First mitochondrial genome for the red crab \( (Charybdis feriata) \) with implication of phylogenomics and population genetics

Hongyu Ma\(^1,2\), Chunyan Ma\(^1,2\), Chenhong Li\(^3\), Jianxue Lu\(^1,2\), Xiong Zou\(^1,2\), Yangyang Gong\(^1,2\), Wei Wang\(^1,2\), Wei Chen\(^1,2\), Lingbo Ma\(^1,2\) & Lianjun Xia\(^1,2\)

In this study, we first described the complete mitochondrial genome for the red crab \( (Charybdis feriata) \), elucidated its phylogenetic relationship among 20 species within Decapoda, and estimated the population genetic diversity. The mitochondrial genome was 15,660 bp in size and encoded 13 protein-coding genes, 22 transfer RNA (tRNA) genes, and two ribosomal RNA genes. The gene arrangement of the mitochondrial genome was the same as that of its sister species, \( C. japonica \). Phylogenomic analysis suggested that genus \( Charybdis \) should be classified into subfamily Portuninae but not into subfamily Thalamitinae. Moreover, a total of 33 haplotypes of complete cytochrome \( c \) oxidase subunit I gene were defined in 70 individuals of \( C. feriata \) derived from three localities. Haplotype diversity and nucleotide diversity values among three localities indicated a high level of genetic diversity in \( C. feriata \). AMOVA analysis suggested a low level of genetic differentiation among the three localities \( (F_{ST} = 0.0023, P > 0.05) \). Neutrality tests and mismatch analysis revealed that \( C. feriata \) might have undergone a population expansion event that possibly occurred in the last 61,498 to 43,814 years. This study should be helpful to better understand the evolutionary status, and population genetic diversity of \( C. feriata \) and related species.

The red crab, \( Charybdis feriata \) (Crustacea: Decapoda: Portunidae) (Linnaeus, 1758), also known as the crucifix crab, is a large swimming crab species that is broadly distributed in the Indo-Pacific sea areas, including Japan, China, Indonesia, Australia, India, Pakistan, Oman, and South Africa\(^1,2\). \( C. feriata \) can be easily distinguished by its striking red and white colour pattern and the distinct cross on the median surface of its carapace\(^3\). Young crabs usually live in the sandy shore, whereas adults prefer to inhabit the areas of sandy and muddy bottoms at depths from 30 m to 60 m\(^4\). Given its fast growth speed, large size, good flavor, and high market demand, this species is considered as one of most valuable fisheries resources and has become a potentially important target for aquaculture, domestication, and stock enhancement\(^5,6\). The wild females usually weigh 200 - 350 g, but the males can grow up to 1 kg\(^7\). In the last few decades, the catching production and the wild resources of \( C. feriata \) have been decreasing on a yearly basis\(^8\) due to over-exploitation and environmental deterioration. Despite its economic importance, studies have been limited to reproductive biology\(^7,8\), larvae characteristics\(^9,10\), and fishery biology\(^11\). Little information could be available about the germplasm resource and population genetic structure for this species by now, except that moderate variation was reported for two wild populations sampled from Shanwei City and Zhoushan City, China based on the cytochrome \( c \) oxidase subunit I (COI) gene and

\(^{1}\)East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China. 

\(^{2}\)Key Laboratory of East China Sea and Oceanic Fishery Resources Exploitation, Ministry of Agriculture, Shanghai 200090, China. 

\(^{3}\)College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China. 

Correspondence and requests for materials should be addressed to L.M. (email: malingbo@vip.sina.com) or L.X. (email: alian1@hotmail.com)

Received: 25 February 2015
Accepted: 26 May 2015
Published: 30 July 2015
The genus Charybdis (De Haan, 1833) is an important group in family Portunidae. For a long time, the taxonomic status of this genus has remained controversial. Several studies have recommended to classify it into subfamily Portuninae of family Portunidae14–16. However, some other studies suggested to assign it to subfamily Thalassininae of family Portunidae17,18. Molecular studies based on COI, 16S rRNA, and RAPD19,20 supported the latter opinion. However, the phylogenetic analysis based on 13 protein-coding genes from the mitochondrial genome21 supported the former opinion. Thus, in order to better solve this problem, more studies need to be carried out in the future.

Mitochondrial genome is a typically closed-circular molecule ranging approximately from 14 to 18 kb in size, and it consists of 13 protein-coding genes, 22 transfer RNA (tRNA) genes, 2 rRNA genes, and a control region. It is thought to be an ideal marker for studies on population genetic diversity, molecular phylogeny, and species identification because of its high mutation rate, simple structure, abundant distribution, and maternal inheritance22–24. Thus far, complete mitochondrial genomes have been reported in many crustacean species, such as Litopenaeus vannamei25, C. japonica26, Scylla serrata27, and S. paramamosain28. Although several mitochondrial gene sequences of C. feriata are present in the GenBank database, the complete mitochondrial genome information is still not available by now. The lack of complete mitochondrial genome has limited the development of population genetic diversity and molecular evolution for this species.

The purpose of this study is to report the complete mitochondrial genome for C. feriata, elucidate its evolutionary status, and estimate population genetic diversity and differentiation. This work should be helpful to better understand the evolutionary status and population genetic diversity of C. feriata and other related crustacean species.

Materials and methods
Sampling and DNA extraction. A total of 70 wild individuals of C. feriata were sampled from the southeastern coasts of China, with 21 from Hainan Island (named HN), 24 from the city of Xiamen (named XM), and 25 from the city of Zhoushan (named ZS) (Fig. 1). Animals were killed by a lethal dose of MS-222. Muscle tissues were collected and fixed in 99% ethanol at room temperature. Genomic DNA was extracted using the traditional proteinase K and phenol–chloroform extraction protocol as described by Ma27.

Primers, PCR, and sequencing. First, partial sequences of three genes (12 S rRNA, 16 S rRNA, and COI) of C. feriata and the complete mitochondrial genomes of three closely related crab species (C. japonica, S. paramamosain, and Portunus trituberculatus) were downloaded from the GenBank database. Then the three genes were confirmed by resequencing. Next, the complete mitochondrial genome of C. feriata was generated by overlapping PCR with specific or degenerate primers (Supplementary Table 1), and sequencing. Furthermore, the complete COI gene sequence was employed to evaluate the population genetic diversity and genetic differentiation of C. feriata population. A pair of primers (COI-f: 5′–AATAAGGAAGTATTTAATACTTGTCTT–3′ and COI-r 5′–GAAGAGAAGATGATCTTTGT–3′) with an anealing temperature of 52°C were successfully designed. Seventy individuals collected from three localities (HN, XM, and ZS) were evaluated in this study.

PCRs were carried out in a 25 μL volume that included 0.4 μM each primer, 0.2 mM each dNTP, 1 × PCR buffer, 1.5 mM MgCl₂, 0.75 unit Taq polymerase, and approximately 100 ng template DNA at the following conditions: one cycle of denaturation at 94°C for 4 min; 37 cycles of 30 s at 94°C, 50 s at a primer-specific annealing temperature (Supplementary Table 1), and 50 s at 72°C. As a final step, the products were extended for 7 min at 72°C. The PCR products were separated on 1% agarose gel and directly sequenced in both directions by using an ABI Prism 3730 automated DNA sequencer (PE Corporation). The sequences were edited and assembled using two softwares, EditSeq and SeqMan (DNASTAR).

Complete mitochondrial genome analysis. The graphical map of the complete mitochondrial genome (Fig. 2) was drawn using the online software OrganellarGenomeDRAW (http://ogdraw.mpimp-golm.mpg.de/28). The genome structure was determined by sequence comparisons with the known complete mitochondrial genomes of the closely related species, including S. paramamosain23 and C. japonica26. tRNAs were identified by their proposed clover-leaf secondary structure and anticodons by using the web-based tRNA-scan-SE 1.21 program (http://lowelab.ucsc.edu/tRNAscan-SE/)29 with default search mode. Protein-coding genes were translated into amino acids by using the software MEGA 4.0.30 The codon usage of protein-coding genes and the nucleotide composition of the mitochondrial genome were also determined using MEGA 4.0. Finally, the complete mitochondrial DNA sequence was deposited into the GenBank database by using the software Sequin 12.30 (http://www.ncbi.nlm.nih.gov/Sequin/).

Phylogenomic analysis. The complete mitochondrial genomes of 19 species under Decapoda were downloaded from the GenBank database, including Charybdis japonica (FJ460517), Callinectes sapidus (NC_006281), Eriocheir japonica (NC_011597), Eriocheir hepuensis (NC_011598), Eriocheir sinensis (NC_011599), and Portunus trituberculatus (NC_006281).
Protein-coding genes were aligned using Clustal W in MEGA 4.0 with default settings. As a result, gene ND6 showed high heterogeneity that consistently causes poor phylogenetic performance\(^31\). Thus, the remaining 12 protein-coding genes alignments were concatenated to a single multiple sequence alignment. Then the multiple sequence alignment was formatted and analyzed using RAxML web-servers (http://embnet.vital-it.ch/raxml-bb/index.php)\(^32\). The CAT model was used to estimate the evolutionary rate of the 12 protein-coding genes. Maximum likelihood (ML) search was carried out after bootstraps. The phylogenetic tree was drawn by the software FigTree v1.4.2.

**Population genetic analysis.** Haplotypes were identified using software Dna SP version 5.0\(^33\) and deposited into the GenBank database. For each locality and overall locality, haplotype diversity (\(h\)) and nucleotide diversity (\(\pi\)) were calculated using Dna SP version 5.0. Molecular variance (AMOVA) analysis was carried out using software Arlequin version 3.11\(^34\) to explain the genetic structure and differentiation among these three localities. Significant level of the test was assessed using 1000 permutations of each pairwise comparison. Neutrality tests including Ewens-Watterson\(^35,36\), Chakraborty\(^37\), Tajima\(^38\), and Fu's\(^39\) \(F_s\) with 1000 permutations were performed using software Arlequin version 3.11. Mismatch analysis\(^40\)
with 10000 bootstrap replicates were also performed using Arlequin 3.11. The histogram of mismatch distribution was constructed using the software Network version 4.6.1.2 (http://www.fluxus-engineering.com/)\textsuperscript{41}. The median-joining network of haplotypes was also constructed using software Network version 4.6.1.2. The rough time of population expansion was estimated using the following equation $t = \tau/2u$\textsuperscript{42}, where $t$ is the time since population expansion, $\tau$ is the mutational time scale, which is calculated using software Arlequin version 3.11, and $2u$ is calculated using the equation $2u = \mu \times$ length of sequence $\times$ generation time, where $\mu$ is the mutation rate. Given the lack of a calibrated mutation rate of COI gene of $C. \text{feriata}$, the mutation rates of COI gene ranging from 1.66% to 2.33% per million years of Sesarma\textsuperscript{43} were used. In addition, a generation time of one year was also used.

Results and Discussion

Genome organization. The mitochondrial genome of $C. \text{feriata}$ was a typically circular molecule with 15,660 bp in size (GenBank accession no. KF386147) and consisted of 13 protein-coding genes, 22 tRNA genes, two rRNA genes, and a control region (Fig. 2) as found in most metazoan species such as $Lutra \text{lutra}$\textsuperscript{44}, $C. \text{Japonica}$\textsuperscript{21}, and $S. \text{paramamosain}$\textsuperscript{23}. This genome was slightly smaller than those of most sequenced crab species under Decapoda but larger than that of $P. \text{gigas}$, whose size was 15,515 bp\textsuperscript{21,23,26,45}. Such a small genome was mainly due to the short size of the control region (762 bp). Moreover, the lengths of other regions of the genome were approximately equal among these species. Heavy strand (H-strand)
encoded 23 genes, whereas light strand (L-strand) encoded the remaining 14 genes. The arrangement and order (Table 1) of the 37 genes were completely identical to those of reported species of Portunids, such as *C. japonica* 21 and *S. paramamosain* 23. However, the location of tRNA His (Table 1) was different from that of most arthropods. In most arthropods, tRNA His was between NAD4 and NAD5, whereas it was found between tRNA Glu and tRNA Phe in *C. feriata*. The phenomenon of gene rearrangements in mitochondrial genome was a relatively common event in crustacean species 25. The tandem duplication

| Gene          | Position | Codon | Intergenic nucleotide (bp) | Strand |
|---------------|----------|-------|----------------------------|--------|
| COI           | 1-1534   | 511 ATG T-- | 0                      | H      |
| tRNA ^{UAU}   | 1535-1601 | 67   | 21                        | H      |
| COII          | 1623-2307 | 228 ATG T-- | 0                      | H      |
| tRNA ^{UAU}   | 2308-2375 | 68   | 0                         | H      |
| tRNA ^{UAU}   | 2376-2440 | 65   | 0                         | H      |
| ATP8          | 2441-2602 | 53   | GTG TAG --7                | H      |
| ATP6          | 2596-3273 | 225 ATT TAA --1          | H      |
| COIII         | 3273-4063 | 263 ATG TA--1           | H      |
| tRNA ^{UAU}   | 4063-4126 | 64   | 0                         | H      |
| ND3           | 4127-4480 | 117 ATT TAA 3            | H      |
| tRNA ^{UAU}   | 4484-4559 | 67   | 3                         | H      |
| tRNA ^{UAU}   | 4554-4618 | 65   | 0                         | H      |
| tRNA ^{UAU}   | 4619-4685 | 67   | 0                         | H      |
| tRNA ^{UAU}   | 4686-4754 | 69   | --1                       | H      |
| tRNA ^{UAU}   | 4754-4819 | 66   | 22                        | H      |
| tRNA ^{UAU}   | 4842-4907 | 66   | 0                         | L      |
| tRNA ^{UAU}   | 4908-4971 | 64   | --1                       | L      |
| ND5           | 4971-6698 | 1728 575 ATG TAA 20       | L      |
| ND4           | 6719-8053 | 1335 444 ATG TAA --7      | L      |
| ND4L          | 8047-8349 | 303 100 ATG TAA 2         | L      |
| tRNA ^{UCG}   | 8352-8416 | 65   | 0                         | H      |
| tRNA ^{UAU}   | 8417-8482 | 66   | 2                         | L      |
| ND6           | 8485-8991 | 507 168 ATG TAA --1       | H      |
| Cytb          | 8991-10125| 1135 378 ATG T--0         | H      |
| tRNA ^{UCG}   | 10126-10192| 67  | 27                        | H      |
| ND1           | 10220-11179| 319 ATA TAA 2            | L      |
| tRNA ^{UAU}   | 11182-11250| 69  | 0                         | L      |
| 16 S rRNA     | 11251-12571| 1321| 0                         | L      |
| tRNA ^{UAU}   | 12572-12645| 74  | 0                         | L      |
| 12 S rRNA     | 12646-13488| 843 | 0                         | L      |
| Control region| 13489-14250| 762 | 0                         |        |
| tRNA ^{UAU}   | 14251-14317| 67  | --3                       | H      |
| tRNA ^{UAU}   | 14315-14383| 69  | 3                         | L      |
| tRNA ^{UAU}   | 14387-14455| 69  | 0                         | H      |
| ND2           | 14456-15463| 335 355 ATG TAA --1      | H      |
| tRNA ^{UAU}   | 15463-15531| 69  | --1                       | H      |
| tRNA ^{UAU}   | 15531-15595| 65  | 0                         | L      |
| tRNA ^{UAU}   | 15596-15660| 65  | 0                         | L      |

Table 1. Gene structure of mitochondrial genome of *Charybdis feriata*. a, T-- or TA- represents incomplete stop codons. b, Numbers correspond to the nucleotides separating adjacent genes. Negative numbers indicate overlapping nucleotides. c, H or L indicates that the gene is encoded by the H or L strand.
Table 2. Comparison of mitochondrial genomes of partial crustacean species.

| Species                   | GenBank accession no. | Length (bp) | A + T (%) | No. of amino acid | A + T (%) | 16S rRNA gene | Length (bp) | A + T (%) | 12S rRNA gene | Length (bp) | A + T (%) | Control region | Length (bp) | A + T (%) |
|---------------------------|-----------------------|-------------|-----------|-------------------|-----------|---------------|-------------|-----------|---------------|-------------|-----------|----------------|-------------|-----------|
| Charybdis feriata        | KF386147              | 15,660      | 70.15     | 3,716             | 68.60     | 16S rRNA      | 1,321       | 74.26     | 12S rRNA      | 843         | 71.89     | 1,473          | 71.76       | 762       |
| Charybdis japonica       | FJ460517              | 15,738      | 69.20     | 3,712             | 67.80     |               | 1,317       | 74.20     |               | 834         | 70.30     | 1,458          | 70.90       | 863       |
| Scylla tranquebarica     | NC_012567             | 15,833      | 73.80     | 3,716             | 72.00     | 16S rRNA      | 1,339       | 77.10     | 12S rRNA      | 869         | 75.90     | 1,486          | 74.40       | 854       |
| Scylla olivacea          | NC_012569             | 15,723      | 69.40     | 3,715             | 67.30     |               | 1,337       | 74.40     |               | 852         | 72.40     | 1,482          | 72.30       | 778       |
| Scylla serrata           | HM590866              | 15,721      | 69.22     | 3,714             | 69.20     | 16S rRNA      | 1,337       | 74.50     | 12S rRNA      | 839         | 71.80     | 1,478          | 72.26       | 788       |
| Scylla paramamosain      | JX457150              | 15,824      | 73.04     | 3,715             | 70.88     | 16S rRNA      | 1,340       | 77.46     | 12S rRNA      | 869         | 75.72     | 1,482          | 74.56       | 833       |
| Portunus trituberculatus | AB093006              | 16,026      | 70.20     | 3,715             | 68.80     |               | 1,332       | 73.80     |               | 840         | 70.10     | 1,468          | 72.00       | 1,104     |
| Callinectes sapidus      | NC_006992             | 16,263      | 69.10     | 3,712             | 67.00     | 16S rRNA      | 1,323       | 71.80     |               | 785         | 70.30     | 1,463          | 71.60       | 1,435     |
| Pseudocarcinus gigas     | NC_006891             | 16,354      | 71.70     | 3,718             | 68.90     |               | 1,311       | 77.40     |               | 899         | 76.60     | 1,473          | 72.40       | 896       |
| Geothelphusa dehaami     | NC_007379             | 18,197      | 74.90     | 3,711             | 71.50     |               | 1,315       | 77.10     |               | 821         | 73.80     | 1,460          | 73.20       | 593       |
| Fenneropenaeus chinesis  | DQ518969              | 16,004      | 68.90     | 3,710             | 67.50     | 16S rRNA      | 1,367       | 72.70     |               | 852         | 69.90     | 1,501          | 65.90       | 997       |
| Litopenaeus vannamei     | DQ534543              | 15,989      | 67.80     | 3,710             | 66.10     |               | 1,371       | 71.80     |               | 853         | 69.60     | 1,493          | 65.30       | 998       |

Protein-coding genes. A total of 13 protein-coding genes were identified, of which 9 (COI, COII, COIII, APT6, ATP8, ND2, ND3, ND6, and Cyt b) were encoded by H-strand, and 4 (ND1, ND4, ND4L, and ND5) were encoded by L-strand. These genes consisted of 11,182 bp in length and coded 3716 amino acids in total. All 13 genes were initiated by the start codon ATN (ATG, ATA, and ATT), with an exception (GTG) in ATP8 (Table 1). The typical stop codon (TAA or TAG) was detected in nine protein-coding genes (ATP6, ATP8, ND1, ND2, ND3, ND4, ND4L, ND5, and ND6), whereas the remaining four genes (COI, COII, COIII, and Cyt b) were ended by incomplete stop codons (T- or TA-). Variable start codons and incomplete stop codons have been reported in many other mitochondrial genomes. For example, four types of start codons (ATT, ATG, ATA, and ACG) were detected in mitochondrial genome of Myrmeleon immanis. Two and one incomplete stop codons were found in Lutjanus russelli and S. paramamosain mitochondrial genomes. For the incomplete stop codon, the missed nucleotides may be produced by post-transcriptional polyadenylation. In addition, ND6 had the highest A + T content (78.74%), whereas the protein-coding region had the lowest A + T content (68.60%).

Transfer and ribosomal RNA genes. A total of 22 rRNA genes ranging from 64 to 74 bp in length were identified from the mitochondrial genome of C. feriata. All of them were capable of folding into a typically clover-leaf secondary structure (Fig. 3). In the closely related crabs C. japonica and S. paramamosain, tRNA\textsuperscript{Ser} (AGN) could not form a secondary structure because it lacked the dihydrouracil (DHU) arms. Fourteen tRNA genes were located on H-strand, whereas the remaining eight were located on L-strand. All tRNA genes had a common length of 7 bp for the aminoacyl stem and an invariant size of 7 bp for the anticodon loop. Variable nucleotide lengths of tRNAs were found at the DHU, T\textsuperscript{C}, and anticodon arms. All these 22 rRNA genes possessed the common anticodons of Decapods mitochondrial genomes, except that tRNA\textsuperscript{A\textsubscript{IV}} and tRNA\textsuperscript{Ser} (AGN) possessed TTT and TCT anticodons rather than CTT and GCT, respectively. Seven unmatched base pairs were found in 22 tRNA genes, which was lower than the number detected from tRNA genes of other mitochondrial genomes. The overall A + T content of 22 rRNA was 71.76%, with the highest content (84.62%) in tRNA\textsuperscript{A\textsubscript{IV}} and the lowest content (57.35%) in tRNA\textsuperscript{A\textsubscript{IV}}.

Both 16S and 12S rRNA genes were located on the L-strand of the mitochondrial genome. They were located between tRNA\textsuperscript{Leu} (CUN) and the putative control region and were separated by tRNA\textsuperscript{Val}. The sizes of 16S rRNA and 12S rRNA genes were 1321 and 843 bp, and the A + T contents were 74.26% and 71.89%, respectively.
| Region                | A     | G     | T     | C     | A + T content (%) |
|-----------------------|-------|-------|-------|-------|-------------------|
| Protein-coding gene   |       |       |       |       |                   |
| COI                   | 26.73 | 15.65 | 38.20 | 19.43 | 64.93             |
| COII                  | 31.97 | 14.01 | 34.60 | 19.42 | 66.57             |
| ATP8                  | 29.01 | 9.26  | 40.12 | 21.60 | 69.14             |
| ATP6                  | 27.73 | 11.36 | 40.27 | 19.65 | 67.99             |
| COIII                 | 26.42 | 15.55 | 38.43 | 19.60 | 64.85             |
| ND3                   | 29.10 | 12.15 | 41.24 | 17.51 | 70.34             |
| ND5                   | 31.89 | 19.16 | 39.24 | 9.72  | 71.12             |
| ND4                   | 29.21 | 18.73 | 42.25 | 9.81  | 71.46             |
| ND4L                  | 26.07 | 21.45 | 42.57 | 9.90  | 68.65             |
| ND6                   | 27.02 | 7.50  | 45.76 | 19.72 | 72.78             |
| Cytb                  | 28.28 | 13.66 | 38.06 | 20.00 | 66.34             |
| ND1                   | 27.50 | 19.38 | 42.08 | 11.04 | 69.58             |
| ND2                   | 27.18 | 8.53  | 42.46 | 21.83 | 69.64             |
| tRNA gene             |       |       |       |       |                   |
| tRNA Lys (UUR)        | 34.33 | 16.42 | 35.82 | 13.43 | 70.15             |
| tRNA Thr             | 26.47 | 20.59 | 30.88 | 22.06 | 57.35             |
| tRNA Asp              | 41.54 | 9.23  | 43.08 | 6.15  | 84.62             |
| tRNA Glu              | 42.19 | 9.38  | 34.38 | 14.06 | 76.56             |
| tRNA Ala             | 35.82 | 14.93 | 35.82 | 13.43 | 71.64             |
| tRNA Cys              | 32.31 | 13.85 | 32.31 | 21.54 | 64.62             |
| tRNA Met              | 41.79 | 14.93 | 29.85 | 13.43 | 71.64             |
| tRNA Ser (AGN)        | 34.78 | 14.49 | 33.33 | 17.39 | 68.12             |
| tRNA Glu             | 33.33 | 13.64 | 39.39 | 13.64 | 72.73             |
| tRNA His             | 31.82 | 22.73 | 36.36 | 9.09  | 68.18             |
| tRNA The             | 40.62 | 15.62 | 35.94 | 7.81  | 76.56             |
| tRNA Thr             | 38.46 | 12.31 | 38.46 | 10.77 | 76.92             |
| tRNA Phe             | 34.85 | 16.67 | 40.91 | 7.58  | 75.76             |
| tRNA Ser (UCN)       | 44.78 | 13.43 | 34.33 | 7.46  | 79.10             |
| tRNA Lys (CCU)       | 39.13 | 14.49 | 37.68 | 8.70  | 76.81             |
| tRNA Val             | 28.38 | 22.97 | 33.78 | 14.86 | 62.16             |
| tRNA Il             | 35.82 | 16.42 | 37.31 | 10.45 | 73.13             |
| tRNA Glu             | 31.88 | 21.74 | 34.78 | 11.59 | 66.67             |
| tRNA Met             | 34.78 | 13.04 | 33.33 | 18.84 | 68.12             |
| tRNA Thr             | 43.48 | 11.59 | 31.88 | 13.04 | 75.36             |
| tRNA Cys             | 36.92 | 15.38 | 36.92 | 10.77 | 73.85             |
| tRNA Tyr             | 35.38 | 18.46 | 35.38 | 10.77 | 70.77             |
| rRNA gene            |       |       |       |       |                   |
| 16 S rRNA             | 38.23 | 16.58 | 36.03 | 9.16  | 74.26             |
| 12 S rRNA             | 37.25 | 17.79 | 34.64 | 10.32 | 71.89             |
| Control region       | 41.99 | 7.87  | 36.75 | 13.39 | 78.74             |
| Overall of protein-coding genes | 28.55 | 15.25 | 40.05 | 16.14 | 68.60 |
| Overall of tRNA genes | 36.25 | 15.61 | 35.51 | 12.63 | 71.76             |
| Overall of rRNA genes | 37.85 | 17.05 | 35.49 | 9.61  | 73.34             |
| Overall of the genome | 34.09 | 11.25 | 36.05 | 18.60 | 70.15             |

Table 3. The base composition for different regions of mitochondrial genome of *Charybdis feriata* (the genes which are encoded by the L-strand are converted to complementary strand sequences).
Figure 3. Putative secondary structures of 22 tRNAs detected from mitochondrial genome of *Charybdis feriata*. 
Non-coding regions. A total of 11 non-coding regions were identified in the mitochondrial genome of *C. feriata*. The major non-coding region (762 bp in length) was found between 12S rRNA and tRNA_{Ile}, which was considered to be the putative control region. The other 10 non-coding regions were small, ranging from 2 to 27 bp in length. The A + T content of control region was higher (78.74%) than that of other regions in mitochondrial genome. The high rate of A + T content was due to the existence of A/T repeated motifs. In control region, TA, AT, TAA, AAT, and TTA were found to be the most abundant motifs. Additionally, microsatellite sequences were detected, such as (TA)$_n$, (AT)$_n$, (TA)$_m$, (AT)$_m$, and (TA)$_{12}$. Microsatellites were also identified from control region of the mitochondrial genome in *M. immanis* and *Nymphes myrmeleonoides*

Phylogenetic relationship. The taxonomic status of genus *Charybdis* within Portunidae has been a highly contentious issue for a long time. In this study, we estimated the evolutionary relationship of *C. feriata* within Decapoda by reconstructing a phylogenetic tree. This tree was created based on 12 concatenated protein-coding genes from the mitochondrial genome of 21 species. From the tree topologies (Fig. 4), we found that *C. feriata* and *C. japonica* first formed a monophyletic group and showed the closest relationship to each other. Together with *C. sapidus* and *P. trituberculatus*, they then formed another monophyletic group. Furthermore, these four species showed a sister relationship with another four species of genus *Scylla*. Thus, our results supported the opinion of classifying genus *Charybdis* into subfamily Portuninae of family Portunidae. The same suggestion was also proposed by Liu and Cui.

Population genetic diversity and differentiation. The complete COI gene sequence (1534 bp) was employed to estimate the genetic diversity and differentiation of *C. feriata* population. A total of 33 haplotypes (Table 4) were identified from 70 individuals, of which 14 were from HN locality, 11 from XM locality, and 16 from ZS locality. H9 was the most abundant haplotype, which was present in each locality. A high level of genetic diversity (Table 5) was found with $h$ ranging from 0.819 to 0.867 and $\pi$ ranging from 0.0011 to 0.0013 per locality. The $h$ value was slightly higher than that (0.787) reported by Huang. In our previous study, moderate genetic variation of *C. feriata* was detected by microsatellite markers. Moreover, a high genetic diversity was also observed in other marine animals, such as *S. serrata* and *Salmo salar*. The following factors, including life history characteristics, environmental heterogeneity, and large population sizes, may help in maintaining a high level of genetic diversity.
AMOVA analysis indicated that 99.77% of the total genetic variation was contributed by within-localities variation, whereas only 0.23% was caused by among-localities variation ($F_{ST} = 0.0023, P > 0.05$) (Table 6). In addition, no significant genetic differentiation was found among three localities (Table 7). The above analysis showed that the genetic differentiation among three localities of *C. feriata* was at a low level, suggesting a single population in East China Sea and South China Sea. The summer southwest monsoonal wind and winter northeast monsoonal wind drive seasonal ocean current among marine animals.\(^{53,54}\)
This homogenous population structure of *C. feriata* might be related with the ocean current circulation and the high larval dispersal ability of this crab species.

Neutrality tests, mismatch analysis and population expansion estimation. Four kinds of neutrality tests, including Ewens-Watterson, Chakraborty, Tajima’s $D$, and Fu’s $F_s$, were carried out in *C. feriata*. All of them, except Ewens-Watterson, suggested a significant deviation from mutation-drift equilibrium (Table 8). Mismatch distributions analysis (Table 9) indicated that the estimated effective population size after population growth was significantly larger than that before population growth. In addition, a star-like topology was produced based on 33 haplotypes (Fig. 5). Of these haplotypes, H9 was located in the center of this topology, and it was closely linked with the majority of other haplotypes, thereby suggesting that it is the ancestral haplotype in *C. feriata*. All above analysis indicated that *C. feriata* might have undergone a population expansion event. Meanwhile, the high $h$ (between 0.819 and 0.867) and low $\pi$ (between 0.0011 and 0.0013) also suggested that *C. feriata* underwent population expansion event.
expansion after a period of low effective population size. Sudden population expansion can affect population genetic diversity and haplotypes, and in this process more haplotypes were generated by mutation than were removed by genetic drift. We further deduced that the population expansion event of C. feriata might have occurred between 61,498 and 43,814 years ago. This period of population expansion is a little bit later than the Last Interglacial complex (140-75 kya). This findings showed that the Last Interglacial complex might have played an important role in demographic history of C. feriata. Further, a big changes of nutrient concentrations and sea water temperature could affect the population distribution of marine organisms too.

Conclusion
This study first described the complete mitochondrial genome of C. feriata, which was 15,660 bp in length, including a typical set of 37 genes and a control region. Phylogenomic analysis results supported that genus Charybdis should be classified into subfamily Portuninae rather than into subfamily Thalamitinae. Furthermore, a high level of genetic diversity and a low level of differentiation of C. feriata were found, and a population expansion event was deduced to have occurred between 61,498 and 43,814 years ago. This study should be helpful for studies on evolution and phylogeny, population genetic structure, and conservaton genetics for C. feriata and related species.

References
1. Stephenson, W. An annotated check list and key to the Indo-west-Pacific swimming crabs (Crustacea: Decapoda: Portunidae). B Roy Soc New Zeal 10, 1–64 (1972).
2. Apel, M. V. & Spiridonov, A. Taxonomy and zoogeography of the portunid crabs (Crustacea: Decapoda: Brachyura: Portunidae) of the Arabian Gulf and adjacent waters. Fauna Arabia 17, 159–331 (1998).
3. Padayatti, P. S. Notes on population characteristics and reproductive biology of the portunid crab Charybdis feriatus (Linnaeus) at Cochin. Indian J Fish 37, 155–158 (1990).
4. Baylon, J. & Suzuki, H. Effects of changes in salinity and temperature on survival and development of larvae and juveniles of the crucifix crab Charybdis feriata (Crustacea: Decapoda: Portunidae). Aquaculture 269, 390–401 (2007).
5. Williams, M. J. & Primavera, J. H. Choosing tropical portunid species for culture, domestication and stock enhancement in the Indo-Pacific. Asian Fish Sci 14, 121–142 (2001).
6. Parado-Estepa, F. D., Quinitio, E. T. & Rodriguez, E. M. Seed production of the crucifix crab Charybdis feriatus. Aquac Res 38, 1452–1458 (2007).
7. Jouleen, J. Captive spawning, hatching and larval development of crucifix crab, Charybdis feriata (Linnaeus, 1758). J Mar Biol Ass India 53, 35–40 (2011).
9. Motoh, H. & Villaluz, A. Larvae of decapod crustaceans of the Philippines. The zoal stages of a swimming crab, *Charybdis cruciata* (Herbst), reared in the laboratory. *Bull Ipn Soc Sci Fish* 42, 523–531 (1976).

10. Fielder, D. R., Greenwood, J. G. & Campbell, G. The megalopa of *Charybdis feriata* (Linnaeus) with additions to the zoal larval descriptions (Decapoda, Portunidae). *Crustaceana* 46, 160–165 (1984).

11. Sarada, P. T. Crab fishery of the Calicut coast with some aspects of the population characteristics of *Portunus sanguinolentus*, *Portunus pelagicus* and *Charybdis cruciata*. *Indian J Fish* 45, 375–386 (1998).

12. Huang, Y. F. Amplification and analysis about *Charybdis feriata* COI gene sequence. *J Grad Sun Yat-Sen Univ* 30, 57–64 (2009).

13. Ma, H. Y. et al. Discovery and characterization of a first set of polymorphic microsatellite markers in red crab (*Charybdis feriata*). *J Genet* 92, e113–e115 (2013).

14. Stephenson, W. & Campbell, B. The Australian portunids (*Crustacea: Portunidae*) IV: remaining genera. *Aust J Mar Freshwater Res* 11, 73–122 (1960).

15. Dai, A. Y., Yang, S. L., Song, Y. Z. & Chen, G. X. Marine crabs of China 189 (China Ocean Press, 1986).

16. Doe, D. P. & Ng, P. K. L. Swimming crabs of the genera *Charybdis* De Haan, 1833 and *Thaeulma* Latreille, 1829 (*Crustacea: Decapoda: Brachyura: Portunidae*) from peninsular Malaysia and Singapore. *Raffles Bull Zool* 1, 1–128 (1995).

17. Ng, P. K. L., Wang, C. H., Ho, P. H. & Shih, H. T. An annotated checklist of brachyuran crabs from Taiwan (*Crustacea: Decapoda*). *Natl Taiwan Mus Spec Publ Ser* 11, 1–86 (2001).

18. Liu, Y. J. Checklist of marine biota of China seas 789–791 (*Science Press*, 2008).

19. Jin, S., Zhao, Q. S., Wang, C. L. & Chen, Y. E. RAPD genetic markers in six species of marine crab. *Zool Res* 25, 172–176 (2004).

20. Zhang, S. et al. The application of mitochondrial DNA in phylogeny reconstruction and species identification of portunid crab. *Mar Sci* 32, 9–18 (2008).

21. Liu, Y. & Cui, Z. Complete mitochondrial genome of the Asian paddle crab *Charybdis japonica* (*Crustacea: Decapoda: Portunidae*): gene rearrangement of the marine brachyurans and phylogenetic considerations of the decapods. *Mol Biol Rep* 37, 2559–2569 (2010).

22. Yu, J. F., Azuma, N. & Abe, S. Genetic differentiation between collections of hatchery and wild masu salmon (*Oncorhynchus masou*) inferred from mitochondrial and microsatellite DNA analysis. *Environ Biol Fish* 94, 259–271 (2012).

23. Ma, H. Y. et al. The complete mitochondrial genome sequence and gene organization of the mud crab (*Scylla paramamosain*) with phylogenetic consideration. *Gene* 519, 120–127 (2013).

24. Baek, S. Y. et al. Complete mitochondrial genomes of *Carcinocorpus rotundicauda* and *Tachypleus tridentatus* (*Xiphosura, Arthropoda*) and implications for chelicerate phylogenetic studies. *Int J Biol Sci* 10, 479–489 (2014).

25. Shen, X. et al. The complete mitochondrial genomes of two common shrimps (*Litopenaeus vannamei* and *Fenneropenaeus chinensis*) and their phylogenetic considerations. *Gene* 403, 98–109 (2007).

26. Jordeung, A., Karinthanyakit, W. & Kaewkhumsan, J. The complete mitochondrial genome of the black mud crab, *Scylla serrata* (*Crustacea: Brachyura: Portunidae*) and its phylogenetic position among (pan) crustaceans. *Mol Biol Rep* 39, 10921–10937 (2012).

27. Ma, H. Y., Yang, J. F., Su, P. Z. & Chen, S. L. Genetic analysis of gynogenetic and common populations of *Verapar mosei* using SSR markers. *Wuhan Univ J Nat Sci* 14, 267–273 (2009).

28. Lohse, M., Drechsel, O. & Bock, R. *OrganellarGenomeDRAW* (*OGDRAW*): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* 52, 267–274 (2007).

29. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25, 955–964 (1997).

30. Kumar, S., Dudley, J., Nei, M. & Tamura, K. *MEGA*: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9, 299–306 (2008).

31. Miyata, M. & Nishida, M. Use of mitogenetic information in teleostean molecular phylogenetics: a base-based exploration under the maximum-parsimony optimality criterion. *Mol Phylogenet Evol* 17, 437–455 (2000).

32. Stamatakis, A., Hoover, P. & Rougemont, J. A rapid bootstrap algorithm for the RAXML web-servers. *Syst Biol* 75, 758–771 (2008).

33. Librado, P. & Rozas, J. *DnaSP v5*: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452 (2009).

34. Excoffier, L., Laval, G. & Schneider, S. *Arlequin* (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform* 1, 47–50 (2005).

35. Edwards, W. J. The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3, 87–112 (1972).

36. Watterson, G. The homozygoisty test of neutrality. *Genetics* 88, 405–417 (1978).

37. Chakraborty, R. Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations. *Am J Hum Genet* 47, 87–94 (1990).

38. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595 (1989).

39. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915–925 (1997).

40. Rogers, A. Genetic evidence for a pleistocene population explosion. *Evolution* 49, 608–615 (1995).

41. Polzin, T. & Daneshmand, S. V. On steiner trees and minimum spanning trees in hypergraphs. *Mol Biol Evol* 9, 552–569 (1992).

42. Schubert, C. D., Diesel, R. & Hedges, S. B. Rapid evolution to terrestrial life in Jamaican crabs. *Nature* 393, 363–365 (1998).

43. Kj. J., Hwang, D., Park, T., Han, S. & Lee, J. A comparative analysis of the complete mitochondrial genome of the Eusarcan otter *Lutra lutra* (*Carnivora; Mustelidae*). *Mol Biol Rep* 37, 1943–1955 (2010).

44. Miller, A. D., Murphy, N. P., Burridge, C. P. & Austin, C. M. Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (*Menippidae*) and *Macrobrachium rosenbergii* (*Palaemonidae*). *Mar Biotechnol* 7, 339–349 (2005).

45. Yamauchi, M. M., Miya, M. U. & Nishida, M. Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (*Crustacea: Decapoda: Brachyura*). *Gene* 311, 129–135 (2003).

46. Wang, Y., Liu, X. & Yang, D. The first mitochondrial genome for Caddisfly (*Insecta: Trichoptera*) with phylogenetic implications. *Int J Biol Sci* 10, 53–63 (2014).

47. Van, Y., Wang, Y., Liu, X., Wintertert, S. L. & Yang, D. The first mitochondrial genomes of antlion (*Neuroptera: Myrmeleontidae*) and split-footed lacewing (*Neuroptera: Nymphidae*) with phylogenetic implications of Myrmeleontiformia. *Int J Biol Sci* 10, 895–908 (2014).

48. Guo, Y., Wang, Z., Liu, C. & Liu, Y. Sequencing and analysis of the complete mitochondrial DNA of Russell’s snapper (*L. russelli*). *Prog Nat Sci* 18, 1233–1238 (2008).

49. Ojala, D., Montoya, J. & Attard, G. RNA punctuation model of RNA processing in human mitochondrial. *Nature* 290, 470–474 (1981).

50. Fratini, S., Ragionieri, L. & Cannicci, S. Stock structure and demographic history of the Indo-West Pacific mud crab *Scylla serrata*. *Estuar Coast Shelf* 86, 51–61 (2010).
52. Karlsson, S., Moen, T. & Hindar, K. Contrasting patterns of gene diversity between microsatellites and mitochondrial SNPs in farm and wild Atlantic salmon. *Conserv Genet* **11**, 571–582 (2010).

53. Nei, M. *Molecular evolutionary genetics* 512 (Columbia University Press, 1987).

54. Avise, J. C. *Phylogeography: The history and formation of species* 447 (Harvard University Press, 2000).

55. Chao, S. Y. Circulation of the East China Sea, a numerical study. *J Oceanogr Soc Japan* **46**, 273–295 (1991).

56. Fang, G. H. et al. A note on the South China Sea shallow interocean circulation. *Adv Atmo Sci* **22**, 946–954 (2005).

57. Avise, J. C., Neigel, J. E. & Arnold, J. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J Mol Evol* **20**, 99–105 (1984).

58. Stirling, C. H., Esat, T. M., Lambeck, K. & McCulloch, M. T. Timing and duration of the Last Interglacial: evidence for a restricted interval of widespread coral reef growth. *Earth Planet Sci Lett* **160**, 745–762 (1998).

59. Hayes, G. C., Richardson, A. J. & Robinson, C. Climate change and marine plankton. *Trends Ecol Evol* **20**, 337–344 (2005).

60. Place, A. R., Feng, X., Steven, C. R., Fourcade, H. M. & Boore, J. L. Genetic markers in blue crabs (*Callinectes sapidus*) II. Complete mitochondrial genome sequence and characterization of genetic variation. *J Exp Mar Biol Ecol* **319**, 15–27 (2005).

61. Sun, H., Zhou, K. & Song, D. Mitochondrial genome of the Chinese mitten crab *Eriocheir japonica sinensis* (Brachyura: Thoracotremata: Grapsoidea) reveals a novel gene order and two target regions of gene rearrangements. *Gene* **349**, 207–217 (2005).

62. Segawa, R. D. & Aotsuka, T. The mitochondrial genome of the Japanese freshwater crab, *Geothelphusa dehaani* (Crustacea: Brachyura): evidence for its evolution via gene duplication. *Gene* **355**, 28–39 (2005).

Acknowledgments

This study was supported by the Special Scientific Research Funds for Central Non-profit Institutes (East China Sea Fisheries Research Institute) (Grant No. 2012M04), the National Basic Research Special Foundation of China (Grant No. 2013FY110700), the Special Fund for Agroscientific Research in the Public Interest (Grant No. 201303047), and the National Infrastructure of Fishery Germplasm Resources.

Author Contributions

H.M., L.M. and L.X. conceived and designed the experiments. H.M., C.M., J.L., X.Z., Y.G., W.W. and W.C. performed the experiments. H.M., C.M., C.L. and L.M. analyzed the data. H.M., L.M. and L.X. contributed reagents and materials. H.M. wrote the paper. H.M., L.M. and L.X. revised the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ma, H. et al. First mitochondrial genome for the red crab (*Charybdis feriata*) with implication of phylogenomics and population genetics. *Sci. Rep.* **5**, 11524; doi: 10.1038/srep11524 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/