Different gene rearrangements of the genus *Dardanus* (Anomura: Diogenidae) and insights into the phylogeny of Paguroidea

Ying Zhang1, Lei Meng2, Liming Wei3, Xinting Lu1, Bingjian Liu1, Liqin Liu1, Zhenming Lü1, Yang Gao2 & Li Gong1

Complete mitochondrial genomes (mitogenomes) can provide useful information for phylogenetic relationships, gene rearrangement, and molecular evolution. In this study, the complete mitogenomes of two hermit crabs, *Dardanus arrosor* and *Dardanus aspersus*, were sequenced for the first time and compared with other published mitogenomes of Paguroidea. Each of the two mitogenomes contains an entire set of 37 genes and a putative control region, but they display different gene arrangements. The different arrangements of the two mitogenomes might be the result of transposition, reversal, and tandem duplication/random loss events from the ancestral pancrustacean pattern. Genome sequence similarity analysis reveals the gene rearrangement in 15 Paguroidea mitogenomes. After synteny analysis between the 15 Paguroidea mitogenomes, an obvious rearranged region is found in *D. aspersus* mitogenome. Across the 13 protein-coding genes (PCGs) tested, COI has the least and ND6 has the largest genetic distances among the 15 hermit crabs, indicating varied evolution rates of PCGs. In addition, the dN/dS ratio analysis shows that all PCGs are evolving under purifying selection. The phylogenetic analyses based on both gene order and sequence data present the monophyly of three families (Paguridae, Coenobitidae, and Pylochelidae) and the paraphyly of the family Diogenidae. Meanwhile, the phylogenetic tree based on the nucleotide sequences of 13 PCGs shows that two *Dardanus* species formed a sister group with five Coenobitidae species. These findings help to better understand the gene rearrangement and phylogeny of Paguroidea, as well as provide new insights into the usefulness of mitochondrial gene order as a phylogenetic marker.

The metazoan mitochondrial genome (mitogenome) is typically a closed circular double-stranded DNA molecule. It is relatively small (15–20 kb) and contains 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs), two ribosomal RNA genes (12S and 16S), and an AT-rich region (also called control region, CR). Owing to some unique features like small genome size, conserved gene content, absence of introns, maternal inheritance, low level of recombination, and fast rate of evolution, the mitogenome has been extensively used in population genetics, comparative genomics, and phylogenetic studies. Besides, comparative analyses of the complete mitogenomes of closely related taxa can deepen the understanding of gene rearrangements and evolutionary relationships.

Generally, the gene order in most vertebrate mitogenomes is considered conserved. However, extensive mitochondrial gene rearrangements have been observed in invertebrate mitogenomes, such as in cephalopods, bivalves, insects, and crabs. The gene rearrangement within a lineage has been supposed to be phylogenetically informative; therefore, comparative analysis of mitochondrial gene order has been proved to be a valuable phylogenetic tool. For example, Yuan et al. compared the mitochondrial gene arrangements of six heterodont bivalves and concluded that *Sanguinolaria* (Psammobiidae) is not monophyletic. By gene order comparisons of echinoderms, Smith et al. provided evidence that the sea cucumbers should group with sea urchins and sea stars with brittle stars. Besides, Morrison et al. reconstructed a phylogeny for decapod taxa based on gene rearrangements and DNA sequences. Their findings supported five independent origins of the crab-like form...
and confirmed the utility of mitochondrial gene rearrangements in phylogenetic reconstruction. So far, three main types of gene rearrangement events have been observed in the mitogenomes of animals, including translocation, shuffling, and inversion. Correspondingly, four mechanisms have been proposed to account for mitogenomic rearrangements, including tandem duplication/random loss (TDRL) model, intramitochondrial recombination model, and tandem duplication/non-random loss model.

The infraorder Anomura consists of approximately 2450 extant species, with a broad range of ecological specializations and with various lineages having successfully adapted to freshwater, terrestrial, and diverse marine environments. The latest classification scheme divides Anomura into Aegloeida, Chirostylidea, Galatheidea, Hippoidea, Lithodoidea, Lomisoidea, and Paguroidea. Hermit crabs (the superfamily Paguroidea Latreille, 1802) consist of Coenobitidae, Diogenidae, Paguridae, Parapaguridae, Pylochelidae and Pylojacquesidae, with more than 1100 species inhabiting diverse biotopes from intertidal to deep seas. They represent an intermediate group of crustaceans from Macrura to Brachyura, which occupy an important position in crustacean evolution.

However, their evolutionary relationships at many taxonomic levels are still waiting for researchers to resolve. Diogenidae Ortmann, 1892 is the second largest family of the superfamily Paguroidea. According to WoRMS, the family Diogenidae has 22 genera and 482 species in total. Among them, the genus *Dardanus* Paulson, 1875 is one of the most diverse genera within Diogenidae. Previous studies of this genus mainly focused on morphology, with almost no attention given to molecular phylogeny. To date, the phylogenetic location of the genus *Dardanus* has not been well resolved.

Accordingly, in the present study, we newly sequenced and annotated the complete mitogenomes of two hermit crabs, *D. arrosor* and *D. aspersus*. These are the first published mitogenomes of the genus *Dardanus*. Combined with 13 available Paguroidea mitogenomes, we performed a comparative mitogenomics analysis, in order to: (a) investigate the characteristics of Paguroidea mitogenomes; (b) estimate the phylogenetic information of mitochondrial gene rearrangements; (c) reconstruct the phylogeny of Paguroidea that could lay a foundation for further evolutionary studies.

**Materials and methods**

**Sampling, DNA extraction, mitogenome sequencing, and assembly.** Specimens of *D. arrosor* and *D. aspersus* were collected from Zhoushan Province, China (29° 45′ 32″ N, 121° 45′ 30″ E). Specimens were immediately preserved in 95% ethanol until DNA extraction. The SQ Tissue DNA Kit (OMEGA) was used to extract the total genomic DNA from muscle tissue following the manufacturer's instructions. The genomic DNA was sent to Shanghai Origingene Biopharm Technology Co., Ltd. for library preparation and high-throughput sequencing. The libraries were constructed by using the VAHTS Universal Plus DNA Library Prep Kit, with an insert size of 150 bp. Paired-end sequencing with a read length of 150 bp was performed on an Illumina Hiseq 6000 platform. Adapters and low-quality bases were removed using cutadapt v1.16 with the following parameters: -q 20 -m 20. Trimmed reads shorter than 50 bp were discarded. Quality control of raw and trimmed reads was performed using FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The filtered clean data were assembled and mapped to complete mitogenome sequence using NOVOPlasty v2.7.2.

**Mitogenome annotation and sequence analysis.** The newly assembled mitogenomes of *D. arrosor* and *D. aspersus* were annotated using the software of Sequin (version 15.10, http://www.ncbi.nlm.nih.gov/Sequin/). The boundaries of protein-coding and ribosomal RNA genes were performed using NCBI-BLAST (http://blast.ncbi.nlm.nih.gov). Transfer RNA genes were manually plotted, according to the secondary structure predicted by the MITOS Web Server and tRNAscan-SE 1.21. The control region was determined by the locations of adjacent genes. Finally, circular mitogenome maps of *D. arrosor* and *D. aspersus* were drawn with the BLAST Ring Image Generator (BRIG) v0.93.

The base composition and relative synonymous codon usage (RSCU) were obtained using MEGA X. The strand asymmetry was calculated using the following formulas: AT-skew = (A − T)/(A + T); GC-skew = (G − C)/(G + C). Furthermore, we chose the complete mitogenome of *Pagurus filholi* as the reference genome for comparative genomic analysis. Genome sequence similarity among the 15 Paguroidea species was carried out using the BRIG tool. Synteny analysis between the genomes was performed using Mauve v2.4.0. The evolutionary-selection constraints on 13 PCGs in the superfamily Paguroidea, the nonsynonymous (dN) and synonymous (dS) substitution rates were calculated using Mega X. The genetic distances of 13 PCGs were also estimated using Mega X based on the Kimura 2-parameter (K2P) substitution model.

**Mitochondrial gene order comparisons and phylogeny.** CREx was used to compare the mitochondrial gene order and infer the gene rearrangement scenarios based on common intervals. CREx considers four types of rearrangement events: reversals (R), transpositions (T), reverse transpositions (RT), and tandem-duplication/random-losses (TDRL). MLGO web server was used to infer a phylogeny from gene order data. To estimate the evolutionary-selection constraints on 13 PCGs in the superfamily Paguroidea, the nonsynonymous (dN) and synonymous (dS) substitution rates were calculated using Mega X. The genetic distances of 13 PCGs were also estimated using Mega X based on the Kimura 2-parameter (K2P) substitution model.

**Phylogenetic analysis.** Phylogeny of the Paguroidea was inferred based on 13 available complete mitogenomes expanded with the two newly determined ones (Table 1). The species *Helicina wuana* and *H. latimera* from Grapsoidea were used as outgroups. Fasta files with the nucleotide sequences for all 13 PCGs were extracted from the GenBank files using PhyloSuite. The MAFFT program integrated into PhyloSuite was executed to align multiple sequences in normal-alignment mode, and ambiguously aligned regions were identified and moved by Gblocks. Subsequently, the sequences were concatenated into a single alignment and converted into input files (Phylip and Nexus format) for phylogenetic analyses. Phylogenetic trees were built under maximum likelihood (ML) and Bayesian inference (BI) methods. The ML analysis was conducted using IQ-TREE, under an ML + rapid bootstrap (BS) algorithm with 1000 replicates. The BI analysis was conducted in MrBayes.
Table 1. Basic information of 15 Paguroidea species and two outgroups used in this paper.

| Species               | Family       | Superfamily | Length (bp) | Accession No | References |
|-----------------------|--------------|-------------|-------------|--------------|------------|
| Pagurus filholi       | Paguridae    | Paguroidea  | 15,674      | LC222528     | 48         |
| Pagurus japonicus     | Paguridae    | Paguroidea  | 16,401      | LC222532     | 48         |
| Pagurus gracilipes    | Paguridae    | Paguroidea  | 16,051      | LC222534     | 48         |
| Pagurus maculatus     | Paguridae    | Paguroidea  | 15,420      | LC222524     | 48         |
| Pagurus nigrofuscus    | Paguridae    | Paguroidea  | 15,423      | LC222531     | 48         |
| Pagurus longicarpus    | Paguridae    | Paguroidea  | 15,630      | NC_003058    | 49         |
| Coenobita rogens      | Coenobitidae | Paguroidea  | 16,427      | KY352235     | 50         |
| Coenobita variebratis  | Coenobitidae | Paguroidea  | 16,421      | KY352236     | 50         |
| Coenobita perlatus     | Coenobitidae | Paguroidea  | 16,447      | KY352234     | 50         |
| Coenobita brevimanus   | Coenobitidae | Paguroidea  | 16,393      | NC_050386    | 50         |
| Birgas latro          | Coenobitidae | Paguroidea  | 16,411      | NC_045091    | 51         |
| Dardanus arrosor       | Diogenidae   | Paguroidea  | 16,592      | MW147148     | This study |
| Dardanus aspersus      | Diogenidae   | Paguroidea  | 16,916      | MW157812     | This study |
| Clinkanias infraspinatus| Diogenidae  | Paguroidea  | 16,594      | NC_025776    | 52         |
| Pylocheles mortenseni  | Pylochelidae | Paguroidea  | 15,093      | KY352242     | 50         |
| Helice wuana           | Varunidae    | Grapoidea   | 16,359      | NC_034995    | Outgroup   |
| Helice latiterra       | Varunidae    | Grapoidea   | 16,246      | NC_033865    | Outgroup   |

3.2.6 with $3 \times 10^6$ metropolis-coupled Markov Chain Monte Carlo (MCMC/CMC) generations, sampling every 1000 generations, and the first 25% of the generations were discarded as burn-in. To guarantee the stationarity had been reached, the average standard deviation of split frequencies was set below 0.01.

Results

General features of *D. arrosor* and *D. aspersus* mitogenomes. The complete mitogenomes of *D. arrosor* and *D. aspersus* are 16,592 bp and 16,916 bp in length, respectively (GenBank accessions MW147148 and MW157812) (Figs. S1, S2, Tables 1, 2). Each mitogenome contains a typical set of 37 genes (13 PCGs, 22 tRNAs, and two rRNAs) and a putative CR. Within these genes, eight PCGs (COI, COII, ND2, ATP8, ATP6, COIII, ND6, and Cyt b) and 11 tRNAs (tRNA-Leu<sup>TAC</sup>, Lys, Met, Ile, Asp, Arg, Asn, Gln, Glu, Thr, Ser<sup>TGA</sup>, and Tyr) are encoded by the heavy (H-) strand, while five PCGs (ND5, ND4, ND4L, ND1, and ND3), 11 tRNAs (tRNA-Phe, His, Pro, Val, Ser<sup>CT</sup>, Ala, Gly, Leu<sup>TAA</sup>, Trp, Gln, and Cys), and two rRNAs (16S and 12S rRNA) are encoded by the light (L-) strand. But the gene order of the two mitogenomes is different.

There are 214 intergenic nucleotides (IGNs) dispersed in 22 locations for *D. arrosor*, 596 IGNs in 22 locations for *D. aspersus*. The longest IGN is 73.3% in *D. arrosor* and between *tRNA-Val* and *tRNA-Pro* (Tables 2, 3). Meanwhile, 27 overlapping nucleotides are located in seven pairs of neighboring genes for both mitogenomes. These overlapping nucleotides vary in length from 1 to 7 bp, and the longest overlap is located between ATP8 and ATP6 as well as ND4 and ND4L (Tables 2, 3).

The base composition of *D. arrosor* is A = 33.3%, T = 34.6%, C = 15.7%, G = 16.4% and that of *D. aspersus* is A = 33.4%, T = 32.6%, C = 15.7%, G = 18.3%. The AT content is 67.9% in *D. arrosor* and between *tRNA-Val* and *tRNA-Pro* (Tables S1, S2).

For ND5 (uses TTG as the start codon) and ND3 (uses TGG as the start codon) in *D. aspersus* mitogenome (Tables 2, 3), the remaining PCGs initiate with typical ATN codons. For the stop codon, the majority of PCGs stop with TAA or TAG except for ND3 (uses a single T as the stop codon) and ND6 (uses GAC as the stop codon) in the two mitogenomes (Tables 2, 3). The GC-skew values of five PCGs (ND5, ND4, ND4L, ND1, and ND3) are positive, indicating they are encoded by the L-strand, whereas the remaining eight exhibit negative values, indicating they are encoded by the H-strand (Tables S1, S2).

For ND5 (uses TTG as the start codon) and ND3 (uses TGG as the start codon) in *D. aspersus* mitogenome (Tables 2, 3), the remaining PCGs initiate with typical ATN codons. For the stop codon, the majority of PCGs stop with TAA or TAG except for ND3 (uses a single T as the stop codon) and ND6 (uses GAC as the stop codon) in the two mitogenomes (Tables 2, 3). The GC-skew values of five PCGs (ND5, ND4, ND4L, ND1, and ND3) are positive, indicating they are encoded by the L-strand, whereas the remaining eight exhibit negative values, indicating they are encoded by the H-strand (Tables S1, S2).

Twenty-two tRNAs of *D. arrosor* and *D. aspersus* mitogenomes are scattered throughout the entire mitogenome (Tables 2, 3). The total length of 22 tRNAs is 1455 bp in *D. arrosor* and 1460 bp in *D. aspersus* (Tables S1, S2). All of the tRNAs can be folded into typical cloverleaf secondary structures except for the tRNA-Ser (TCT) in the two mitogenomes (Figs. S3, S4). The lack of DHU arm in tRNA-Ser (TCT) is thought to be a common phenomenon in metazoan mitogenomes. The 16S rRNA and 12S rRNA genes of *D. arrosor* and *D. aspersus* are located between ND1 and tRNA-Val and between tRNA-Val and CR, respectively. The AT content of the two rRNAs is 73.3% in *D. arrosor*, which is higher than that of *D. aspersus* (70.1%) (Tables S1, S2).

Codon usage bias in Paguroidea mitogenomes. Codon usage bias is a phenomenon in which specific codons are used more frequently than other synonymous codons by certain organisms during the translation of genes to proteins. In this study, the relative synonymous codon usage (RSCU) of 15 hermit crabs is roughly identical. Except for *Pagurus longicarpus* and *Pylocheles mortenseni*, which miss codons, the other 13 species have all 62 available codons. The lost codons usually belong to GC-rich codon-families (Fig. S5, Table S3). The RSCU values for the codons NNU and NNA are usually greater than one, suggesting a strong AT bias in the third
codon position (Fig. S5, Table S3). This result supports the hypothesis that the codon usage bias in PCGs and the AT bias of the third codon position are positively correlated54,55.

Comparative genomic analysis of Paguroidea species. Using the *P. filholi* mitogenome as the reference sequence, all available mitogenomes in the superfamily Paguroidea were compared using BRIG. The results reveal the gene rearrangement in 15 Paguroidea mitogenomes (Fig. 1). The mitogenomes of the family Paguridae are observed to be fairly conserved, with about 80% sequence identity in most regions (six innermost rings in Fig. 1). However, the mitogenomes of the species under the families Coenobitidae, Diogenidae, and Pylochelidae are quite different from the family Paguridae, as can be seen from the larger gap regions in the BRIG map (nine outermost rings in Fig. 1).

By using the Mauve analysis, we identified five large genomic homologous regions (marked A–E in Fig. 2). These homologous regions are commonly presented in all 15 Paguroidea mitogenomes. For the B region, it has the maximum length diversification and is greatly contributed to the genome size variation between Paguroidea mitogenomes (Fig. 2). In addition, we found that homologous regions D and E are rearranged in *D. aspersus* mitogenome. The two homologous regions show an E–D order in *D. aspersus* mitogenome, while the other hermit crabs display a D–E order (Fig. 2).

| Gene  | Position From | Position To | Length (bp) | Amino acid | Start/stop codon | Anticodon | Intergenic region | Strand |
|-------|---------------|-------------|-------------|------------|------------------|-----------|------------------|--------|
| COI   | 1             | 1539        | 1539        | 512        | ATG/TAA          | −5        | H                |
| Leu (L1) | 1535        | 1597        | 63          |            | TAG              | 4         | H                |
| COII  | 1602          | 2294        | 693         | 230        | ATG/TAG          | 6         | H                |
| Iys (K) | 2301          | 2367        | 67          |            | TTT              | 8         | H                |
| Met (M) | 2376          | 2443        | 68          |            | CAT              | 7         | H                |
| Ile (I) | 2451          | 2514        | 64          |            | GAT              | 52        | H                |
| ND2   | 2567          | 3568        | 1002        | 333        | ATT/TAG          | 4         | H                |
| Asp (D) | 3573          | 3637        | 65          |            | GTC              | 0         | H                |
| ATP8  | 3638          | 3796        | 159         | 52         | ATT/TAG          | −7        | H                |
| ATP6  | 3790          | 4464        | 675         | 224        | ATG/TAA          | −1        | H                |
| COIII | 4464          | 5255        | 792         | 263        | ATG/TAA          | 11        | H                |
| Arg (R) | 5267          | 5333        | 67          |            | TCG              | −1        | H                |
| Arg (N) | 5333          | 5397        | 65          |            | GAT              | 3         | H                |
| Glu (E) | 5401          | 5466        | 66          |            | TAC              | 1         | H                |
| Phe (F) | 5468          | 5532        | 65          |            | GAA              | 8         | L                |
| ND5   | 5541          | 7260        | 1720        | 573        | ATG/T             | 0         | L                |
| His (H) | 7261          | 7325        | 65          |            | GTG              | 64        | L                |
| ND4   | 7390          | 8820        | 1431        | 476        | ATG/TAA          | −7        | L                |
| ND4L  | 8814          | 9116        | 303         | 100        | ATG/TAA          | 2         | L                |
| Thr (T) | 9119          | 9185        | 67          |            | TGT              | 12        | H                |
| ND6   | 9198          | 9716        | 519         | 172        | ATG/GAC          | −5        | H                |
| Cyt b  | 9712          | 10,848      | 1137        | 378        | ATG/TAG          | −1        | H                |
| Ser (S) | 10,848        | 10,913      | 66          |            | TGA              | 2         | H                |
| Pro (P) | 10,916        | 10,982      | 67          |            | TGG              | 3         | L                |
| ND1   | 10,986        | 11,915      | 930         | 309        | ATT/TAG          | 0         | L                |
| I6S   | 11,916        | 13,322      | 1407        |            | 0                | 0         | L                |
| Val (V) | 13,323        | 13,391      | 69          |            | TAC              | 1         | L                |
| I2S   | 13,393        | 14,190      | 798         |            | 0                | 0         | L                |
| CR    | 14,191        | 15,686      | 1496        |            | 0                | 0         | H                |
| Ser (S) | 15,687        | 15,753      | 67          |            | TCT              | 0         | L                |
| Ala (A) | 15,754        | 15,815      | 62          |            | TGC              | 5         | L                |
| ND3   | 15,821        | 16,168      | 348         | 115        | ATT/TAG          | 3         | L                |
| Gly (G) | 16,172        | 16,237      | 66          |            | TAC              | 1         | L                |
| Leu (L2) | 16,239        | 16,304      | 66          |            | TAA              | 0         | L                |
| Tyr (T) | 16,305        | 16,373      | 69          |            | GTA              | 8         | H                |
| Trp (W) | 16,382        | 16,450      | 69          |            | TCA              | 0         | L                |
| Gin (Q) | 16,451        | 16,515      | 65          |            | TTG              | 8         | L                |
| Cys (C) | 16,524        | 16,590      | 67          |            | GCA              | 1         | L                |

Table 2. Features of the mitochondrial genome of *D. arrosors*. 
To estimate the evolutionary-selection constraints on 13 PCGs in the superfamily Paguroidea, we perform dN/dS analysis for each PCG. The dN/dS ratios for all PCGs are less than 1, indicating that these genes are evolving primarily under purifying selection. Among them, the lowest dN/dS value (0.113) for \textit{COI} gene indicates the strongest purifying selection, whereas the highest dN/dS value (0.707) for \textit{ATP8} gene shows a highly relaxed purifying selection (Fig. 3). In general, the dN/dS values indicate that the evolution of Paguroidea mitogenomes has been dominated by purifying selection. Besides, we conduct genetic distance analysis for 13 PCGs. \textit{COI} gene possesses the least genetic distance (average 0.237), and \textit{ND6} gene captures the largest value (average 0.494), representing the most conserved and variable genes, respectively (Fig. 3).

### Mitochondrial gene order and rearrangements.

The gene arrangement in the mitogenomes of \textit{D. arrosor} and \textit{D. aspersus} is shown in Fig. 4. The gene order of the two mitogenomes belonging to the same genus is different. Compared with the gene order in ancestral crustaceans (the pancrustacean ground pattern) mitogenomes\(^6\), the gene order in \textit{D. arrosor} and \textit{D. aspersus} mitogenomes underwent large-scale gene rearrangements. For \textit{D. arrosor}, at least six gene clusters (or genes) significantly differ from the typical order, involving 12 tRNAs (L\(_2\), G, A, S\(_1\), P, L\(_1\), I, Q, M, W, C, and Y), and two PCGs (ND3 and ND2). Of these six gene rearrangements, a single L\(_2\) is inverted from the downstream of \textit{COI} in the H-strand to downstream of the G in the

| Gene | Position From | Position To | Length (bp) | Amino acid | Start/stop codon | Anticodon | Intergenic region | Strand |
|------|---------------|-------------|-------------|------------|-----------------|-----------|--------------------|--------|
| COI | 1539 | 1539 | 512 | ATG/TAA | −5 | H |
| Lys (K) | 2297 | 230 | 68 | TAG | 5 | H |
| Met (M) | 2373 | 2306 | 72 | TTT | 9 | H |
| Thr (T) | 2449 | 2383 | 67 | CAT | 7 | H |
| His (H) | 2523 | 2457 | 67 | GAT | 51 | H |
| ND2 | 3576 | 2575 | 1002 | ATT/TAA | 3 | H |
| ATP8 | 3804 | 3646 | 159 | ATC/TAG | −7 | H |
| ATP6 | 4472 | 3798 | 675 | ATG/TAA | −1 | H |
| COII | 5263 | 4472 | 263 | ATG/TAG | 17 | H |
| Arg (R) | 5346 | 5281 | 66 | TCG | 0 | H |
| Asp (D) | 5408 | 5347 | 62 | GTC | 0 | H |
| Glu (E) | 5480 | 5415 | 66 | GTC | 163 | H |
| Thr (T) | 5711 | 5644 | 68 | | 9 | H |
| ND6 | 6239 | 5721 | 519 | ATG/GAC | −5 | H |
| Cyt b | 7371 | 6235 | 1137 | ATG/TAG | −1 | H |
| Ser (S\(_2\)) | 7438 | 7371 | 68 | | 3 | H |
| ND5 | 7233 | 7508 | 1726 | ATG/TAG | −1 | L |
| His (H) | 9299 | 9234 | 66 | GTG | 87 | L |
| ND4 | 1345 | 9387 | 467 | ATG/TAA | −7 | L |
| ND4L | 11,086 | 10,784 | 303 | ATG/TAA | 176 | L |
| Pro (P) | 11,329 | 11,263 | 67 | TGG | 2 | L |
| ND1 | 11,306 | 11,332 | 309 | ATC/TAA | −1 | L |
| 16S | 14,530 | 12,621 | 1400 | | 0 | L |
| Val (V) | 15,991 | 14,531 | 1461 | TGG | 0 | L |
| 12S | 15,954 | 13,730 | 801 | | 0 | L |
| CR | 15,991 | 14,531 | 1461 | TCT | 6 | L |
| Ser (S\(_1\)) | 16,554 | 15,992 | 63 | TGG/TAG | 6 | L |
| Ala (A) | 16,125 | 16,061 | 65 | TGC | 8 | L |
| ND3 | 16,478 | 16,134 | 345 | TTG/TAG | 6 | L |
| Gly (G) | 16,550 | 16,485 | 66 | TGC | 9 | L |
| Leu (L\(_2\)) | 16,627 | 16,560 | 68 | TAA | 0 | L |
| Tyr (Y) | 16,697 | 16,628 | 70 | GTA | 8 | H |
| Trp (W) | 16,775 | 16,706 | 70 | TCA | 0 | L |
| Gin (Q) | 16,838 | 16,776 | 63 | TTG | 9 | L |
| Cys (C) | 16,914 | 16,848 | 67 | GCA | 1 | L |

Table 3. Features of the mitochondrial genome of \textit{D. asperses}.

---

To estimate the evolutionary-selection constraints on 13 PCGs in the superfamily Paguroidea, we perform dN/dS analysis for each PCG. The dN/dS ratios for all PCGs are less than 1, indicating that these genes are evolving primarily under purifying selection. Among them, the lowest dN/dS value (0.113) for \textit{COI} gene indicates the strongest purifying selection, whereas the highest dN/dS value (0.707) for \textit{ATP8} gene shows a highly relaxed purifying selection (Fig. 3). In general, the dN/dS values indicate that the evolution of Paguroidea mitogenomes has been dominated by purifying selection. Besides, we conduct genetic distance analysis for 13 PCGs. \textit{COI} gene possesses the least genetic distance (average 0.237), and \textit{ND6} gene captures the largest value (average 0.494), representing the most conserved and variable genes, respectively (Fig. 3).

**Mitochondrial gene order and rearrangements.** The gene arrangement in the mitogenomes of \textit{D. arrosor} and \textit{D. aspersus} is shown in Fig. 4. The gene order of the two mitogenomes belonging to the same genus is different. Compared with the gene order in ancestral crustaceans (the pancrustacean ground pattern) mitogenomes\(^6\), the gene order in \textit{D. arrosor} and \textit{D. aspersus} mitogenomes underwent large-scale gene rearrangements. For \textit{D. arrosor}, at least six gene clusters (or genes) significantly differ from the typical order, involving 12 tRNAs (L\(_2\), G, A, S\(_1\), P, L\(_1\), I, Q, M, W, C, and Y), and two PCGs (ND3 and ND2). Of these six gene rearrangements, a single L\(_2\) is inverted from the downstream of \textit{COI} in the H-strand to downstream of the G in the
L-strand (Fig. 4A①). The G- ND3- A- S1 cluster is inverted from the downstream of COIII in the H-strand to downstream of the CR in the L-strand (Fig. 4A②). A single P moves from the downstream of T to downstream of the S2 (Fig. 4A③). A single L1 is inverted from the downstream of ND1 in the L-strand to downstream of the COI in the H-strand (Fig. 4A④). The I- Q- M- ND2 cluster is divided into two sections, one (I, M, and ND2) is shifted to downstream of K. The other (Q) is shifted to the end of the linear mitogenome (Fig. 4A⑤). The W- C- Y cluster order is changed into Y- W- C order (Fig. 4A⑥). For D. aspersus, there are also at least six gene clusters (or genes) that differ significantly from the typical order, but the genes involved are different from D. arrosor.

Relative to the gene arrangement of D. arrosor mitogenome, the T- P- ND6- Cyt b- S2 cluster is divided into two sections, one (T, ND6, Cyt b, and S2) is shifted to downstream of E. The other (P) is shifted to downstream of ND4L (Fig. 4A⑤). Based on the CREx analysis, transposition, reversal, and TDRL may be involved in the large-scale gene rearrangements in D. arrosor and D. aspersus mitogenomes (Figs. S6, S7).

The 15 hermit crabs exhibit six types of gene organization (Fig. 5). The mitogenomes of the family Paguridae possess three types of gene order (Type I, Type II, and Type III in Fig. 5). Relative to the remaining three types of gene order (Type IV, Type V, and Type VI in Fig. 5), these three types of gene order are more similar. The mitogenomes of the family Diogenidae possess two types of gene order (Type IV and Type V in Fig. 5). Only the gene arrangement of one gene cluster (T- ND6- Cyt b- S2) is found to be different between the two gene orders. For the remaining two families, Coenobitidae and Pylochelidae, each has only one type of gene order (Type IV and Type VI in Fig. 5). Among them, Coenobitidae shares one of the two gene orders of Diogenidae (Type IV). These results are consistent with the conclusion from the gene order-based phylogenetic tree (Fig. 5). In the gene order tree, all Paguridae species cluster into a clade, showing the closest relationship (Clade I). Species of the family Coenobitidae and Diogenidae are clustered together as a group (Clade II). As the only representative of the family Pylochelidae (Clade III), P. mortensenii forms a separate branch. Our results support that comparisons of mitochondrial gene rearrangements, to some extent, are a useful tool for phylogenetic studies.
Phylogenetic analysis. In the present study, the phylogenetic relationships among Paguroidea were reconstructed based on the nucleotide sequences of 13 PCGs using maximum likelihood (ML) and Bayesian (BI) methods. The phylogenetic trees (ML tree and BI tree) show an identical topology; thus, only one topology (BI) with both support values is displayed (Fig. 6). The results show that *D. arrosor* and *D. aspersus* are most closely related, forming part of the family Diogenidae. Three Diogenidae species are separated into two clades, two
Dardanus species cluster together as a clade, and then form a sister group with five Coenobitidae species. While the remaining one (Clibanarius infraspinatus) forms a separate clade, resulting in the paraphyly of Diogenidae. Besides, of the four families included in the phylogenetic tree, almost all families except Diogenidae form a monophyletic clade. However, the paraphyly of the family Pylochelidae was originally proposed by Richter and Scholtz57 and has been confirmed by many previous researches58,59. Since there is only one representative of the family Pylochelidae in our study, the monophyly of this taxon should be treated with caution.

Discussion

In the present study, our phylogenetic reconstruction based on the nucleotide sequences of 13 PCGs recovered a robust tree (Fig. 6). For a long time, the phylogenetic status of Diogenidae has been controversial. Most previous researches based on morphological features considered it to be a paraphyletic clade. But Forrest60,61 suggested that the family Diogenidae is an ancient monophyletic group. In recent years, an increasing number of molecular studies, including ours, have supported the paraphyly of this taxon. For example, Tsang et al.62 used two nuclear protein genes to conduct phylogenetic inference and clearly pointed out that the family Diogenidae is a paraphyletic clade. In their study, the genus Coenobita (Coenobitidae) is embedded within the branch of the family Diogenidae. Based on mitochondrial gene sequences, Landschoff and Gouws's research recovered the paraphyly of the family Diogenidae as well63. However, there are few studies on the phylogenetic relationships among the genera of the family Diogenidae. Previous studies mainly focused on exploring the phylogenetic relationships of the infraorder Anomura, involving only a few genera and species of the family Diogenidae. Accordingly, increased taxon sampling is required to conclusively resolve the phylogenetic relationships within the family Diogenidae and the superfamily Paguroidea.

Besides, we reconstructed the phylogeny of Paguroidea based on the gene order (Fig. 5). In the family-level relationships within Paguroidea, the gene order information seems to be reliable for phylogenetic inference. A good illustration is that the monophyly of three families (Paguridae, Coenobitidae, and Pylochelidae) and the paraphyly of the family Diogenidae are reconfirmed in the gene order tree (Figs. 5, 6). Our conclusion is in accordance with previous workers, who suggested that gene rearrangements, to some extent, contain phylogenetic information. For example, Shao et al.64 compared the mitochondrial gene arrangements of 12 anomurans and found that Munidopsis lauensis and M. verrilli are most closely related to Shinkaia crosnieri. Based on the comparative analysis of mitochondrial gene arrangement within Coleoidea, Akasaki et al.65 concluded that order Octopoda might be the most ancestral among this subclass Coleoidea. However, the potential to resolve the phylogenetic relationships within families based on gene order alone is clearly inferior to sequence-based approaches. One example is that the monophyly of the genera (Dardanus and Coenobita) is not recovered in the gene order tree (Fig. 5). In future studies, it may be possible to resolve some long-standing phylogenetic controversies by integrating gene order and sequence data.

For most families of the order Decapoda, congeners belonging to the same family share the same gene arrangement generally. Accordingly, it is acceptable to apply gene rearrangement as a molecular marker for phylogenetic inference. A good illustration is that the monophyly of three families (Paguridae, Coenobitidae, and Pylochelidae) and the paraphyly of the family Diogenidae are reconfirmed in the gene order tree (Figs. 5, 6). Our conclusion is in accordance with previous workers, who suggested that gene rearrangements, to some extent, contain phylogenetic information. For example, Shao et al.64 compared the mitochondrial gene arrangements of 12 anomurans and found that Munidopsis lauensis and M. verrilli are most closely related to Shinkaia crosnieri. Based on the comparative analysis of mitochondrial gene arrangement within Coleoidea, Akasaki et al.65 concluded that order Octopoda might be the most ancestral among this subclass Coleoidea. However, the potential to resolve the phylogenetic relationships within families based on gene order alone is clearly inferior to sequence-based approaches. One example is that the monophyly of the genera (Dardanus and Coenobita) is not recovered in the gene order tree (Fig. 5). In future studies, it may be possible to resolve some long-standing phylogenetic controversies by integrating gene order and sequence data.

For most families of the order Decapoda, congeners belonging to the same family share the same gene arrangement generally. Accordingly, it is acceptable to apply gene rearrangement as a molecular marker for phylogenetic inference. A good illustration is that the monophyly of three families (Paguridae, Coenobitidae, and Pylochelidae) and the paraphyly of the family Diogenidae are reconfirmed in the gene order tree (Figs. 5, 6). Our conclusion is in accordance with previous workers, who suggested that gene rearrangements, to some extent, contain phylogenetic information. For example, Shao et al.64 compared the mitochondrial gene arrangements of 12 anomurans and found that Munidopsis lauensis and M. verrilli are most closely related to Shinkaia crosnieri. Based on the comparative analysis of mitochondrial gene arrangement within Coleoidea, Akasaki et al.65 concluded that order Octopoda might be the most ancestral among this subclass Coleoidea. However, the potential to resolve the phylogenetic relationships within families based on gene order alone is clearly inferior to sequence-based approaches. One example is that the monophyly of the genera (Dardanus and Coenobita) is not recovered in the gene order tree (Fig. 5). In future studies, it may be possible to resolve some long-standing phylogenetic controversies by integrating gene order and sequence data.

For most families of the order Decapoda, congeners belonging to the same family share the same gene arrangement generally. Accordingly, it is acceptable to apply gene rearrangement as a molecular marker for phylogenetic inference. A good illustration is that the monophyly of three families (Paguridae, Coenobitidae, and Pylochelidae) and the paraphyly of the family Diogenidae are reconfirmed in the gene order tree (Figs. 5, 6). Our conclusion is in accordance with previous workers, who suggested that gene rearrangements, to some extent, contain phylogenetic information. For example, Shao et al.64 compared the mitochondrial gene arrangements of 12 anomurans and found that Munidopsis lauensis and M. verrilli are most closely related to Shinkaia crosnieri. Based on the comparative analysis of mitochondrial gene arrangement within Coleoidea, Akasaki et al.65 concluded that order Octopoda might be the most ancestral among this subclass Coleoidea. However, the potential to resolve the phylogenetic relationships within families based on gene order alone is clearly inferior to sequence-based approaches. One example is that the monophyly of the genera (Dardanus and Coenobita) is not recovered in the gene order tree (Fig. 5). In future studies, it may be possible to resolve some long-standing phylogenetic controversies by integrating gene order and sequence data.
utility of gene rearrangement as a molecular marker in phylogenetic studies. So it triggers a thought-provoking question that why the mitogenome gene arrangement differs between very closely related species? One possible hypothesis is that the mitogenome gene rearrangement is a continuous and dynamic process and may occur very recently even after speciation events. In future studies, more relevant data are essential to verify this hypothesis.

Figure 5. Phylogeny of Paguroidea reconstructed using gene order data, and arrangement of mitochondrial genes in the ancestral pancrustacean pattern and the superfamily Paguroidea. PCGs and CR are indicated with boxes, and tRNAs are indicated with columns. COI has been designated the start point for the linear representation of the gene arrangement.
Conclusion

In the present study, we reported the mitogenomes of *D. arrosor* and *D. aspersus*, supplementing the limited mitogenome information of the family Diogenidae (Anomura: Paguroidea). By analyzing the mitogenomes of *D. arrosor* and *D. aspersus* and comparing them with other published Paguroidea mitogenomes, we can draw the following conclusions: (a) the gene content of the two mitogenomes belonging to the same genus is conserved, whereas the gene arrangement is different; (b) CREx analysis reveals that transposition, reversal, and TDRL may be involved in the large-scale gene rearrangements in *D. arrosor* and *D. aspersus* mitogenomes; (c) the dN/dS analysis indicates that the evolution of Paguroidea mitogenomes has been dominated by purifying selection; (d) the phylogenetic analyses based on both gene order and sequence data reveal the monophyly of three families (Paguridae, Coenobitidae, and Pylochelidae) and the paraphyly of the family Diogenidae. In future studies, large-scale taxonomic samplings are still needed to further investigate the taxonomical and phylogenetic studies of Paguroidea.

Received: 30 July 2021; Accepted: 11 October 2021
Published online: 08 November 2021

References

1. Boore, J. L. Animal mitochondrial genomes. *Nucl. Acids Res.* **27**, 1767–1780 (1999).
2. Gyllensten, U., Wharton, D., Josefsson, A. & Wilson, A. C. Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**, 255–257 (1991).
3. Sato, M. & Sato, K. Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochim. Biophys. Acta (BBA) Mol. Cell Res.* **1833**, 1979–1984 (2013).
4. Sanchez, G. et al. Population genetics of the jumbo squid *Doxidus gigas* (Cephalopoda: Ommastrephidae) in the northern Humboldt Current system based on mitochondrial and microsatellite DNA markers. *Fish. Res.* **175**, 1–9. https://doi.org/10.1016/j.fishres.2015.11.005 (2016).
5. Ma, Z. et al. Comparative mitogenomics of the genus *Odontobutis* (Perciformes: Gobioidei: Odontobutidae) revealed conserved gene rearrangement and high sequence variations. *Int. J. Mol. Sci.* **16**, 25031–25049. https://doi.org/10.3390/ijms161025031 (2015).
6. Tan, M. H. et al. ORDER within the chaos: Insights into phylogenetic relationships within the Anomura (Crustacea: Decapoda) from mitochondrial sequences and gene order rearrangements. *Mol. Phylogenet. Evol.* **127**, 320–331. https://doi.org/10.1016/j.ympev.2018.05.015 (2018).
7. Ren, L. et al. Comparative analysis of mitogenomes among the subfamily Sarcaphaginae (Diptera: Sarcophagidae) and phylogenetic implications. *Int. J. Biol. Macromol.* **161**, 214–222 (2020).
8. Chen, L.-P. et al. Comparative analysis of mitogenomes among six species of grasshoppers (Orthoptera: Acridoidea: Catantopidae) and their phylogenetic implications in wing-type evolution. *Int. J. Biol. Macromol.* **159**, 1062–1072 (2020).
9. Jiang, L., Kang, L., Wu, C., Chen, M. & Liu, Z. A comprehensive description and evolutionary analysis of 9 Loliginidae mitochondrial genomes. *Hydrobiologia* **808**, 115–124 (2018).
10. Wu, X. et al. New features of Asian *Crassostrea* oyster mitochondrial genomes: A novel alloacceptor tRNA gene recruitment and two novel ORFs. *Gene* **507**, 112–118. https://doi.org/10.1016/j.gene.2012.07.032 (2012).
11. Liu, Q. N. et al. A transfer RNA gene rearrangement in the lepidopteran mitochondrial genome. *Biochem. Biophys. Res. Commun.* **489**, 149–154. https://doi.org/10.1016/j.bbrc.2017.05.115 (2017).
of the manuscript. Z.Y. and L.G. wrote the manuscript. All authors read and approved the final version.

Acknowledgements
This work was supported by the Natural Science Foundation of Zhejiang Province (LY21C190007), Basic Scientific Research Operating Expenses of Zhejiang Provincial Universities (2021JZ003), Zhoushan Science and Technology Bureau (2021C21007) and Open Foundation from Marine Sciences in the First-Class Subjects of Zhejiang.

Author contributions
Y.Z. and L.M. conceived and designed the research. Y.Z., L.W., X.L., B.L., L.L., Z.L., and Y.G. conducted experiments and analyzed data. Z.Y. and L.M. wrote the manuscript. All authors read and approved the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-01338-8.

Correspondence and requests for materials should be addressed to L.G.

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.

© The Author(s) 2021