Gene 555 (2015) 169–177

Contents lists available at ScienceDirect

Gene

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

Complete mitochondrial genome of *Cuora trifasciata* (Chinese three-striped box turtle), and a comparative analysis with other box turtles

Wei Li [a,b], Xin-Cheng Zhang [a], Jian Zhao [a,b], Yan Shi [a], Xin-Ping Zhu [a,*]

[a] Key Laboratory of Tropical & Subtropical Fishery Resource Application & Cultivation of Ministry of Agriculture, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510380, PR China
[b] College of Life Science and Fisheries, Shanghai Ocean University, Shanghai 201306, PR China

ARTICLE INFO

Article history:
Received 19 May 2014
Received in revised form 22 October 2014
Accepted 31 October 2014
Available online 4 November 2014

Keywords:
Mitochondrial genome
Cuora trifasciata
Genus *Cuora*
Comparative mitogenomic analysis
Phylogenetic analysis

ABSTRACT

Cuora trifasciata has become one of the most critically endangered species in the world. The complete mitochondrial genome of *C. trifasciata* (Chinese three-striped box turtle) was determined in this study. Its mitochondrial genome is a 16,575-bp-long circular molecule that consists of 37 genes that are typically found in other vertebrates. And the basic characteristics of the *C. trifasciata* mitochondrial genome were also determined. Moreover, a comparison of *C. trifasciata* with *Cuora cyclornata*, *Cuora pani* and *Cuora aurocapitata* indicated that the four mitogenomics differed in length, codons, overlaps, 13 protein-coding genes (PCGs), *ND3*, rRNA genes, control region, and other aspects. Phylogenetic analysis with Bayesian inference and maximum likelihood based on 12 protein-coding genes of the genus *Cuora* indicated the phylogenetic position of *C. trifasciata* within *Cuora*. The phylogenetic analysis also showed that *C. trifasciata* from Vietnam and China formed separate monophyletic clades with different *Cuora* species. The results of nucleotide base compositions, protein-coding genes and phylogenetic analysis showed that *C. trifasciata* from these two countries may represent different *Cuora* species.

1. Introduction

A typical animal mitochondrial genome (mitogenome, MT) is a circular DNA of 16–18 kb and usually contains 13 proteins, 22 tRNAs, and two rRNAs (Garesse and Vallejo, 2001). In addition, there is a noncoding (control or D-loop) region, which controls replication and transcription (Boore, 1999). In most turtles, with the exception of *Platysternon megacephalum* and *Malacochersus tornieri*, the genome organization and gene order are similar to those of typical vertebrates (Peng et al., 2006). Since Kumazawa and Nishida (1995) first sequenced the complete mitochondrial genome of the green sea turtle (*Chelonia mydas*), 59 species of Testudines have been reported in the GenBank database. mtDNA has been widely used as a molecular marker for evolutionary, phylogenetic, and population genetic studies (e.g. Zardoya and Meyer, 1998; Cao et al., 2000; Olmo et al., 2002; Ma et al., 2013).

Cuora trifasciata (Chinese three-striped box turtle) is a freshwater turtle in the genus *Cuora* that was previously widely distributed in Vietnam, Laos and southwestern China. However, because of habitat destruction, over-exploitation, traditional medicine, and the pet trade, *C. trifasciata* has catastrophically declined in the wild (Rhodin et al., 2011). *C. trifasciata* was listed as an endangered species on the International Union for Conservation of Nature (IUCN) Red List in 1996, and its alert status was set to the critically endangered level in 2000 (IUCN, 2008). In 2011, a report by the Wildlife Conservation Society (WCS) and Turtle Survival Alliance (TSA), "Turtles in Trouble", listing the world's 25 most endangered turtles, in which *C. trifasciata* was ranked ninth (Rhodin et al., 2011).

Since the early 2000s, *C. trifasciata* has been the topic of disputes concerning the range and phylogenetic relationships of the genus *Cuora*. The first dispute concerns the relationship of *C. trifasciata* with *Cuora aurocapitata* and *Cuora pani* (Stuart and Parham, 2004; He et al., 2007; Spinks and Shaffer, 2007). The second dispute concerns the controversial classification of *Cuora cyclornata*, which was determined to be independent from *C. trifasciata* in the past few years (Blanck et al., 2006). At the same time, Spinks et al. (2009) recommended it as a junior synonym of *C. trifasciata* and then suggested that further data are needed (2012). In addition, *C. aurocapitata* and *C. pani* have been suspected

Abbreviations: A + T, percentage of adenosine and thymidine; aa, amino acid; Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; ATP, adenosine 5-triphosphate; bp, base pair(s); C, cytosine; COX, cytochrome c oxidase; CR, control region; CSB, conserved sequence blocs; Cys, cysteine; Cytb, cytochrome b; DNA, deoxyribonucleic acid; G, guanine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; H, heavy; His, histidine; Ile, isoleucine; kb, kilobase; L, light; Leu, leucine; Met, methionine; min, minute; mtDNA, mitochondrial DNA; NAD, nicotinamide-adenine dinucleotide; NADH, nicotinamide-adenine dinucleotide (reduced); ND, NADH dehydrogenase; ng, nanogram; PCG, protein coding gene; PCR, polymerase chain reaction; Phe, phenylalanine; pM, picomole; Pro, proline; RNA, ribonucleic acid; rRNA, ribosomal ribonucleic acid; s, second; S, subunit; Ser, serine; T, thymine; Thr, threonine; Tm, melting temperature; tRNA, transfer ribonucleic acid; Trp, tryptophan; Tyr, tyrosine; Val, valine; U, uracil.

* Corresponding author.
E-mail address: zhuxinping_1964@163.com (X.-P. Zhu).

http://dx.doi.org/10.1016/j.gene.2014.10.060
0378-1119/

to be the same species. However, there has been no detailed mitogenomic analysis of *C. aurocapitata* and *C. pani*, although their complete sequences have been submitted to the NCBI. A key way of resolving these problems is to determine the complete mitochondrial genome of *C. trifasciata*, which has not yet been reported.

To further identify and protect these turtles, we determined the complete mitogenome sequence of *C. trifasciata*. In this study, the complete mitogenome of *C. trifasciata* was determined for the first time, and its genome organization, gene arrangement and characterization are described. Moreover, *C. trifasciata* (Chinese three-striped box turtle, which is called *C. trifasciata* CHN below) was compared with *C. cyclornata* (Li et al., 2014) (Vietnamese three-striped box turtle, which is called *C. trifasciata* VIE below), *C. pani* and *C. aurocapitata* based on different mitogenomic aspects. In addition, 11 complete mitochondrial genomes of the genus *Cuora* were collected. Using the maximum likelihood (ML) and Bayesian inference (BI) methods, 12 protein-coding gene regions of the mtDNAs were used to separately perform molecular phylogenetic analyses to provide more information on the relationships among the genus *Cuora*.

2. Materials and methods

2.1. Sample collection and DNA extraction

The tissue samples were obtained from *C. trifasciata* (which were collected from Guangdong Provence, China) specimens at the Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, China. The total genomic DNA was extracted using a Tissue DNA Kit (Omega, USA) according to the manufacturer's instructions.

2.2. Mitochondrial DNA amplification and sequence analysis

Fifteen pairs of primers (Table 1) were designed to amplify contiguous and overlapping fragments of the complete mitochondrial genome of *C. trifasciata* according to other published mitogenomes of the genus *Cuora*. PCR was performed in a total volume of 40 μL containing 2.4 μL of dNTP (2.5 mM), 1 μL of template DNA, 0.5 μL of each primer (10 μmol), 0.5 μL of Taq DNA polymerase (5 U/μL) and 4 μL of 10× Buffer (TaKaRa, Dalian, China). Using Bio-RAD iCyclers (TaKaRa, Dalian, China), the amplifications were performed under the following conditions: predenaturation at 95 °C for 2 min followed by 35 cycles of 94 °C for 30 s, a linear gradient of 51 °C to 58 °C for 30 s and 72 °C for 2 min and a final extension at 72 °C for 10 min. The PCR products were electrophoresed on a 1.0% agarose gel, and the DNA fragments of the intended sizes were purified using a Gel Extract Purification Kit (Omega, USA). The products were then sequenced by Invitrogen Biotechnology (Shanghai, China). We repeated the PCR process and sequencing to make sure the accuracy of the result.

2.3. Sequence editing and analysis

The complete mitochondrial genome sequence of *C. trifasciata* was manually aligned and corrected using EditSeq (DNASTAR INC.) and ClustalX 1.8 (Thompson et al., 1997). The locations of 13 protein-coding genes, two *rRNA* genes, and 22 *tRNA* genes were determined using the programs DOGMA (Wyman et al., 2004), tRNAscan-SE 1.21 (Lowe and Eddy, 1997) and ARWEN (online version) (Laslett and Canbäck, 2008). The tRNA genes were also identified according to their secondary structures and positions in other turtles.

The nucleotide base compositions and the relative synonymous codon usage (RSCU) values for each PCG were calculated with the MEGA 5 program (Tamura et al., 2011). All PCGs were aligned at the amino acid level using the default settings in Clustal *X*2 (Larkin et al., 2007). The alignments were back-translated to the corresponding nucleotide sequences. Because of high variability, the stop codons in PCGs were excluded in the alignment (Zhang et al., 2014; Shuang-Shuang et al., 2014).

2.4. Phylogenetic analysis

Phylogenetic analysis was performed using the complete mitochondrial genomes of 10 members of the Geoemydidae family and *C. trifasciata* (CHN) in this study. *Pelodiscus sinensis* of the Trionychidae family was used as an outgroup (Table 2). Twelve protein-coding genes were chosen for the construction of Bayesian and ML trees. Because the *ND6* gene was the only one of the 13 protein-coding genes located in the L-strand, its nucleotide composition and amino acid composition were obviously different from those of the other 12 protein-coding genes, and the *ND6* gene was excluded from the phylogenetic analysis (Carapelli et al., 2000). The third codon position of the 12 protein-coding genes was analyzed with DAMBE 5.3.17 (http://dambe.bio.uottawa.ca/), and the results revealed that its transition and transversion rates were both saturated. Finally, 1 + 2 codons of 7238 bp were used for the phylogenetic analysis, and every gene was aligned and concatenated using ClustalX Version 2.1 (Thompson et al., 1997).

The phylogenies were determined using both Bayesian inference (BI) and the maximum likelihood (ML) method. The BI analysis was performed using MrBayes version 3.2.1 (Ronquist and Huelsenbeck, 2003), and the best-fit model (GTR + I + G) of nucleotide evolution was selected using the hLRT and AIC test in MrModeltest 2.3 (Posada and Crandall, 1998). According to Markov Chain Monte Carlo analysis, four chains (one cold and three heated chains) were set to run simultaneously for 3,000,000 generations, by which time the average standard deviation of the split frequency between the runs had decreased to 0.015, displaying good agreement and implying convergence. Each set

Table 1
PCR primers used in the analysis of the *C. trifasciata* mitochondrial genome.

Code	Upper primer sequence (5′–3′)	Lower primer sequence (5′–3′)	Approximate product length (bp)	Tm (°C)
1	AAAGCATGGCACTGAAGTTGC	TTTCATCTTTCCCTTGCGGTAC	1234	53–56
2	AAAGCATTCAGCTTACACCTGA	AAGTTCCACAGGGTCTTCTCG	1145	55–68
3	TAATGCCTGCCCAGTGACA	TGATTCCGAGGGTTACTTC	1244	55–58
4	TCAGGCTGAGCTTCAAACTC	GTAGTTGGGTCTGATTTAGCCC	1336	57–60
5	ACCTGACAAAAACTAGCACCA	ACTATACCTGCTCAGGCCCCG	1118	59–62
6	TACCTGTGTTTTTAACCCGCTGAT	TGGGTGAATCCTGCTATAA	1173	53–58
7	GCTATCCCAACAGGAGTAAAAG	GCTATCCTGTTTAGCTTCTATAG	1508	55–59
8	AAGCAGATGCCGTTCCGGGACG	GTTATGAGTAATGCTGCTGCTGC	1223	59–63
9	GCCTCTACCTACAAGAAAAC	GAAAAATCGAATTGAGAAAGG	1094	55–57
10	AGTACAAGTGACTTCCAATCAa	TTTGGCCGCCTCAACGTGTA	1162	57–59
11	GAACCAACTCCACGAAAACG	GCTGTTTTTACAGTTGTTTTTG	1467	55–59
12	AGGATAGAAGCAATCCACTGG	TATCTTTCGAATGTCTTTTTC	1141	55–59
13	TATACACGCCTTCTTCAAAGC	CTAATAGTGATCCGAAGTTTCAT	1524	53–56
14	AACCACCGTTGTATTCAACTA	CAGTTTCACTGAGTCGGCAG	18890	57–60
15	AGGCCTCTGGTTAATGTGTT	TTGGGCTGTCATGGTGTGCCT	901	58–62