Gene 548 (2014) 263–269

Contents lists available at ScienceDirect

Gene

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

Complete mitochondrial genome of Malaysian Mahseer (*Tor tambroides*)

M.Y. Norfatimah [a,b,*], L.K. Teh [c], M.Z. Salleh [c], M.N. Mat Isa [d], M.N. SitiAzizah [b,e]

[a] Faculty of Applied Sciences, Universiti Teknologi MARA, Shah Alam 40450, Malaysia
[b] School of Biological Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia
[c] Pharmacogenomics Centre, Faculty of Pharmacy, Universiti Teknologi MARA, 42300 Puncak Alam, Malaysia
[d] Malaysia Genome Institute, Kajang, 43000, Malaysia
[e] Centre for Marine and Coastal Studies, Penang 11800, Malaysia

ARTICLE INFO

Article history:
Received 7 January 2014
Received in revised form 8 July 2014
Accepted 14 July 2014
Available online 15 July 2014

Keywords:
NGS
Tor tambroides
Mitochondrial genome

ABSTRACT

This is the first documentation of the complete mitochondrial genome sequence of the Malaysian Mahseer, *Tor tambroides*. The 16,690 bp mitogenome with GenBank accession number JX444718 contains 13 protein genes, 22 tRNAs, two rRNAs, and a noncoding control region (D-loop) as is typical of most vertebrates. The phylogenomic reconstruction of this newly generated data with 21 Cypriniformes GenBank accession ID concurs with the recognized status of *T. tambroides* within the subfamily Cyprininae. This is in agreement with previous hypotheses based on morphological and partial mitochondrial analyses.

1. Introduction

The Malaysian Mahseer, *Tor* spp. or locally known as kelah is an ornamental fish with high commercial value in Malaysia as well as in many parts of Southeast Asia. As a consequence, their populations are depleting and highly threatened due to commercial exploitation. In Malaysia, it is widely presumed that there are three species collectively known as Malaysian Mahseer present, namely *Tor tambroiders*, *Tor tambra* and *Tor dourenensis* although there still remains some uncertainty (Esa et al., 2011; Nguyen et al., 2008; Siraj et al., 2007). Considering their high economic value, we believe that genomic data may provide essential information to preserve species genetic resources in order to support an adequate management program for wild and culture populations as well as in systematics study.

Mahseers in the genus *Tor* are defined as carps with big scales, fleshy lips, continuous at the angles of the mouth with an interrupted fold or groove across the lower jaw, two pairs of big barbels, lateral-line scales ranging from 22 to 28, and length of head equal to or greater or less than the depth of body (Desai, 2003). They have been placed in the subfamily Cyprininae by various authors (Rainboth, 1996; Roberts and Khaironizam, 2008). This subfamily is defined by various researchers based on its spinous anal-fin ray (or anal spine), serrated/non serrated spine and occasionally the number of branched dorsal fin ray (Yang et al., 2010). According to Roberts and Khaironizam (2008), *T. tambroides* first described from Java was grouped under *T. tambra* subspecies. Based on phenotypic features, *T. tambroides* is distinctive from other *Tor* spp. by the presence of a median lobe, large scales with dark vertical bands especially on the scales above the lateral line (Poh, 2006; Roberts and Khaironizam, 2008). However, several authors have also suggested variations of two different colors (reddish and silver–bronze) as a method for *T. tambroides* recognition. On the other hand, other workers believe that these characteristics are under environmental influences (Esa et al., 2006; Siraj et al., 2007; Vrijenhoek, 1998).

Although *T. tambroiders* is one of the most economically valuable fish in Malaysia, relatively very few scientific literatures are available. The taxonomic status of the Malaysian Mahseer, *T. tambroides* species within the *Tor* complex has still remained in conflict despite numerous related studies conducted since it was first recognized by Bleeker in 1854. This is largely in part due to its morphological variations and plasticity in response to habitat differences (Roberts and Khaironizam, 2008; Wimberger, 1994). Chronologically, the study on the taxonomy of the genus *Tor* began in 1854 by Bleeker and continued with Rainboth (1996) who classified *T. tambroides* under family Cyprinidae, subfamily of Cyprininae, tribe Cyprinini and subtribe of Tores. These results were based on traditional morphological, morphometric and meristic data. Later, various independent studies came to a parallel conclusion on the taxonomic status of *Tor*; family Cyprinidae, subfamily Cyprininae, tribe Barbini (interchangeably referred to as subfamily Barbinae by various authors) but offered no information on the subtribe status (Cavender and Coburn, 1992; Gilles et al., 1998; Mayden et al., 2009;

Abbreviations: ATP 6 and 8, adenosine triphosphate synthases 6 and 8; bp, base pairs; COI–III, cytochrome *c* oxidase subunits I–III; cyt *b*, cytochrome *b*; D-loop, displacement loop; mtDNA, mitochondrial DNA; mitogenome, mitochondrial genome; ND1–6 and 4L, NADH dehydrogenase subunits 1–6 and 4L; NGS, next generation sequencing; rRNA, ribosomal RNA; tRNA, transfer RNA.

* Corresponding author at: Faculty of Applied Sciences, Universiti Teknologi MARA, Shah Alam 40450, Malaysia.
E-mail address: norfatimah9778@salam.uitm.edu.my (M.Y. Norfatimah).

http://dx.doi.org/10.1016/j.gene.2014.07.044
0378-1119/

Nguyen and Ngo, 2001; Thai et al., 2007; Wang et al., 2007) based on morphological and molecular studies. More recent and extensive studies (Yang et al., 2010, 2012) based on five partial mitochondrial genes, cyt *b* and 16S rRNA, CO1, ND4 and ND5 respectively also resolved a monophyletic Cyprininae and recovered seven major clades from this subfamily. Their study limited the tribe Cyprinini sensu stricto to four genera, *Cyprinus*, *Carassius*, *Carassioides* and *Procypris*. *Tor* and its allied genus *Neolissochilus* in both studies formed a monophyletic assemblage within the tribe Barbini. The authors noted however that the Barbini was not a natural grouping. Other authors (Miya et al., 2001) have suggested that the phylogenetic analyses of mitogenomes may have a better resolving efficiency in persistent controversies over higher-level relationships of teleosts. In a contribution to the systematics and phylogenetics of this group, here we describe the mitogenome sequence of the Malaysian Mahseer, *T. tambroides*. Our findings would provide a framework for further mitogenomic investigations of other species within this genus towards a better taxonomic understanding of the group. Herein we attempt to fill the knowledge gap by sequencing the whole mitochondrial genome data to complement previous studies (Saitoh et al., 2006; Wu et al., 2010; Yang et al., 2010). Furthermore, our study will provide genomic data for prediction of functional genes and biomarkers. Such data would be immensely beneficial for an aquaculture program of this threatened and highly valued species.

The main aim of this study was to decipher and establish *T. tambroides* mitogenomic data. This data is an invaluable contribution of the Malaysian Mahseer for the elucidation of the taxonomic status, phylogenomic reconstruction, as well as conservation and management of wild and cultured populations of the species.

2. Materials and methods

2.1. Sample and DNA extraction

A caudal fin sample of a single *T. tambroides* individual was obtained from Keniam River, at the National Park in Pahang, Malaysia with coordinates of latitude: 4°34′15.27″N and longitude: 102°27′57.39″E. The sample was stored in 96% ethanol for preservation. Total genomic DNA was isolated using a Tissue DNA Extraction Kit (Invitek GmbH, Berlin, Germany) and produced in replicates of 5 μg in 50 μl concentrations and ratios of 1.8–2.0 at OD A260/280.

2.2. Next generation sequencing and sequence quality control

The DNA extract was then prepared using Illumina Truseq kit and sequenced by Illumina Genome Analyzer IIX with 2 × 100 bp read length (Paired-End) in a single lane to obtain a high amount of data. Prior to assembly, the sequence reads were subjected to quality control procedures to eliminate low-quality bases. The quality check was based on the quality value determination. Trimming of the reads for base quality assessment and the sequence length (minimum size 50 bp) was done using CLC Genomics Workbench 4.9, and Dynamic Trim (Cox et al., 2010). Quality filtering was reviewed using the FastQC tool. The quality value generated by the Consensus Assessment of Sequence and Variation (CASAVA 1.8) software was based on the conventional Sanger/Phred quality values.

2.3. Assembly and gene annotation

Two assembly approaches namely, Velvet 1.2.07 (linux based platform) (Zerbino and Birney, 2008) and CLC Genomics Workbench 4.9 (automated commercial software) were used and results were compared. BLAST searches against the *T. tambroides* mitochondrial genome draft were performed with six Cyprinidae complete mitogenome references namely; *Cyprinus carpio* (NC_001606), *Labeo bata* (NC_015193), *Procypris rabaudi* (EU082030), *Barbus barbus* (AB238965), *Sinocyclocheilus grahami* (GQ148557.1) and *Sinocyclocheilus altishoulderus* (FJ984568). These mitogenomes were used for reference mapping assembly and also to confirm manual annotations. In order to confirm the precise sequence of specific regions with low coverage and to produce a high quality complete mitogenome data, six sets of primer were designed using Primer Premier version 6.0. All the primers were used to amplify and genome finishing of *T. tambroides* mitochondrial genome (data not shown here). Annotation was conducted using the Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al., 2004) and Artemis (Rutherford et al., 2000) and Artemis comparison tool (ACT) Software (Carver et al., 2005).

2.4. Phylogenomic analysis

The newly generated *T. tambroides* mitogenome together with previously reported mitogenome sequence of 20 Cyprinidae species from GenBank were used to conduct a phylogenomic analysis. Sequences included in this study with their accession numbers and references are given in Table 1. Sequence alignment was conducted using CLUSTAL W. A model test was conducted using Model test 3.06 PPC (Posada and Crandall, 1998) to identify the best model of evolution for each gene region and the combined dataset for the phylogenomic construction. The model with the best Maximum Likelihood score using the Akaike Information Criterion (AIC) was chosen (Akaike, 1973) as it was best suited for the combined dataset for Maximum Likelihood and single model Bayesian approaches (Fig. 3). Models for individual gene region were used in the mixed model Bayesian analyses (Ronquist et al., 2011, 2012). The phylogenomic analysis involving 22 nucleotide sequences (20 Cyprinidae and two outgroups, *Danio rerio* and *Esomus metallicus*) was implemented using MEGA 5.0 (Tamura et al., 2011). Codon positions included were 1st, 2nd and 3rd coding and noncoding regions. All positions containing gaps and missing data were eliminated. There were a total of 15,435 positions in the final dataset. Initial tree(s) for the heuristic search in Maximum Likelihood analysis were obtained in the default output as follows. When the number of common sites was <100 or less than one fourth of the total number of sites, the Maximum Parsimony (MP) method was used; otherwise Neighbor Joining (BIONJ) method with Maximum Composite Likelihood (MCL) distance matrix was used. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.6056), Fig. 4).

3. Results and discussion

3.1. NGS data analysis of Tor tambroides mitochondrial genome

The total number of reads after trimming using both VELVET and CLC was 75 million from the total expected reads of 80 million with the average length of 94.85 bp and total number of bases of 7.1 Gb. Results showed that only 2.2 Mb and 1.6 Mb were mapped to the reference mitogenome of *C. carpio* by CLC and Velvet, respectively. Total reference coverage between both approaches differed with a mean of 134× compared to 95× for CLC and Velvet, respectively. Table 2 shows the representative reference assembly with *C. carpio*. Both pipelines (automated and Linux based) successfully generated a draft of *T. tambroides* with different views and information. The combination of Velvet and CLC Genomics Workbench assembled the Illumina reads effectively. However, the CLC Genomics Workbench provided an integrated bioinformatics environment and powerful tool as compared to Velvet.

3.2. Tor tambroides mitochondrial genome organization

The complete mitochondrial genome sequence of *T. tambroides* was determined to be 16,690 bases in length and has been deposited in GenBank (accession no. JX444718). As shown in Fig. 1, the organization of mitochondrial genome of *T. tambroides* is similar to that of the typical vertebrate mitochondrial genome, consisting of 13 protein-coding