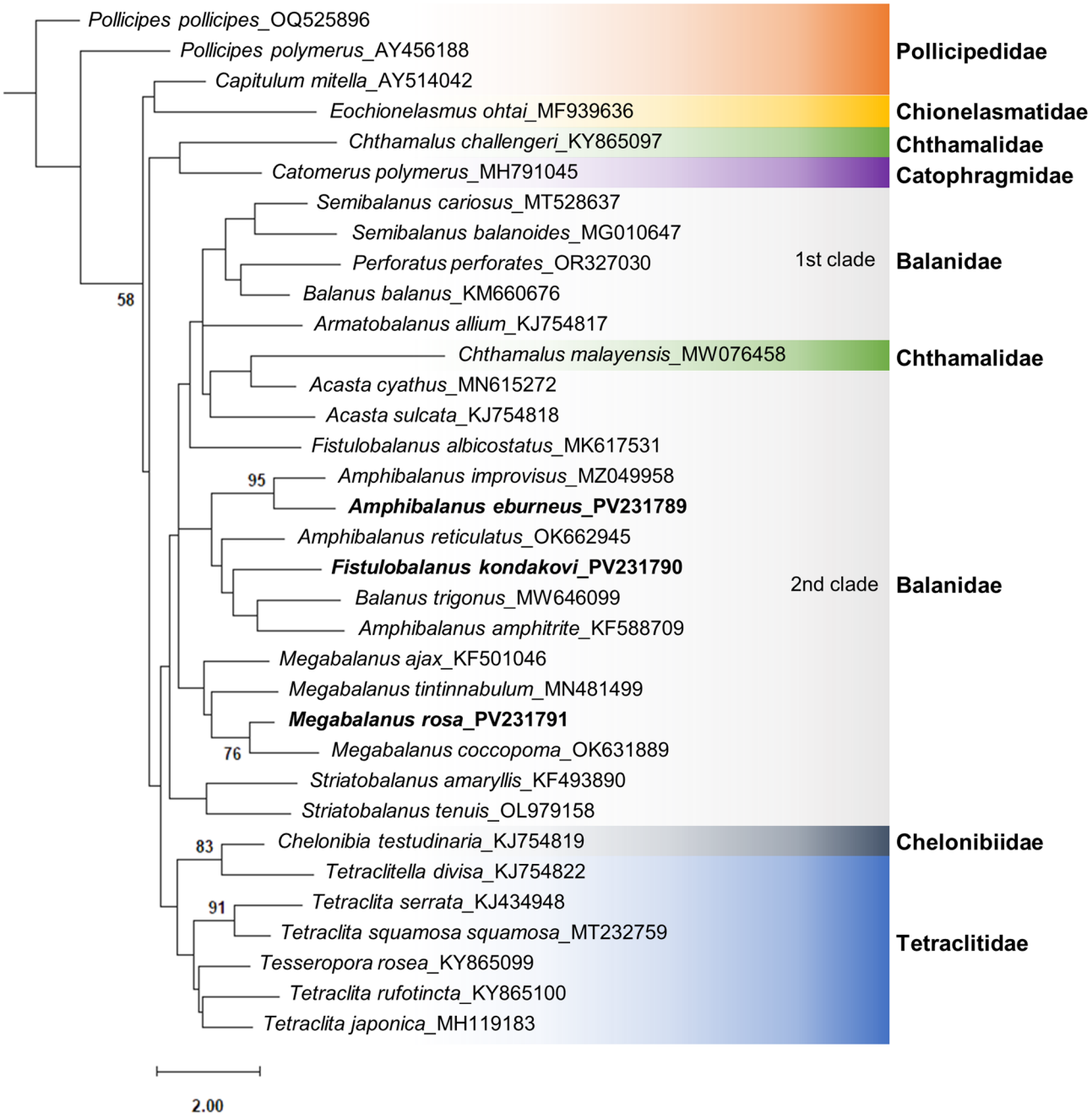

Fig. 4. Maximum likelihood phylogenetic tree based on the universal COI marker region (658 bp). Bootstrap values ≥ 50% are shown at nodes. Species analyzed in this study are highlighted in bold. The scale bar represents substitutions per site.

belonging to different families that are phylogenetically distant but share the same gene order[7] and in crustaceans, where similar gene translocations occur independently in unrelated lineages[19]. The Gene order patterns across 10 different barnacle clades provide valuable insights into mitochondrial genome evolution, despite the limitations of phylogenetic reconstructions based solely on gene order.

In regions I and II, recombination-induced inversions occurred from trnK to *ND5* and from trnP to trnF, respectively, with approximately 58% (8,863–8,885) of the total sequence length (15,230–15,358 bp) inverted in region I. This large-scale gene inversion (> 50%) likely resulted from mitochondrial recombination mechanisms involving double-strand breaks, similar to the patterns observed by Uda et al.[53] in Octocorallia and Sun et al.[4] in Holothuroidea. Additionally, the breakpoints in regions I and II (319 and 100, respectively) are highly elevated compared to the entire genome (87), confirming them as hotspots with a high likelihood of genomic restructuring. These hotspots may arise from tandem duplication events followed by random gene loss, where structural instabilities in regions with high tRNA density promote recombination mediated rearrangements[8,54].

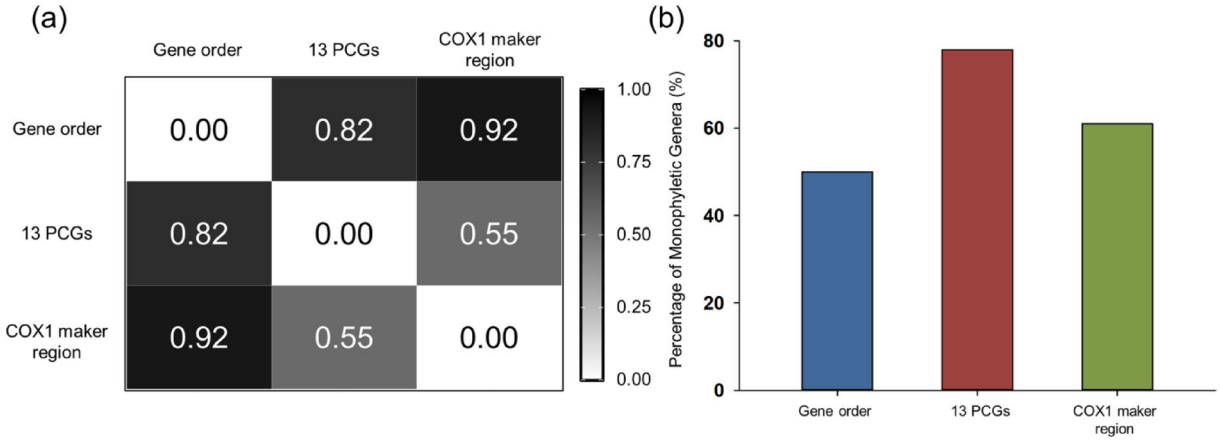

Fig. 5. Comparative analysis of phylogenetic approaches. (**A**) Robinson–Foulds distance heatmap showing topological differences between phylogenetic methods. (**B**) Monophyletic preservation rates across methods (Gene order: 50.0%, 13 PCGs: 78.8%, COI marker region: 61.3%).

In region III, the trnK and trnQ blocks were located in the intergenic space between rrnS and trnI, whereas in C4–C5, they were located between trnY and trnC. These distant differences in tRNA positions are likely related to targeted tRNA remodeling events, rather than simple transpositions. This phenomenon, in which tRNAs change their identity through mutations in the anticodon region while maintaining their secondary structure[55]has been reported in the mitochondrial genomes of amphipods[56] and sessile marine gastropods[57]. Although Region III does not meet traditional hotspot criteria based on breakpoint density, its widespread occurrence across multiple clades (C6–C10) indicates that this region represents an evolutionarily significant region for genome reorganization events.

Evolutionary patterns from genomic rearrangement regions

These hotspots provide insights into mitochondrial genome evolution patterns across barnacle lineages. Previous studies have demonstrated a correlation between rearrangement rates and nucleotide sequence evolution rates in animal mitochondrial genomes[58]. Significant mitochondrial gene rearrangements have also been observed in other marine invertebrates (annelids), showing an association with differential evolutionary patterns[59]. In the case of bivalves, differences in ecological habitats were associated with mitochondrial gene arrangement, supporting the hotspots are key points for observing differential evolutionary patterns[60].

Although region III does not meet traditional hotspot criteria based on breakpoint density, its consistent occurrence across phylogenetically diverse clades (C6–C10) suggests a different evolutionary significance. Unlike regions I and II, which show extensive rearrangement activity, region III appears to represent a shared genomic feature maintained across taxonomically distinct barnacle families. Previous studies have demonstrated that barnacle species from different families can share identical gene rearrangements[7]and such shared genomic features can provide valuable phylogenetic information within Cirripedia[8]. This suggests that region III may represent a rare genomic event that reflects shared evolutionary relationships rather than an active rearrangement zone.

Despite the limitations in gene order phylogenetics observed in this study, the identification of these hotspot regions demonstrates that gene order analysis can reveal meaningful evolutionary patterns. In particular, gene order data showed significant phylogenetic signals within Tetraclitidae (C10) and certain genera, such as *Megabalanus*, suggesting that gene order data can provide useful phylogenetic information for closely related species where rearrangements are unlikely[7]. These limitations highlight the need for caution when applying gene order analysis to deep phylogenetic relationships in barnacles, where rearrangement patterns may obscure true evolutionary relationships.

Comparative analysis of PCG and COX1 marker region phylogenetic approaches

The PCG-based analysis performed best in distinguishing between the phylogenetic relationships of barnacles: Monophyletic preservation was greater using the 13 PCGs (78.8%) than the gene order (50.0%) and *COX1* marker region (61.3%). The tree based on 13 PCGs preserved each family, except Balanidae, as a single clade, whereas the *COX1* marker region showed inconsistent performance among the families. The consistent results obtained using these methods, which did not clearly distinguish between Balanidae and Chthamalidae, suggested that this was an evolutionary pattern rather than a methodological error. This is in contrast with previous studies supporting the monophyly of Balanidae[32,35].

The good performance of the 13 PCGs approach is attributed to the following advantages. The relatively diverse taxa effectively capture patterns in gene loci with varying evolutionary rates[61] and mitigate potential biases affecting individual genes[62]. Strong conservation was observed at all the classification levels, from deep divergences to recent relationships. In contrast, the poor performance of the *COX1* marker region is attributed to its relatively short nucleotide sequence length (658 bp), functional constraints that increase the likelihood of

convergent evolution[63]and rapid evolutionary rates that make it more prone to reaching replacement saturation at deeper divergence points[64].

Comparative assessment of phylogenetic methods

This study provides new insights that complement existing reports on barnacle phylogeny. Tsang et al.[7] examined barnacle mitochondrial genomes using gene order and sequence analysis and discovered that, although both approaches were useful, they may yield different phylogenetic results. My study builds on their work by empirically confirming the performances of these methods. The results of this study are consistent with several key taxonomic findings. Notably, Balanidae did not form a monophyletic group across all three methods, highlighting the need for taxonomic revision. The discrepancy between traditional classification and molecular evidence is particularly intriguing, given that Balanidae have primarily been defined based on shell morphology and plate structure[65]. The allocation of *B. trigonus* and *B. balanus* to different clades in all phylogenetic trees was consistent with previous studies[28,35]suggesting a taxonomic inconsistency at the genus level that necessitates further review.

This study highlights the similarities and differences between the phylogenetic approaches in barnacle classification. In particular, all approaches consistently maintain Tetraclitidae as a single monophyletic group, supporting morphological taxonomic classification. In contrast, Chthamalidae showed conflicting phylogenetic placements, contributing to the non-monophyletic status of Balanidae[66]. These complementary approaches are likely to enrich our understanding of barnacle evolution through multiple lines of evidence.

Overall, the comparative results of the phylogenetic trees emphasize the importance of method selection in barnacle phylogenetics. Although gene order analysis has limitations in inferring the overall phylogeny, it revealed interesting patterns of mitochondrial genome evolution. In particular, the differential distribution of gene rearrangements in regions I, II, and III may provide homomorphic phylogenetic features for specific barnacle groups. Although the complete mitochondrial genome approach provides excellent resolution and monophyletic preservation for comprehensive evolutionary studies, the *COX1* marker region is considered more valuable for rapid species identification than for phylogenetic studies because of its extensive reference database[67,68]. Finally, the findings highlight the relevance of comparative analyses using multifaceted molecular data for developing classification systems that accurately reflect evolutionary history.

Conclusion

In the mitochondrial phylogenetics of barnacles, the use of 13 PCGs is demonstrated higher monophyletic preservation rates than the gene order or *COX1* marker region. A comprehensive comparison of the three phylogenetic approaches using complete mitochondrial Genomes of three newly and 39 previously sequenced species revealed that the 13 PCGs-based method demonstrated higher monophyletic preservation ability compared with *COX1* marker region and gene order analysis. Gene rearrangements are theoretically useful for phylogenetics; however, owing to the apparent homogeneity of mitochondrial structures, they perform surprisingly poorly in restoring established relationships. Nevertheless, the identified hotspots provide insights into genomic evolutionary mechanisms and highlight the potential influence of molecular and potential ecological factors on rearrangement patterns. The consistent non-monophyly of Balanidae across all three methods provides evidence that this family requires a taxonomic reevaluation, though these findings are based on mitochondrial data alone and may not be generalizable to other marine invertebrate lineages. Future studies should focus on incorporating nuclear genome data and morphological characteristics to resolve these taxonomic inconsistencies.

Data availability

The datasets used and analyzed during the current study available from the corresponding author on reasonable request. The complete mitochondrial genome supporting the results of this study is available in the NCBI GenBank database under accession number PV231789 (Amphibalanus eburneus), PV231790 (Fistulobalanus kondakovi), and PV231791 (Megabalanus rosa **).**.

Received: 12 May 2025; Accepted: 2 September 2025
Published online: 06 October 2025

References

1. Avise, J. C. et al. Intraspecific phylogeography: the mitochondrial DNA Bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**, 489–522 (1987).
2. Gissi, C., Iannelli, F. & Pesole, G. Evolution of the mitochondrial genome of metazoa as exemplified by comparison of congeneric species. *Heredity* **101**, 301–320 (2008).
3. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
4. Sun, S. et al. Highly rearranged gene orders and rapid evolutionary rates in the mitochondrial genomes of apodida (Echinodermata: Holothuroidea). *Front. Mar. Sci.* **10**, 1219437 (2023).
5. Xu, T., Qi, L., Kong, L. & Li, Q. Mitogenomics reveals phylogenetic relationships of patellogastropoda (Mollusca, Gastropoda) and dynamic gene rearrangements. *Zoöl Scr.* **51**, 147–160 (2022).
6. Gissi, C. et al. Hypervariability of Ascidian mitochondrial gene order: exposing the myth of deuterostome organelle genome stability. *Mol. Biol. Evol.* **27**, 211–215 (2010).
7. Tsang, L. M., Shen, X., Cheang, C. C., Chu, K. H. & Chan, B. K. K. Gene rearrangement and sequence analysis of mitogenomes suggest polyphyly of archaeobalanid and Balanid barnacles (Cirripedia: Balanomorpha). *Zool. Scr.* **46**, 729–739 (2017).
8. Zhang, M., Cai, Y., Ji, N., Chan, B. K. K. & Shen, X. The mitochondrial genome of the globally invasive barnacle *Megabalanus coccopoma* Darwin 1854 (Crustacea: Balanomorpha): Rearrangement and Phylogenetic Consideration within Balanomorpha. *Diversity* **15**, 117 (2023).

9. Alfaro, R., Mantilla, G. & Marín, A. A complete mitochondrial genome-based phylogeny of gooseneck barnacles confirms the sister group relationship between the Eastern Pacific *Pollicipes elegans* and the Eastern Atlantic *P. pollicipes*. *Gene Rep.* **40**, 102238 (2025).
10. Feng, J. et al. Characterization of four mitochondrial genomes of family neritidae (Gastropoda: Neritimorpha) and insight into its phylogenetic relationships. *Sci. Rep.* **11**, 11748 (2021).
11. Williams, S. T., Foster, P. G. & Littlewood, D. T. J. The complete mitochondrial genome of a turbinid vetigastropod from miseq illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. *Gene* **533**, 38–47 (2014).
12. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotech.* **3**, 294–299 (1994).
13. Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. Biological identifications through DNA barcodes. *Proc. Royal Soc. Lond. Ser. B Biol. Sci.* **270**, 313–321 (2003).
14. Podsiadlowski, L., Braband, A., Struck, T. H., von Döhren, J. & Bartolomaeus, T. Phylogeny and mitochondrial gene order variation in lophotrochozoa in the light of new mitogenomic data from nemertea. *BMC Genom.* **10**, 364 (2009).
15. Griggio, F. et al. Ascidian mitogenomics: comparison of evolutionary rates in closely related taxa provides evidence of ongoing speciation events. *Genome Biol. Evol.* **6**, 591–605 (2014).
16. Stokkan, M. et al. Species delimitation and mitogenome phylogenetics in the subterranean genus *Pseudoniphargus* (Crustacea: Amphipoda). *Mol. Phylogenetics Evol.* **127**, 988–999 (2018).
17. Romanova, E. V. et al. Evolution of mitochondrial genomes in Baikalian amphipods. *BMC Genom.* **17**, 1016 (2016).
18. Bukin, Y. S., Petunina, J. V. & Sherbakov, D. Yu. The mechanisms for genetic diversity of Baikal endemic amphipod Gmelinoides fasciatus: relationships between the population processes and paleoclimatic history of the lake. *Russ J. Genet.* **54**, 1059–1068 (2018).
19. Kilpert, F., Held, C. & Podsiadlowski, L. Multiple rearrangements in mitochondrial genomes of isopoda and phylogenetic implications. *Mol. Phylogenetics Evol.* **64**, 106–117 (2012).
20. Mao, N. et al. Complete mitochondrial genome of *Striatobalanus tenuis* hoek, 1883 (Balanomorpha: Balanidae) and a novel molecular phylogeny within cirripedia. *Mitochondrial DNA Part. B.* **9**, 29–32 (2024).
21. Tsang, L. M., Shen, X., Chu, K. H. & Chan, B. K. K. Complete mitochondrial genome of the acorn barnacle *Striatobalanus amaryllis* (Crustacea: Maxillopoda): the first representative from Archaeobalanidae. *Mitochondrial DNA*. **26**, 761–762 (2015).
22. Chan, B. K. K., Aguilar, L., Hou, B. K., Kang, H. M. & Kim, S. J. Complete mitochondrial genome of the catophragmid barnacle *Catomerus polymerus* (Cirripedia, thoracica, balanomorpha, Catophragmidae). *Mitochondrial DNA Part. B.* **3**, 1286–1287 (2018).
23. Jeong, J. S., Kim, H. K., Park, J. S., Hwang, H. S. & Kim, I. Complete mitochondrial genome of the European common barnacle *Perforatus perforatus* bruguière, 1789 (balanomorpha: balanidae). *Mitochondrial DNA Part. B.* **9**, 823–827 (2024).
24. Kim, S. J., Lee, W. K., Kim, R. O. & Ju, S. J. Complete mitochondrial genome of the hydrothermal vent barnacle *Eochionelasmus Ohtai* (Cirripedia, Thoracica). *Mitochondrial DNA Part. B.* **3**, 46–47 (2018).
25. Ge, T. et al. The first mitochondrial genome of *Tetraclita japonica* (Crustacea: Sessilia) from china: phylogeny within cirripedia based on mitochondrial genes. *Mitochondrial DNA Part. B.* **4**, 2008–2010 (2019).
26. Feng, M. et al. The first mitochondrial genome of *Megabalanus tintinnabulum* (Sessilia: Balanidae) from china: phylogeny within cirripedia based on mitochondrial genes. *Mitochondrial DNA Part. B.* **4**, 4016–4018 (2019).
27. Song, J. et al. The first mitochondrial genome of *Fistulobalanus albicostatus* (Crustacea: maxillopoda: Sessilia) and phylogenetic consideration within the superfamily Balanoidea. *Mitochondrial DNA Part. B.* **5**, 2776–2778 (2020).
28. Liu, S. et al. The first mitochondrial genome of *Balanus trigonus* darwin, 1854 (Sessilia: Balanidae) and molecular phylogeny within cirripedia. *Mitochondrial DNA Part. B.* **6**, 1740–1742 (2021).
29. Shen, X., Chu, K. H., Chan, B. K. K. & Tsang, L. M. The complete mitochondrial genome of the fire coral-inhabiting barnacle *Megabalanus ajax* (Sessilia: Balanidae): gene rearrangements and atypical gene content. *Mitochondrial DNA Part. A.* **27**, 1173–1174 (2016).
30. Bae, S., Kim, P. & Yi, C. H. The complete mitochondrial genome of *Balanus trigonus* (Thecostraca, balanomorpha, Balanidae) from South Korea. *Mitochondrial Dna Part. B Resour.* **6**, 2736–2738 (2021).
31. Shen, X., Chan, B. K. K. & Tsang, L. M. The complete mitochondrial genome of common fouling barnacle *Amphibalanus amphitrite* (Darwin, 1854) (Sessilia: Balanidae) reveals gene rearrangements compared to pancrustacean ground pattern. *Mitochondrial DNA*. **26**, 773–774 (2015).
32. Shen, X., Tsang, L. M., Chu, K. H., Achituv, Y. & Chan, B. K. K. Mitochondrial genome of the intertidal acorn barnacle *Tetraclita serrata* darwin, 1854 (Crustacea: Sessilia): Gene order comparison and phylogenetic consideration within Sessilia. *Mar. Genom.* **22**, 63–69 (2015).
33. Song, J., Shen, X., Chu, K. H. & Chan, B. K. K. Mitochondrial genome of the acorn barnacle *Tetraclita rufotincta* pilsbry, 1916: highly conserved gene order in Tetraclitidae. *Mitochondrial DNA Part. B.* **2**, 936–937 (2017).
34. Cai, Y. F., Shen, X., Zhou, L., Chu, K. H. & Chan, B. K. K. Mitochondrial genome of *Tesseropora rosea*: molecular evidence for non-monophyly of the genus Tetraclita. *Mitochondrial DNA Part. B.* **3**, 92–94 (2018).
35. Pérez-Losada, M. et al. Molecular phylogeny, systematics and morphological evolution of the acorn barnacles (Thoracica: sessilia: Balanomorpha). *Mol. Phylogenetics Evol.* **81**, 147–158 (2014).
36. Meng, G., Li, Y., Yang, C. & Liu, S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.* **47**, e63–e63 (2019).
37. Wick, R. R., Holt, K. E. & Polypolish Short-read Polishing of long-read bacterial genome assemblies. *PLoS Comput. Biol.* **18**, e1009802 (2022).
38. Grant, J. R. & Stothard, P. The CGView server: a comparative genomics tool for circular genomes. *Nucleic Acids Res.* **36**, W181–W184 (2008).
39. Hu, F., Lin, Y. & Tang, J. MLGO: phylogeny reconstruction and ancestral inference from gene-order data. *BMC Bioinform.* **15**, 354 (2014).
40. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal Omega. *Mol. Syst. Biol.* **7**, MSB201175 (2011).
41. Edler, D., Klein, J., Antonelli, A. & Silvestro, D. RaxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol. Evol.* **12**, 373–377 (2021).
42. R core team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2023).
43. Schliep, K. P. Phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2010).
44. Wickham, H. ggplot2, Elegant Graphics for Data Analysis. (2009). https://doi.org/10.1007/978-0-387-98141-3
45. Paradis, E. & Schliep, K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2018).
46. Blanchette, M., Kunisawa, T. & Sankoff, D. Gene order breakpoint evidence in animal mitochondrial phylogeny. *J. Mol. Evol.* **49**, 193–203 (1999).
47. Boore, J. L. & Brown, W. M. Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. *Genetics* **138**, 423–443 (1994).
48. Crozier, R. H. & Crozier, Y. C. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**, 97–117 (1993).

49. Feng, M. et al. Complete mitochondrial genome of *Tetraclita squamosa squamosa* (Sessilia: Tetraclitidae) from China and phylogeny within cirripedia. *Mitochondrial DNA Part. B*. **5**, 2121–2123 (2020).
50. Lim, J. T. & Hwang, U. W. The complete mitochondrial genome of *Pollicipes mitella* (Crustacea, maxillopoda, Cirripedia): Non-Monophylies of Maxillopoda and Crustacea. *Mol. Cells*. **22**, 314–322 (2006).
51. Okimoto, R., Macfarlane, J. L. & Wolstenholme, D. R. Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans*. *Nucleic Acids Res.* **18**, 6113–6118 (1990).
52. Gan, Z. et al. Phylogeny and adaptative evolution to chemosynthetic habitat in barnacle (Cirripedia: Thoracica) revealed by mitogenomes. *Front. Mar. Sci.* **9**, 964114 (2022).
53. Uda, K. et al. Complete mitochondrial genomes of two Japanese precious corals, *Paracorallium japonicum* and *Corallium konojoi* (Cnidaria, octocorallia, Coralliidae): Notable differences in gene arrangement. *Gene* **476**, 27–37 (2011).
54. Moritz, C. & Brown, W. M. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proc. Natl. Acad. Sci.* 84, 7183–7187 (1987).
55. Rawlings, T. A., Collins, T. M. & Bieler, R. Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. *Proc. Natl. Acad. Sci.* **100**, 15700–15705 (2003).
56. Romanova, E. V. et al. Hidden cases of tRNA gene duplication and remolding in mitochondrial genomes of amphipods. *Mol. Phylogenetics Evol.* **144**, 106710 (2020).
57. Rawlings, T. A., MacInnis, M. J., Bieler, R., Boore, J. L. & Collins, T. M. Sessile snails, dynamic genomes: gene rearrangements within the mitochondrial genome of a family of caenogastropod molluscs. *BMC Genom*. **11**, 440 (2010).
58. Xu, W., Jameson, D., Tang, B. & Higgs, P. G. The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. *J. Mol. Evol.* **63**, 375–392 (2006).
59. Su, X. et al. Substantial mitochondrial gene order rearrangements and differential evolution rates within the family capitellidae (Annelida). *Zoosystematics Evol.* **101**, 955–967 (2025).
60. Zhang, K., Sun, J., Xu, T., Qiu, J. W. & Qian, P. Y. Phylogenetic Relationships and Adaptation in Deep-Sea Mussels: Insights from Mitochondrial Genomes. *Int. J. Mol. Sci.* 22, (2021). (1900).
61. Miao, J. et al. Sequence comparison of the mitochondrial genomes of five brackish water species of the family neritidae: phylogenetic implications and divergence time Estimation. *Ecol. Evol.* **12**, e8984 (2022).
62. Yuan, Y., Li, Q., Yu, H. & Kong, L. The complete mitochondrial genomes of six heterodont bivalves (Tellinoidea and Solenoidea): Variable Gene Arrangements and Phylogenetic Implications. *PLoS ONE*. **7**, e32353 (2012).
63. Pentinsaari, M., Salmela, H., Mutanen, M. & Roslin, T. Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Sci. Rep.* **6**, 35275 (2016).
64. Park, E. & Poulin, R. Extremely divergent COI sequences within an amphipod species complex: A possible role for endosymbionts? *Ecol. Evol.* **12**, e9448 (2022).
65. Pitombo, F. B. Phylogenetic analysis of the Balanidae (Cirripedia, Balanomorpha). *Zoöl Scr.* **33**, 261–276 (2004).
66. Pochai, A., Kingtong, S., Sukparangsi, W. & Khachonpisitsak, S. The diversity of acorn barnacles (Cirripedia, Balanomorpha) across thailand's coasts: the Andaman sea and the Gulf of Thailand. *Zoosyst Evol.* **93**, 13–34 (2017).
67. Heller, P., Casaletto, J., Ruiz, G. & Geller, J. A database of metazoan cytochrome c oxidase subunit I gene sequences derived from GenBank with CO-ARBitrator. *Sci. Data*. **5**, 180156 (2018).
68. Porter, T. M. & Hajibabaei, M. Over 2.5 million COI sequences in GenBank and growing. *PLoS ONE*. **13**, e0200177 (2018).

Funding

This work was supported by a grant from the National Marine Biodiversity Institute of Korea (grant number: 2025M00300).

Declarations

Competing interests

The authors declare no competing interests.

Supplementary Information

Supplementary Information can be found in Supplementary File 1.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-18483-z.

Correspondence and requests for materials should be addressed to S.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025