Gene xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Gene

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

Short communication

The complete mitochondrial genome of threatened chocolate mahseer (*Neolissochilus hexagonolepis*) and its phylogeny

Prabhati Kumari Sahoo [a], Chirag Goel [a], Rohit Kumar [a], Nisha Dhama [a], Shahnawaz Ali [a], Dandadhar Sarma [b], Prasanta Nanda [c], Ashoktaru Barat [a,*]

[a] *ICAR—Directorate of Coldwater Fisheries Research, Bhimtal, 263136 Nainital, Uttarakhand, India*
[b] *Gauhati University, Guwahati, 781014 Assam, India*
[c] *Dera Natung Government College, Itanagar, 791111 Arunachal Pradesh, India*

ARTICLE INFO

Article history:
Received 23 May 2015
Received in revised form 1 July 2015
Accepted 9 July 2015
Available online xxxx

Keywords:
Neolissochilus hexagonolepis
Complete mitochondrial genome
Mahseer group fishes
Phylogenetic status

ABSTRACT

The chocolate mahseer (*Neolissochilus hexagonolepis*) is an important food and game fish of North Eastern India. To study the phylogenetic status we sequenced the complete mitochondrial genome of *N. hexagonolepis*. The mitogenome is 16,563 bp in length and composed of 13 protein coding genes, 22 tRNAs, 2 rRNAs and one putative control region. The overall base composition was A 31.8%, T 25.0%, G 15.8%, C 27.4% and A + T content 56.9%, G + C content 43.1%. The phylogenetic analysis using the complete mitochondrial genome revealed that the chocolate mahseer belonged to same clade of mahseer group of fishes but different from genera *Barbus* and *Acrossocheilus*. The present study will be helpful for the evolution and conservation genetic studies of *N. hexagonolepis*.

1. Introduction

Neolissochilus hexagonolepis, a cyprinid fish is most commonly known as "the chocolate mahseer" and is endemic to North East states of India, particularly to the Brahmaputra river basin. It is a very popular fish for both food and sport and also a rich protein source for remote villagers of Arunachal Pradesh, Assam and Meghalaya. But it is reported to be declining in the wild and presently designated as nearly threatened species (IUCN, 2014). Therefore, its propagation by artificial breeding was attempted as a measure of conservation by several research and educational Institutions (Mahapatra and Vinod, 2011; Sarma et al., 2015). It is also being tried to study the genetic variability of different populations with the aim of future genetic improvement program. But usable genetic markers for population as well as species characterization are lacking in this species. Furthermore, its nomenclature and taxonomic position was viewed by several authors in different genera (Dinesh et al., 2010; Eschmeyer and Fong, 2014) despite its morphological distinction to other mahseer species (Laskar et al., 2013). *N. hexagonolepis* (chocolate mahseer) was initially included as the member of the genus *Barbus* (McClelland, 1839) but again reclassified in the genus *Lissochilus* (Weber & de-Beaufort, 1916). Later on its taxonomic position was proposed in the genus *Acrossocheilus* (Oshima, 1919).

Recently, studies on complete mitochondrial genome sequences have been increased many fold due to its smaller size, conserved gene content and organization in many teleostean fishes (Fischer et al., 2013; Yue et al., 2006). Hence, we studied the complete mitochondrial genome organization for the first time in *N. hexagonolepis* and also tried to draw phylogenetic relationship with other members of family Cyprinidae (including mahseer group of fishes) using their respective complete mitochondrial genome sequences. This data will help in defining proper taxonomic position as well as in the conservation and evolution studies of this species. In addition, our study will provide genomic data for the prediction of functional genes and biomarkers.

2. Materials and methods

2.1. Sample collection and DNA extraction

Caudal fin tissue samples of *N. hexagonolepis* were collected from the River Jia Bhoreli (27°02′N; 92°35′E) near Bhalukpong (Assam–Arunachal Pradesh), India and were preserved in 90% ethanol. Voucher specimens

Abbreviations: atp6, ATP synthase F0 subunit 6; atp8, ATP synthase F0 subunit 8; cyb, cytochrome b; coxI, cytochrome c oxidase subunit 1; coxII, cytochrome c oxidase subunit 2; coxIII, cytochromec oxidase subunit 3; nad1, NADH dehydrogenase subunit 1; nad2, NADHdehydrogenase subunit 2; nad3, NADH dehydrogenase subunit 3; nad4, NADH dehydrogenase subunit 4; nad4l, NADH dehydrogenase subunit 4 l; nad5, NADH dehydrogenase subunit 5; nad6, NADH dehydrogenase subunit 6; Phe, phenylalanine; rrnL, mitochondrial large subunit ribosomal RNA gene; rrnS, mitochondrial small subunit ribosomal RNA gene; PCGs, protein coding genes; IUCN, International Union for Conservation of Nature; mtDNA, mitochondrial DNA; UPGMA, Unweighted Pair Group Method with Arithmetic Mean; TAS, termination associated sequence; CSB, conserved sequence blocks; OH, origin of heavy strand replication; NCBI, National Center for Biotechnology Information.

* Corresponding author. Tel.: +91 5942 247280.
E-mail address: abarat58@hotmail.com (A. Barat).

http://dx.doi.org/10.1016/j.gene.2015.07.024
0378-1119/

were deposited at a reference laboratory (corresponding author's address). Total genomic DNA was extracted from fin tissue using Phenol:Chloroform:Iso-amyl alcohol method (Sambrook and Russell, 2001).

2.2. PCR amplifications and DNA sequencing

The mtDNA was selectively amplified from the genomic DNA using REPLI-g Mitochondrial DNA Kit (Qiagen, Venlo, NL). Complete mtDNA was amplified as concatenated sequences using selectively amplified mtDNA template and seventeen primer pairs derived from the literature (Cheng et al., 2012; Mabuchi et al., 2006; Wang et al., 2000). The PCR was performed in a total reaction volume of 50 μl containing 1× PCR Buffer, 200 nM of each primer, 400 μM dNTP, 1 μl template (as per kit) and 1 U Taq DNA Polymerase (Invitrogen, Waltham, USA). The reaction conditions were initial denaturation at 94 °C for 4 min; 35 cycles of denaturation at 94 °C (45 s), annealing at 48 °C–50 °C (45 s) and extension at 72 °C (90 s) followed by a final extension at 72 °C for 10 min. Annealing temperatures were varied within the above-listed ranges in order to optimize the efficiency of different primers. Contiguous, overlapping fragments of the entire mitochondrial genome was sequenced using BigDye Terminator cycle sequencing chemistry v.3.1 in ABI 3730xl automatic sequencer (Life Technologies, Carlsbad, CA).

2.3. Assembly and annotation of the mitochondrial genome

Chromatographs were proofread and assembled into contigs using CLC Genomics Workbench v.7.5.2 (CLC Bio, Aarhus, Denmark). Contigs so obtained were assembled on the reference *Tor putitora* (NC_021755) complete mitochondrial genome. In addition, CLC Genomics workbench v.7.5.2 was used for the prediction of open reading frame (ORFs), codon usage and for the preparation of graphical representation of the annotated mitochondrial genome. The ORF finder tool of the workbench was set to find ATG or GTG as start codons and non-triplet 3′ ends of the protein-coding genes immediately adjacent to the beginning of downstream features and reading T or TA as truncated stop codons. Transfer RNA analysis was conducted using tRNAscan-SE v.1.21 (Schattner et al., 2005) with a cut-off score of 0.1. Origin of replication of the light strand (O_L) and the heavy strand (O_H) and conserved blocks in the non-coding control region were identified by multiple sequence alignments with annotated sequences of other species as well as by secondary structure search. The ribosomal RNA secondary structure was predicted using mfold web server (Zuker, 2003). The mtDNA sequence of *N. hexagonolepis* was deposited in NCBI GenBank under accession number NC_026106.1 (Fig. 1).

2.4. Phylogeny construction

The phylogenetic tree (Fig. 2) was constructed using mitogenome sequences retrieved from NCBI GenBank. The study was aimed to delineate the relationship of chocolate mahseer with other mahseer as well as to validate its taxonomic position. The sequences included in the present analysis were; mahseer (*Tor tor*, *T. putitora and Tor sinensis*); genus *Acrossocheilus* (*Acrossocheilus monticola*, *Aphanius fasciatus*, *Acrossocheilus wenchowensis*, *Acrossocheilus kreyenbergii*, *Acrossocheilus hemispinus*, *Acrossocheilus stenotaeniatus*, *Acrossocheilus barbodon*); genus *Puntius* (*Puntius semifasciolatus and Puntius snyderi*); genus *Cyprinus* (*Cyprinus carpio* 'color', *Cyprinus carpio carpio*, *Cyprinus carpio xingguonensis and Cyprinus carpio wuyuanensis*), *Barbus barbus* and *Danio rerio* as an out group. The 12 concatenated protein coding genes (excluding *ND6*) encoded on heavy strand were aligned using CLC Genomics v.7.5.2 with default parameters. Best fitted model was selected based on hierarchical likelihood ratio test (hLRT) and Bayesian information criteria (BIC) parameter and finally general time reversible (GTR) model was selected for phylogeny construction. UPGMA dendrogram was constructed based on maximum likelihood phylogeny (1000 replicates; maximum log likelihood = −69452.92).

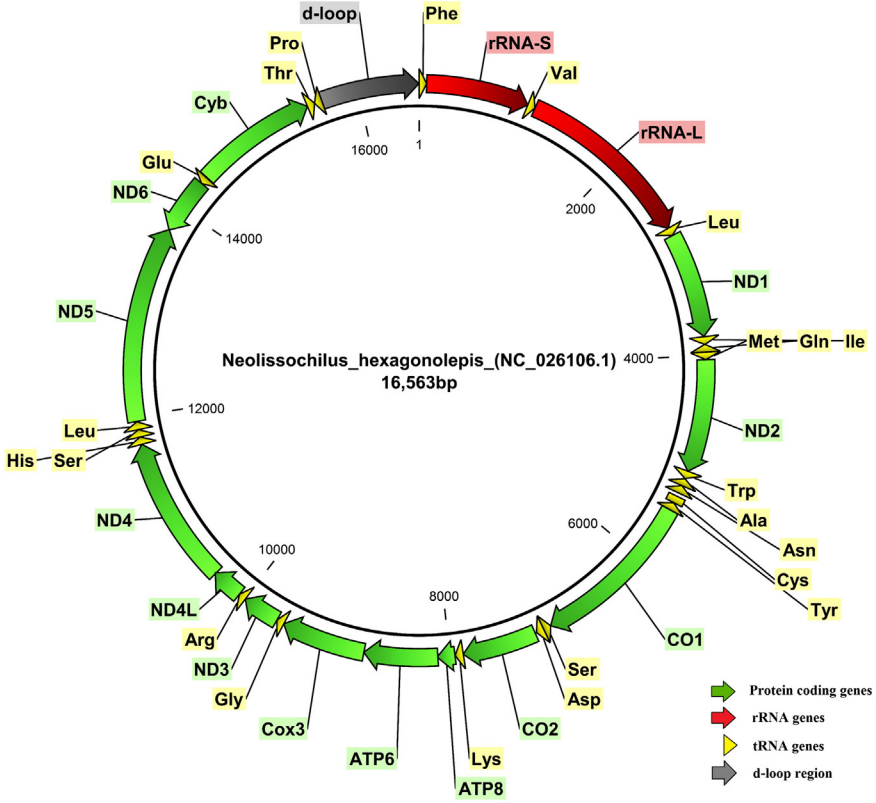

Fig. 1. Complete mitochondrial genome organization and gene arrangement of chocolate mahseer.

Please cite this article as: Sahoo, P.K., et al., The complete mitochondrial genome of threatened chocolate mahseer (*Neolissochilus hexagonolepis*) and its phylogeny, Gene (2015), http://dx.doi.org/10.1016/j.gene.2015.07.024