

## The complete mitochondrial genome of *Nematobrycon palmeri* (Characiformes:Nematobrycon) and phylogenetic studies of Characidae

Qi Wang<sup>a,b</sup>, Tao Zhang<sup>c</sup>, Xiaolong Yin<sup>d</sup>, Fang Meng<sup>a,b</sup>, Youkun Huang<sup>a,b</sup>, Bingjian Liu<sup>a,b</sup> and Yifan Liu<sup>a,b</sup>

<sup>a</sup>National Engineering Research Center for Marine Aquaculture, Zhejiang Ocean University, Zhoushan, China; <sup>b</sup>National Engineering Laboratory of Marine Germplasm Resources Exploration and Utilization, Marine Science and Technology College, Zhejiang Ocean University, Zhoushan, China; <sup>c</sup>Zhejiang Province Key Lab of Mariculture and Enhancement, Marine Fisheries Research Institute of Zhejiang, Zhoushan, China; <sup>d</sup>Zhoushan Fisheries Research Institute of Zhejiang Province, Zhoushan, China

### ABSTRACT

Complete mitochondrial genome of the Characiform fish *Nematobrycon palmeri* was characterized in the present study. The whole mitogenome was 17,340 bp in size and the proportion of coding sequences with a total length of 11,448 bp was 66.02%, which encodes 3805 amino acids. The base composition of the genome was 30.92% for A, 23.92% for C, 14.88% for G, and 30.28% for T. All protein-coding genes were started with ATG, CO1 and ATP8 ended by AGG, TAG respectively, whereas CO2, ATP6, ND4 ended by a single T, the other PCGs commonly ended by TAA. The length of 12S and 16S ribosomal RNA was 949 bp and 1675 bp, respectively. The control region (D-loop) ranging from 15,654 bp to 17,340 bp was 1687 bp in size. It showed negative GC skew value (−0.2329) and positive AT skewness (0.0105). Phylogenetic analysis showed that *N. palmeri* was most closely related to *Gephyrocharax atracaudatus*. The complete mitochondrial genome sequence would provide a new insight into taxonomic classification, and help to draw a more complete picture of species diversity within the Characidae.

### ARTICLE HISTORY

Received 13 July 2020  
Accepted 14 September 2020

### KEYWORDS

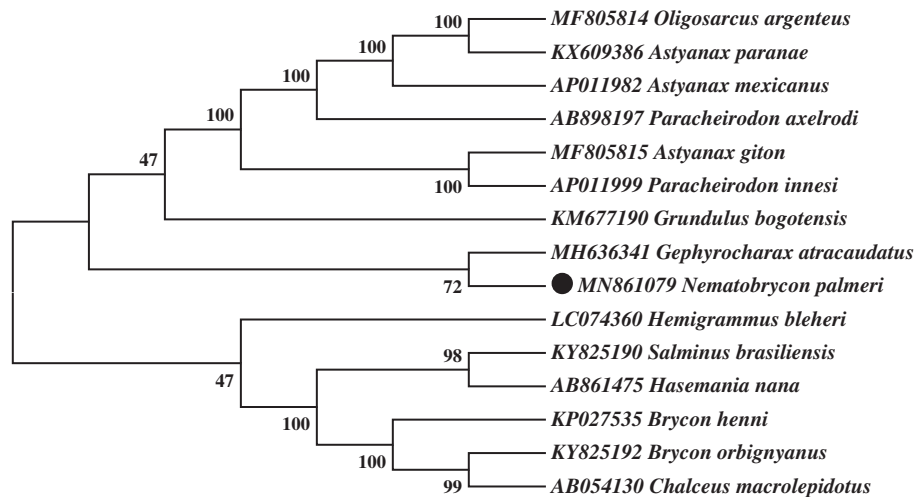
*Nematobrycon palmeri*;  
mitochondrial genome;  
evolutionary relationships

*Nematobrycon palmeri* belongs to the family Characidae and the order Characiformes, a species of characid fish found in the Atrato and San Juan river basins in western Colombia. Though diverse species based on anatomical diversity have been added to each genus, but there is hardly any sequences on the *Nematobrycon*. We are at loss here to assess properly the status of species of *Nematobrycon* relevant to our review because their descriptions are incomplete and/or misleading. In this study, we determined the complete mitochondrial genome of *N. palmeri*, which would supplement the limited data on molecular level and even be reference for systematics.

Specimens of *N. palmeri* were collected from San Juan river basin of Colombia (4°02'42"N, 77°26'29"W) and stored in a refrigerator of −80 °C in Zhejiang Engineering Research Center for Mariculture and Fishery Enhancement Museum (Accession number: PH181922). Total genomic DNA was extracted from muscle of three different individuals using the phenol–chloroform method (Barnett and Larson 2012). The calculation of base composition and phylogenetic construction were conducted by MEGA6.0 software (Tamura et al. 2013). The transfer RNA (tRNA) genes were generated with the program tRNAscan-SE (Lowe and Eddy 1997). The mitochondrial genome sequence of *N. palmeri* was sequenced by Sanger dideoxy and assembled by CodonCode Aligner 5.1.5 (CodonCode Corporation, Dedham, MA), the annotated genes was deposited in GenBank with the accession number MN861079.

Similar to the typical mitogenome of vertebrates, the mitogenome of *N. palmeri* is a closed double-stranded circular molecule of 17,340 bp including 13 protein-coding genes, 2 ribosomal RNA genes, 22 tRNA genes and two main non-coding regions (Boore 1999; Zhu et al. 2018). The contents of A, C, G, and T are 30.92%, 23.92%, 14.88%, and 30.28%, respectively. Most mitochondrial genes are encoded on the H-strand except for ND6 and eight tRNA genes (Gln, Ala, Asn, Cys, Tyr, Ser, Glu, and Pro), which are encoded on the L-strand. The proportion of coding sequences with a total length of 11,448 bp is 66.02%, 13 protein-coding genes (PCGs) encode 3805 amino acids in total. A-T and G-C contents of mitochondrial genome were 61.21% and 38.79%, respectively, thereby with a high AT bias. Besides, it showed negative GC skew value (−0.2329), indicating that C base was more common than G base, whereas AT skewness was positive (0.0105), suggesting A base occurs more frequently than T base in the *N. palmeri* mitochondrial genome.

All the protein-coding genes use the initiation codon ATG, which is quite common in vertebrate mtDNA (Miya et al. 2001; Liu et al. 2017). ND1, ND2, CO3, ND3, ND4L, ND5, ND6 and CytB end by TAA as a stop codon, CO1 and ATP8 end by AGG, TAG respectively, and three incomplete termination codons (T) were found in the other genes (CO2, ATP6, and ND4). The content of Cys in 13 PCGs was the lowest accounting for 0.79% and there were 5 amino acids (from high to low: Leu 16.35%, Ala 8.36%,



**Figure 1.** Neighbor-joining (NJ) tree of 15 Characidae species based on 12 PCGs. The bootstrap values were based on 10,000 resamplings. The number at each node is the bootstrap probability. The number before the species name is the GenBank accession number. The genome sequence in this study is labeled with a black spot.

Thr 7.73%, Ile 7.28%, Phe 6.89%) used in a high frequency. In particular, the content of Leu was the highest.

The lengths of 12S ribosomal RNA and 16S ribosomal RNA were 949 bp and 1675 bp, which were both located in the typical positions between tRNA-Phe and tRNA-Leu (UUA), separated by tRNA-Val (Petrillo et al. 2006). The overall A + T content was 57.81%, the AT and GC skew value for the rRNAs were 0.2037 and  $-0.0894$ , respectively. The 22 *N. palmeri* mt-tRNA genes varied in size from 67 nucleotides to 75 nucleotides. Additionally, the length of control region (D-loop) was 1687 bp, ranging from 15,654 bp to 17,340 bp. The control region is also called A + T-rich region because of its high content of adenine and thymine, 620 nucleotides for A, 565 nucleotides for T, both of them were accounting for 70.24% of the whole D-loop, which was a usual large number higher than overall AT content (61.21%).

Neighbor-joining (NJ) tree of 15 Characidae species was constructed based on 12 PCGs. The bootstrap values were based on 10,000 resamplings. The number at each node is the bootstrap probability. The result suggested that *N. palmeri* was most closely related to *Gephyrocharax atracaudatus* among all the Characidae species included in the analysis. This result was consistent with conventional morphological taxonomy. In **Figure 1**, Characidae were divided into three branches, *Nematobrycon* + *Gephyrocharax* was resolved as sister group of a clade formed by *Oligosarcus* + *Astyanax* + *Paracheirodon* + *Grundulus* in the combined data set within the tribe Characidae. The other genera were far more evolutionarily related than the above two clades. Our findings suggested that, the phylogenetic placement of *N. palmeri* was well supported within Characidae. We believe that more similar discoveries of these species would further promote better understand of the Characidae family.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was supported by National Natural Science Foundation of China [No. 31702321], Zhejiang Provincial Natural Science Foundation of China [No. LY20C190008], Science and Technology Project of Zhoushan [No. 2020C21016], Open Foundation from Key Laboratory of Tropical Marine Bio-resources and Ecology, Chinese Academy of Sciences [No. LMB20201005].

## Data availability statement

The data that support the findings of this study are openly available at NCBI (<https://www.ncbi.nlm.nih.gov>), GenBank accession no. MN861079. And the data that support the findings of this study are also available from the corresponding author, Dr. Liu, upon reasonable request.

## References

- Barnett R, Larson G. 2012. A phenol-chloroform protocol for extracting DNA from ancient samples. *Methods Mol Biol.* 840:13–19.
- Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27: 1767–1780.
- Liu J, Ding Q, Gao L. 2017. The complete mitochondrial genome of North Island brown kiwi (*Apteryx mantelli*). *Mitochondrial DNA Part B.* 2(1):1–2.
- Lowe T, Eddy S. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5): 955–964.
- Miya M, Kawaguchi A, Nishida M. 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Mol Biol Evol.* 18:1993–2009.
- Petrillo M, Silvestro G, Nocera PPD, Boccia A, Paoletta G. 2006. Stem-loop structures in prokaryotic genomes. *BMC Genomics.* 7:170.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30(12):2725–2729.
- Zhu K, Lu Z, Liu L, Gong L, Liu B. 2018. The complete mitochondrial genome of *Trachidermus fasciatus* (Scorpaeniformes:Cottidae) and phylogenetic studies of Cottidae. *Mitochondrial DNA Part B.* 3(1):301–302.