Complete mitochondrial genome of the copepod Sinergasilus undulates (Copepoda: Poecilostomatoida)

Cong-jie Hua, Ming-yang Su, Zhi-wei Sun, Yuan-hong Lu and Jin-mei Feng

Wuhan Institute of Biomedical Sciences, School of Medicine, Jianghan University, Wuhan, People’s Republic of China; Department of Pathogenic Biology, School of Medicine, Jianghan University, Wuhan, People’s Republic of China

ABSTRACT
The total mitochondrial genome size of Sinergasilus undulates is 14,239 bp in length, including 13 protein-coding genes (PCGs), two rRNA genes, 22 transfer RNA genes, and a non-coding control region (D-loop). The overall nucleotide composition of the mitochondrial DNA of S. undulates is 34.9% A, 35.5% T, 15.7% C, 13.9% G, and 70.4% AT, respectively. Phylogenetic analysis suggests that the genus Sinergasilus is monophyletic, and S. undulates is closely related to S. polycolpus. The complete mitochondrial genome of S. undulates would be useful for species identification, epidemiology, and phylogenetics among Copepods.

Copepods in the genus Sinergasilus Yin, 1949 are considered as the most important parasites in freshwater fish. In this genus, three species are described with restricted host specificity. Among them, S. undulates is reported only infecting on common carp, Cyprinus carpio, and crucian carp, Carassius carassius, which are important economic fish species in China (Nie and Yao 2000). The first valid mitochondrial genome in the genus Sinergasilus Yin, 1949 is Sinergasilus polycolpus (Peng et al. 2010; Feng et al. 2015). Here, we assembled the complete mitochondrial genome of S. undulates, which is the second reported one in this genus. The complete mitochondrial genome of S. undulates would be useful for species identification, epidemiology, and phylogenetics among Copepods.

In this study, the S. undulates was collected from a naturally infected C. carpio, obtained in the Lake Dongsong, Yueyang (29°38’N, 113°09’E), China. The complete mitochondrial gene sequences were determined using the Sanger sequencing method. Voucher specimens are permanently stored in absolute ethanol under accession number WIBS20150911, in Wuhan Institute of Biomedical Sciences, School of Medicine, Jianghan University, Wuhan, China. The total genomic DNA was extracted from a single specimen (WIBS 20150911_1) using SDS/Protease K according to the instructions of TIAnamp Genomic DNA Kit (Tiangen Biotech, Beijing, China) following the manufacturer’s protocol. Subsequently, based on the existing mitochondrial genes of S. polycolpus (EU621723, NC_028085), pairs of primers were designed to amplify fragments from 16S rRNA (F: 5’-CCTAAATTCAACATCGAGGTC-3’, R: 5’-CATCCTCGCAGGTC-3’), cox1 (F: 5’-CGHCNGGTATGCGTCTGACC-3’, R: 5’-GATTGGTTTTACACATCCITGG-3’), and cyt b (F: 5’-GTATGTCTCACTCCTGGDNCA-3’, R: 5’-TCTACTGGCGTCHCCGATT-3’). Then, other pairs of primers were designed to amplify the remaining mitochondrial genome sequences of S. polycolpus, 16S-cox3 (F: 5’-GAAATTGATTAGTCTATTTTT-3’, R: 5’-CATCCTGGGGTCTTATTACA-3’). The complete mitochondrial genome of S. undulates was assembled manually and aligned against other published mitochondrial genome sequences of Copepods using the program MAFFT (Katoh and Standley 2013) to determine the gene boundaries. BLAST and ORF Finder NCBI tools were also used to identify and annotate the protein-coding genes (PCGs) and rRNAs. Transfer RNA (tRNA) genes and their secondary structures were identified using tRNAscan-SE (Lowe and Eddy 1997), MITOS (Bernt et al. 2013), and ARWEN (Laslett and Canback 2008). Nucleotide composition (%) of the complete mitochondrial sequences was calculated using PhyloSuite v1.2.2 (Zhang et al. 2020). The phylogenetic
analysis and analysis of other nucleotide and amino acid components were performed by using PhyloSuite v1.2.2 (Zhang et al. 2020). The complete mitochondrial genome sequence of *S. undulatus* (GenBank accession number MW080644) is 14,239 bp in length, including 13 PCGs, two rRNA genes, 22 tRNA genes, and a non-coding control region (D-loop). The overall nucleotide composition of its mitochondrial DNA is 34.9% A, 35.5% T, 15.7% C, 13.9% G, and 70.4% AT, respectively. Seven PCGs started with ATA, three PCGs started with ATG, but *cytb* uses ATC, *nad3* uses ATT as the start codon; nine PCGs were finished with TAA, but *nad4* and *nad6* were finished with TAG. The incomplete stop codon (T––) was found in two genes (cox2 and cox3). Fourteen overlaps exist among mitochondrial genes. All 22 tRNAs distributed on the H and L strands were between 53 and 74 bp in length. Fifteen tRNA genes were encoded on the L and seven on the H strands. Most of the tRNAs could form a common cloverleaf secondary structure, with the exception of three tRNAs that lacked the DHU arm (trnR, trnS1, and trnS2), and the trnC lacking the TψC arm. Two rRNA genes, 12S and 16S were 657 bp and 919 bp in size, respectively.

A phylogenetic tree was reconstructed using 18 mitogenomes from the subclass Copepoda and two species of Branchiopoda. In PhyloSuite, amino acid sequences and two rRNA genes were aligned in batches and ambiguously aligned fragments removed using two plug-in programs: MAFFT (Katoh and Standley 2013) and Gblocks 0.91b (Talavera and Castresana 2007), respectively. Subsequently, the optimized alignments were concatenated by PhyloSuite. Bayesian information criterion in ModelFinder (Kalyaanamoorthy et al. 2017) was used to select the optimal evolutionary model (GTR + F + I + G4). Bayesian inference analysis, conducted using MrBayes 3.2 (Ronquist et al. 2012), was used for phylogenetic reconstruction: 2,000,000 generations, four Markov chain Monte Carlo chains, and the trees were sampled every 1000 generations, and the initial 25% of trees were discarded as burn-in. The phylogenetic tree suggested that the genus *Sinergasilus* is monophyletic, and the *S. undulatus* is the sister group to *S. polycopus* (Figure 1).

**Disclosure statement**

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

**Funding**

This study was supported by the Natural Science Foundation of China to JMF [Grant No. 32070420] and Science Foundation of Jianghan University to HCJ [Grant No. 1010-08390001].

**Data availability statement**

Mitogenome data supporting this study are openly available in GenBank at: https://www.ncbi.nlm.nih.gov/nuccore/MW080644.

**References**

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 69(2): 313–319.
Nie P, Yao W. 2000. Seasonal population dynamics of parasitic copepods, Sinergasilus spp. on farmed fish in China. Aquaculture. 187(3–4):239–245.

Peng G-X, Gao Q, Song Y, Zhao Q-P, Luo Y-L, Nie P. 2010. Mitochondrial genes of Sinergasilus polycopus (Copepoda, Ergasilidae) parasitizing the gills of fish. Acta Hydrobiol Sin. 34(1):177–183.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 56(4):564–577.

Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 20(1):348–355.