Gene 591 (2016) 369–375

Contents lists available at ScienceDirect

Gene

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

Research paper

Characterization of the complete mitochondrial genome of *Cynoglossus gracilis* and a comparative analysis with other Cynoglossinae fishes

Min Wei [a,b], Yang Liu [a,b], Hua Guo [b], Fazhen Zhao [a,b], Songlin Chen [a,b,*]

[a] Wuxi Fisheries College, Nanjing Agricultural University, Wuxi 214081, China
[b] Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences (CAFS), Key Laboratory for Sustainable Development of Marine Fisheries, Ministry of Agriculture, Qingdao 266071, China

ARTICLE INFO

Article history:
Received 25 January 2016
Received in revised form 10 June 2016
Accepted 11 June 2016
Available online 13 June 2016

Keywords:
Mitochondrial genome
Cynoglossus gracilis
Cynoglossinae
Phylogenetic analysis

ABSTRACT

Mitochondrial genomes can provide basic information for phylogenetic analysis and evolutionary studies. We present here the mitochondrial genome of *Cynoglossus gracilis*, which is 16,565 bp in length. Numerous distinct regions were identified, including 13 protein-coding genes (PCGs), 22 tRNA genes, two rRNA genes, a light-strand replication origin, and a control region. Interestingly, we detected rearrangement of genes in *C. gracilis*, including a control region translocation, $tRNA^{Gln}$ gene inversion, and $tRNA^{Ile}$ gene shuffling. Additionally, a phylogenetic analysis based on the nucleotide sequences of the 13 PCGs using maximum likelihood and Bayesian inference methods reveals that *C. gracilis* is closely related to *Cynoglossus semilaevis*. This study provides important mitogenomic data for analyzing phylogenetic relationships in the Cynoglossinae.

1. Introduction

Cynoglossus gracilis (*Pleuronectiformes*: Cynoglossidae) is a migratory, marine fish. It is currently distributed across the northwestern Pacific from China to Korea (Menon, 1977). Its wild resources have been depleted and continue to be highly threatened due to massive water conservancy constructions, overfishing, and severe environmental contamination (Chen et al., 2002; Cheng et al., 2015).

In animals, mitochondrial DNA (mtDNA) is a small, circular genome ranging in size from 15 to 20 kb. It generally encodes 13 protein-coding genes (PCGs), 22 tRNA genes, and two rRNA genes (Boore, 1999). Additionally, it contains two functional non-coding regions: the light-strand (L-strand) replication origin (O_L) and the control region (CR). Gene order in vertebrate mtDNA is thought to be quite conserved, especially in fishes (Clayton, 1992; Manchado et al., 2007). However, an increasing number of gene rearrangements in mtDNA have been identified in many taxa, including reptiles (Mueller and Boore, 2005; Okajima and Kumazawa, 2010), birds (Bensch and Harlid, 2000; Schirtzinger et al., 2012; Verkuil et al., 2010), amphibians (Macey et al., 1997; Sano et al., 2005), and fishes (Gong et al., 2013a; Inoue et al., 2003; Kong et al., 2009; Ki et al., 2008; Mabuchi et al., 2004; Ponce et al., 2008; Shi et al., 2013, 2014a, 2015a).

Thus far, mainly small-scale gene rearrangements have been reported for teleosteans. In flatfish (*Pleuronectiformes*), however, large-scale gene rearrangements appear to be rather common (Shi et al., 2013). For example, in *Crossorhombus azureus* (Bothidae), a large-scale gene rearrangement has been found and characterized in a cluster involving the *ND6* gene and seven *tRNA* genes (Shi et al., 2013). Additional large-scale gene rearrangements include a control region translocation, *tRNA* gene inversions, and shuffling in *Cynoglossus semilaevis* (Kong et al., 2009). Furthermore, a typical gene order and large-scale gene rearrangements have been found within the teleost genus *Symphurus* (Shi et al., 2015a). In contrast, no gene rearrangements have been identified in soles (Soleidae), even though they are the closest family to the Cynoglossidae (Shi et al., 2014b; Manchado et al., 2007; Wang et al., 2015).

Here we present the complete mitogenome sequence of *C. gracilis*. Analysis of this genome indicates the presence of gene rearrangements, including a control region translocation, $tRNA^{Gln}$ gene inversion, and $tRNA^{Ile}$ gene shuffling. We also constructed a phylogenetic tree based on mitogenome data from the Cynoglossinae using both maximum likelihood (ML) and Bayesian inference (BI) methods. These data support a close relationship between *C. gracilis* and *C. semilaevis* and provide important genomic data for analyzing phylogenetic relationships in the Cynoglossinae.

Abbreviations: PCR, polymerase chain reaction; mitogenome, mitochondrial genome; mtDNA, mitochondrial DNA; *ATP6* and *ATP8*, genes for adenosine triphosphate synthases 6 and 8 respectively; *COXI-III*, genes for cytochrome oxidase subunits I-III; *Cytb*, a gene for cytochrome b; *ND1-6* and *ND4L*, genes for NADH dehydrogenase subunits 1–6 and 4L respectively; rRNA, ribosomal RNA; *tRNA**, for transfer RNA (where * is replaced by three letter amino acid code of the corresponding amino acid); 16S, large rRNA subunit; 12S, small rRNA subunit; PCGs, protein coding genes.

* Corresponding author at: Wuxi Fisheries College, Nanjing Agricultural University, Wuxi 214081, China.
E-mail address: chensl@ysfri.ac.cn (S. Chen).

http://dx.doi.org/10.1016/j.gene.2016.06.023
0378-1119/

2. Materials and methods

2.1. Sample and DNA extraction

Three wild adult *C. gracilis* specimens (weights 43.7, 52.6, and 55.5 g; lengths 21.1, 23.2, and 25.3 cm) were collected from the Yangtze River in Anhui Province, China. One of the three specimens was used for whole mitogenome sequencing. Whole genomic DNA was extracted using a tissue DNA kit (Omega, USA) according to the manufacturer's recommended protocol.

2.2. PCR amplification and sequencing

Sixteen pairs of primers (Table 1) were used to amplify contiguous overlapping fragments of the complete mitogenome of *C. gracilis*. The primers CG5F/R (Table 1) were also used to amplify the gene rearrangements fragment (from *ND1* to *ND2*). Primer design was based on the published mitogenome of *C. semilaevis* (GenBank: EU366230). The reactions were conducted using a total volume of 30 μL consisting of 22.45 μL sterile deionized water, 0.15 μL ExTaq DNA polymerase (5 U/μL, Takara), 3 μL 10× PCR Buffer (15 mM Mg^{2+}), 2.4 μL dNTPs (2.5 mM each), 1 μL DNA template (40 ng/μL), and 0.5 μL each primer (10 μM) with the following cycling parameters: initial denaturation at 94 °C for 5 min; followed by 23 cycles of 94 °C for 45 s, 56–60 °C (Table 1) for 40 s, and 72 °C for 90 s. The PCR products were purified using Gel Extraction Kit (CWBIO, China) according to the manufacturer's recommended protocol, and purified PCR products were sequenced by Genewiz Biotechnology (Beijing, China).

2.3. Sequence analysis

Overlapping sequenced fragments were assembled to create the complete mitogenome of *C. gracilis* using the DNASTAR software package. Annotation and boundary determination of PCGs and rRNA genes were conducted using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and NCBI-BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). tRNA genes were identified using tRNAscan-SE Search Server v1.21 (Lowe and Eddy, 1997) and comparative mitogenomics with the complete mitogenomes of 12 Cynoglossinae species deposited in GenBank (Table 2). The *C. gracilis* mitogenome map was generated using CGView (Grant and Stothard, 2008). The stem and loop structure of the O_L was predicted using Mfold Web Server (Zuker, 2003). In addition, base composition and codon usage were predicted using the program MEGA v6.0 (Tamura et al., 2013).

Table 2
List of species from subfamily Cynoglossinae used in this study.

No	Species name	Accession ID	References
1	*Cynoglossus gracilis*	KT809367	This study
2	*Cynoglossus abbreviatus*	GQ380410	Unpublished
3	*Cynoglossus bilineatus*	JQ349000	Shi et al. in press
4	*Cynoglossus itinus*	JQ639062	Unpublished
5	*Cynoglossus lineolatus*	JQ349004	Unpublished
6	*Cynoglossus puncticeps*	JQ349003	Yang et al., 2016
7	*Cynoglossus semilaevis*	EU366230	Kong et al., 2009
8	*Cynoglossus sinicus*	JQ348998	Shi et al., 2015b
9	*Cynoglossus trigrammus*	KP057581	Mu et al., 2015
10	*Paraplagusia bilineata*	JQ349001	Unpublished
11	*Paraplagusia blochii*	JQ349002	Li et al., 2016
12	*Paraplagusia japonica*	JQ639066	Gong et al., 2013b

2.4. Phylogenetic analysis

To position *C. gracilis* in the Cynoglossinae phylogeny, phylogenetic analyses of the *C. gralilis* mitogenome and mitogenomes of 11 other species of the Cynoglossinae, downloaded from GenBank (Table 2), were conducted. *Symphurus plagiusa* (GenBank: JQ639061) was used as the outgroup. Nucleotide sequences of each of the 13 PCGs were aligned using the ClustalW method (MEGA v6.0) and then concatenated together by Editseq (DNASTAR) after filtering ambiguous positions. PartitionFinder v1.1.1 (Lanfear et al., 2012) was used to select the optimal partition schemes and nucleotide substitution models with the "greedy" algorithm, with branch lengths estimated as "linked" and Bayesian information criterion (BIC). The optimal partition schemes and substitution models obtained from PartitionFinder are shown in Table S1.

ML and BI analyses were used to infer phylogenetic trees. RAxML v7.2.6 (Stamatakis, 2006) was used for ML analysis. Node support was calculated with 1000 bootstrap replications using the rapid bootstrap feature of RAxML (random seed value 12,345). The Bayesian analyses were conducted by MrBayes v3.2 (Ronquist and Huelsenbeck, 2003). Two independent analyses were run for one million generations with trees sampled every 100 generations. The first 2500 trees (25%) were discarded as burn-in, and the remaining trees were used to construct Bayesian consensus trees which share a majority consensus rule. Stationarity was determined when the average standard deviation of split frequencies was below 0.01.

Table 1
Primer pairs used for amplification and sequencing of mitogenome.

Primer label	sense (5′ → 3′)	Primer label	Anti-sense (5′ → 3′)	Annealing temperature
CG1F	TTGATTAGTCCTACGAGCAC	CG1R	TACTTGATTTATTGTTCCAGG	56
CG2F	TTACACCGAGAAGTTATCCG	CG2R	CTCTAGTGTAAGGTTAACGG	58
CG3F	CAAGTGTAACTCGTAATGGAC	CG3R	AATTGGTTGCAGGAGACCG	60
CG4F	GACATCCTAATGGTGCAGC	CG4R	ATGAATCACATTAAGCATAGG	58
CG5F	GATTCATCTCAACCATTGCTG	CG5R	GTTGCTAGTCCAAATAGTCCG	60
CG6F	GATCTCCTAAGTGAAGGTCC	CG6R	CCTAAGTCTCTTTATAGAAGG	60
CG7F	ACACACCCACCATCTCTGC	CG7R	TGTCAACATCTATACCCACC	58
CG8F	AGTTTCCTCTATCTTAGGAGC	CG8R	TAGTAGAGGTTATTAGTGTGG	60
CG9F	TACTTTAAATCAAGTGTGAGC	CG9R	AATTAGCTTCTTAGTGATTCG	56
CG10F	CTCACCCTTAGTGAAATGC	CG10R	ATGCTCAGAAGAAGCCAGC	58
CG11F	GCACTATTACTAACATCAGG	CG11R	AGCAGAAATAGAATTGCGAC	56
CG12F	TACTAATCTCACACCTCCTG	CG12R	AGCAAGGATTATTGAACCTG	58
CG13F	TCCAAACTAATACAGGATCAC	CG13R	CACAGGCAGGTTGTTATTGC	60
CG14F	GTGATCTAGGTCTCTTCGTC	CG14R	ATTGATGTACTCTGATGTAGG	60
CG15F	TTTCGTCTCAATTAATCGACC	CG15R	GCATTATCTACTGAGAATCCG	60
CG16F	TATCATTTTGAGGTGCAACAG	CG16R	ATGCTTTAACCACTCTTTACG	58