Gene 675 (2018) 27–35

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Characterization of four new mitogenomes from Ocypodoidea & Grapsoidea, and phylomitogenomic insights into thoracotreme evolution

Jianqin Chen*, Yuhui Xing, Wenjia Yao, Chenling Zhang, Zhenhua Zhang, Gongcheng Jiang, Zhengfeng Ding

Institute of Aquatic Biology, Jiangsu Key Laboratory of Biofunctional Molecule, School of Life Sciences, Chemistry & Chemical Engineering, Jiangsu Second Normal University, 77 West Beijing Road, Nanjing 210013, China

ARTICLE INFO

Keywords:
Mitochondrial genome
Genomic rearrangement
Phylogenetic relationship
Thoracotremata
Ocypodoidea
Grapsoidea

ABSTRACT

Four new complete mitochondrial genomes (mitogenomes) from the two superfamilies Ocypodoidea and Grapsoidea were sequenced, which represented *Uca* (*Gelasimus*) *borealis* (Ocypodidae: Ucinae), *Dotilla wichmani* (Dotillidae), *Metopograpsus quadridentatus* (Grapsidae: Grapsinae), and *Gaetice depressus* (Varunidae: Gaeticinae). All of the mitogenomes shared the complete set of 37 mitochondrial genes. Mitogenome lengths were 15,659, 15,600, 15,517, and 16,288 bp, respectively, with A + T contents of 69.41%, 68.46%, 70.30%, and 72.96%, respectively. Comparative genomic analyses suggested that they exhibited different genomic rearrangements. In particular, *G. depressus* shared a major rearrangement pattern present in *Eriocheir* crabs, while the remainder shared the brachyuran ground genomic rearrangement patterns. Phylomitogenomic inferences provided new evidence for the strongly supported nesting of Thoracotremata within Heterotremata clades. A close phylogenetic relationship was observed between Varunidae and Macrophthalmidae crabs, and between Dotillidae and Grapsidae crabs, which was consistent with mitochondrial genomic rearrangement similarities. Altogether, these results suggest the presence of reciprocal paraphyly for Ocypodoidea and Grapsoidea.

1. Introduction

Brachyuran crabs are one of the most species-rich clades among extant crustaceans, with over 7250 described species (Davie et al., 2015b), and they are economically important. Guinot (1977) defined two major groups of Brachyurans based on plesiomorphic conditions of gonopore positions: Heterotremata Guinot, 1977, and Thoracotremata Guinot, 1977. The two groups are collectively referred to as the Eubrachyura section (Guinot, 1977; de Saint-Laurent, 1980; Ahyong et al., 2007; Tsang et al., 2014), and have been interpreted as independent sister groups within the monophyletic Eubrachyura (Guinot et al., 2013; Tsang et al., 2008, 2014). However, controversies still exist involving the monophyly of both subsections in their current compositions, and the interrelationships of these groups. Increasing studies have proposed that Thoracotremata nested within Heterotremata (Von Sternberg et al., 1999; Von Sternberg and Cumberlidge, 2001a, 2001b; Brösing et al., 2007; Ahyong et al., 2007; Ji et al., 2014; Xing et al., 2016, 2017). Further, within the Thoracotremata, the monophyly of the Ocypodoidea Rafinesque, 1815, and Grapsoidea MacLeay, 1838 superfamilies in their current compositions remain controversial (reviewed in Davie et al., 2015a). Martin and Davis (2001) recognized four valid families within Ocypodoidea, including the Camptandriidae, Mictyridae, Palicidae, and Ocypodidae. Subsequently, the other four subfamilies (Dotillinae, Heloeciinae, Macrophthalminae, and Ucidinae) that were previously assigned to Ocypodidae have been reassigned to the family level, and Xenophthalmidae was reassigned to the superfamily, and Palicidae was excluded (Ng et al., 2008). Therefore, the Ocypodoidea provisionally retained eight families (Ng et al., 2008; Davie et al., 2015b). The Grapsoidea is traditionally recognized as monophyletic, with its current composition including eight families: Gecarcinidae, Glyptograpsidae, Grapsidae, Plagusiidae, Sesarmidae, Varunidae, Xenograpsidae, and Percnidae (Martin and Davis, 2001; Ng et al., 2008; Davie et al., 2015b). However, molecular phylogenetic studies have indicated that both Ocypodoidea and Grapsoidea (O & G) are not monophyletic in their current compositions (Schubart et al., 2006). Many molecular work based on partial mtDNA sequences showed that the two superfamilies are polyphyletic (Kitaura et al., 2002, 2010; Tsang et al., 2014; Davie et al., 2015a).

Abbreviations: mt, mitochondrial; O & G, Ocypodoidea and Grapsoidea; IGNs, non-coding intergenic nucleotides; mNCR, main non-coding region; AA(or aa), amino acid(s); PCG, protein-coding gene(s); rRNA, ribosomal RNA; M & V, Macrophthalmidae and Varunidae; LBA, long-branch attraction

* Corresponding author.
E-mail address: jqchen1102@126.com (J. Chen).

https://doi.org/10.1016/j.gene.2018.06.088
Received 16 April 2018; Received in revised form 1 June 2018; Accepted 25 June 2018
Available online 27 June 2018
0378-1119/

Mitochondrial (mt) genomes contain rich evolutionary signals in both sequence characters and in genomic arrangements (Boore, 1999; Sun et al., 2003, 2005; Tan et al., 2017). Genomic rearrangements that may act as synapomorphies for phylogenetic estimation have been identified in decapod crustaceans (Shen et al., 2015; Tan et al., 2015, 2017; Xing et al., 2017). Further, phylomitogenomic data has been used to generate a robust molecular phylogeny for eubrachyurans (Ji et al., 2014). For example, recent phylogenetic inference that was based on mt genomes suggested non-monophyly of O & G, with moderate to low support (Basso et al., 2017). However, inadequate taxon sampling and phylogenetic data usage may mislead tree reconstruction for decapod crustaceans (Qian et al., 2011; Xing et al., 2017). Consequently, the phylogenetic relationships of O & G that are inferred from mitogenomes require reanalysis by integrating more taxa from divergent lineages of both superfamilies.

In this study, we enhanced taxonomic coverage of brachyurans by sequencing four new complete mt genomes from different families and subfamilies including *Uca* (*Gelasimus*) *borealis* Crane, 1975 (Ocypodidae: Ucinae), *Dotilla wichmani* De Man, 1892 (Dotillidae), *Metopograpsus quadridentatus* Stimpson, 1858 (Grapsidae: Grapsinae), and *Gaetice depressus* (De Haan, 1835) (Varunidae: Gaeticinae). The overall goal of this research was to sequence and analyze the complete mitochondrial genomes from the four above taxa in order to resolve the phylogenetic position of the Thoracotremata, the phylogenetic relationships among O & G by combining with genomic rearrangements.

2. Materials and methods

2.1. Sample collection and identification

U. (*G.*) *borealis*, *D. wichmani*, *M. quadridentatus*, and *G. depressus* were collected by hand from Beihai in the Guangxi Province, China. Specimens were stored in 95% ethanol and deposited at the Institute of Aquatic Biology, School of Life Sciences, Chemistry & Chemical Engineering, Jiangsu Second Normal University, Nanjing, China. Identification was performed morphologically using morphological information in *Crabs of China Seas* (Dai and Yang, 1991), with a stereo dissection microscope. Morphological identifications were verified using the 16S mtDNA and cytochrome oxidase subunit I (COI) gene molecular markers.

2.2. Genomic DNA extraction and sequencing

Total genomic DNA was extracted from the samples using a Cell and Tissue DNA Extraction Kit (Generay Biotech, Shanghai, China). The DNA was then sent to Novogene, Beijing, China for sequencing on the Illumina sequencing platform using 150-bp paired-end runs (2 × 150). The average insert size of sequencing libraries was approximately 300 bp.

2.3. Mt genome assembly and annotation

Sequencing resulting in a total of 14,867,477 (*U.* (*G.*) *borealis*) 21,127,653 (*D. wichmani*), 18,613,326 (*M. quadridentatus*), and 19,519,420 (*G. depressus*) paired sequences. Mt genome assemblies were conducted in Geneious 10.1.2 using parameter settings described previously (Kearse et al., 2012; Xing et al., 2017). Mitochondrial contigs were then extracted as single sequences. Protein-coding genes (PCGs) and ribosomal RNA genes (rRNAs) were annotated by alignment of homologous genes from other published thoracotreme mt genomes. The PCG coding regions, initiation codons, and termination codons were identified using the NCBI ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder). Transfer RNA (tRNA) genes were mainly identified by tRNAscan-SE (Lowe and Chan, 2016). The remaining unidentified tRNA genes were annotated according to their tRNA-like secondary structures, anticodon sequences, and alignment with homologous tRNA genes. Finally, complete mitochondrial genome sequences were further confirmed with the MITOS web server, using the invertebrate genetic code for mitochondria (Bernt et al., 2013).

2.4. Mt genomic characterization and rearrangement analyses

Nucleotide composition was calculated in MEGA 6.0 (Tamura et al., 2013). Tandem repeats in the larger, main non-coding regions were predicted using the Tandem Repeats Finder program (http://tandem.bu.edu/trf/trf.basic.submit.html; Benson, 1999). Hierarchical cluster analysis was performed in Spass 22.0 using intergenic spacer data distributions. Genomic rearrangement analyses were performed using the CREx program (http://pacosy.informatik.uni-leipzig.de/crex/form) and common intervals and breakpoints (Bernt et al., 2007).

2.5. Phylogenetic analysis

Phylogenies were reconstructed using data from the four newly sequenced mt genomes, in addition to brachyuran mitogenomes that were available and belonged to 59 species, 24 families, and 14 superfamilies (taxa are listed in Table S1). After removing all termination codons, the putative amino acid (AA) sequences of the 13 mt PCGs, and sequences for two rRNAs, were individually aligned using MAFFT 7.215 (Katoh and Standley, 2013). Ambiguous or randomly similar sites from each gene were removed in Aliscore 2.0 (Misof and Misof, 2009; Kuck et al., 2010) using default settings. The pruned, aligned AA sequences were then used as a backbone to align the corresponding nucleotide (NT) sequences using DAMBE 5.3.15 (Xia and Xie, 2001). The final alignment consisted of a concatenation of the 13 PCGs and two rRNAs.

PartitionFinder 1.1.1 (Lanfear et al., 2012) was used to determine the best-fit partitioning schemes and models using a greedy search with Bayesian information criterion (BIC). Maximum likelihood (ML) analysis was performed using RAxML-HPC2 on XSEDE 8.0.0 (Stamatakis, 2014) through the Cipres Science Gateway (Miller et al., 2010). Given the advantages of a CAT versus Gamma distribution when there are more than 50 sequences (Stamatakis, 2006), the nucleotide substitution model comprising GTR + CAT, and the best-fit partitioning schemes identified by PartitionFinder, were used. Nodal support was evaluated using 1000 bootstrap replicates in RAxML. Bayesian inference (BI) analysis was performed using MrBayes 3.2.2 (Huelsenbeck and Ronquist, 2001), where the best partitioning scheme and the best-fit nucleotide substitution model for each partition, as calculated by PartitionFinder, were used. The analysis was started with one million generations with every 1000 generations being sampled, and the first 25% of the generations were discarded as burn-in. The standard deviation of split frequencies was 0.0036 (below 0.01) after 1,000,000 generations, reflecting that two runs were very good indication of convergence.

3. Results and discussion

3.1. Mt genomic characterization

We successfully obtained complete mt genomes from the four species, *U.* (*G.*) *borealis*, *D. wichmani*, *M. quadridentatus*, and *G. depressus*, which were 15,659 bp, 15,600 bp, 15,517 bp and 16,288 bp in length, respectively. The genomes have been deposited under the following GenBank accession numbers: MH183126, MH183129, MH183127 and MH183128 (Table 1). The mitogenome of *G. depressus* was the largest among the four species, owing to the highest observed size in total intergenic nucleotides (IGNs: 577 bp) and main non-coding region (mNCR: 897 bp). The overall A + T contents varied from 69.41% (*U.* (*G.*) *borealis*) to 72.96% (*G. depressus*; Table 2). AT bias is normal in brachyuran mt genomes, which have been reported to range from 64.85% to 77.32% (Sun et al., 2005; Bai et al., 2016). The A + T contents of the 13 PCGs were calculated from the coding strand of each