Gene 605 (2017) 12–19

Contents lists available at ScienceDirect

Gene

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

Identification of *Triplophysa* species from the Qinghai-Tibetan Plateau (QTP) and its adjacent regions through DNA barcodes

Jiuxuan Li [a,1], Ying Wang [a,b,d,1], Huifang Jin [a], Wujiao Li [a], Chaochao Yan [a], Pengfei Yan [a], Xiuyue Zhang [a], Shunping He [b], Zhaobin Song [a,c,*]

[a] Sichuan Key Laboratory of Conservation Biology on Endangered Wildlife, College of Life Sciences, Sichuan University, Chengdu 610065, PR China
[b] Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, PR China
[c] Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, PR China
[d] School of Life Sciences, Jianghan University, Wuhan 430056, PR China

ARTICLE INFO

Article history:
Received 25 July 2016
Received in revised form 21 October 2016
Accepted 30 November 2016
Available online xxxx

Keywords:
Triplophysa
DNA barcoding
Intraspecific and interspecific genetic distance
Phylogenetic tree
Species identification

ABSTRACT

The genus *Triplophysa* is the largest and most difficult to identity morphologically fish group of superfamily Cobitoidea with 140 currently valid species, and is mainly distributed in the Qinghai-Tibetan Plateau (QTP) and adjacent regions. Most species within this genus possess highly similar morphological characteristics for adaption to the highland environment and are very difficult to be identified only based on morphology. The traditional species identification, mainly based on external morphological diagnostic characters, leads to inconsistent results in many cases. Herein, we provided a molecular method based on mitochondrial cytochrome *c* subunit I (COI) for the identification of *Triplophysa* fishes. Thirty-three *Triplophysa* species, 244 individuals, were used to determine whether barcoding was effective in discriminating species for this genus. The mean intraspecific and interspecific K2P distances ranged from 0 to 14.9% (mean, 2.9%) and 0 to 23.4% (mean, 9.7%), respectively. The tree-based analysis displayed most of species formed discrete clusters with strong bootstrap support values (>90%). The results showed that most of *Triplophysa* species could be identified by DNA barcode and indicated DNA barcode could be used as a molecular marker for these species.

1. Introduction

The genus *Triplophysa* was first proposed in 1933 by Rendahl (Prokovfiev, 2001). It is the largest genus within the family Nemacheilidae, which belongs to superfamily Cobitoidea (Prokovfiev, 2010). Taxonomically, *Triplophysa* is the most complicated group of superfamily Cobitoidea. *Triplophysa* is mainly distributed in rivers and lakes in the Qinghai-Tibetan Plateau (QTP) and its adjacent regions (Zhu, 1989; Wu and Wu, 1992; He, 2008; Prokovfiev, 2010). It contains approximately 170 currently described species, including 140 valid species, most of which are distributed in China (Prokovfiev, 2006; He et al., 2011; Froese and Pauly, 2015). Nearly 100 new species of *Triplophysa* were described during over two decades (He et al., 2011) and new species of *Triplophysa* are constantly being discovered in recent years (He et al., 2008, 2012; Zheng et al., 2009, 2010; Yang et al., 2011, 2012; Li et al., 2015). Most new species were not compared with previously described species and that makes it more difficult to identify the species of *Triplophysa*.

The traditional species identification of *Triplophysa* usually needs the specialized morphological and anatomical knowledge (Zhu, 1989; Wu and Wu, 1992; He, 2008; Prokovfiev, 2009; Prokovfiev, 2010), which is mainly based on external morphological and internal anatomical diagnostic characters, including the shape characters of trunk, caudal peduncle and head, eye diameter and its position in head, shape and relative position of fins, structures of air-bladder and intestine and etc. These morphological characters also should be combined together to identify *Triplophysa* species. Meanwhile, the species identification of *Triplophysa* usually need complete samples for obtaining their morphological and anatomical characters. Most of *Triplophysa* species possess similar morphological characters, easily leading to confusing results in many cases (Ding, 1994; Prokovfiev, 2007; He, 2008). Some morphological characteristics used for species delimitation are closely size-associated or vary with the environment (He et al., 2011). Some countable or measurable characters, shape characters of trunk, body bars and barbel lengths, usually possess of great ecological plasticity and distinct differences for the same species from different rivers and even from the same river (He, 2008). The barbel length and body bars sometimes are

Abbreviations: QTP, Qinghai-Tibetan Plateau; COI, cytochrome *c* subunit I; NJ, Neighbour-joining; BI, Bayesian; AIC, Akaike Information Criterion; MCMC, Markov Chain Monte Carlo; BM, best match; BCM, best close match.

* Corresponding author.
E-mail address: zbsong@scu.edu.cn (Z. Song).

[1] These two authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2016.11.045
0378-1119/

also used as main characters in *Triplophysa* identification, but they were often unstable and could not be regarded as the final diagnostic characters (He et al., 2011). In addition, some different species possess similar or identical external morphological characters while show significantly internal anatomical difference (He, 2008). Thus, the high morphological variability and species diversity of *Triplophysa* cause great difficulties in the classification and identification of this group fish. Owing to some characters, the same species from different habitats ever was identified as different species or subspecies (Prokofiev, 2001; He, 2008). These phenomena implied that it is difficult to accurately discriminate the *Triplophysa* species only based on their morphological characters. Therefore, different classification criteria about this group have been proposed by many authors and the phylogenetic relationships of *Triplophysa* species remain controversial (He, 2008; Prokofiev, 2010).

The valid identification of fish species is an essential step for both fundamental science and fish conservation or management. In order to overcome the limitations of morphological diagnostic characters for identification, we introduced DNA barcoding to identify *Triplophysa* species in this study. DNA barcoding has been widely used as a biological tool for species identification based on short and standardized DNA sequences (Hebert et al., 2003a; Savolainen et al., 2005; Hubert et al., 2010; April et al., 2011; Asgharian et al. 2011; Zhang and Hanner, 2012; Knebelsberger et al., 2014a, 2014b), which are easily amplified with universal primers (Ward et al., 2005; Ivanova et al., 2007). The mitochondrial DNA cytochrome oxidase gene I (COI) is remarkably effective to discriminate animal species (Fazekas et al., 2009). The powerful and reliable tool has been used for the identification of fish species, containing all major marine taxa as well as freshwater species from different geographic regions (Ward et al., 2005, 2008; Hubert et al., 2008; Steinke et al., 2009; Mabragaña et al., 2011; Costa et al., 2012; Zhang and Hanner, 2012; Keskin and Atar, 2013; Mccusker et al., 2013; Geiger et al., 2014; Knebelsberger et al., 2014a, 2014b). Herein, we employed DNA barcoding to discriminate the *Triplophysa* species and improve the species identification and classification of this group fish. And, previous studies revealed that the high morphological and genetic variability occurred among *Triplophysa* species or even within the same *Triplophysa* species from different populations due to their diverse habitats (Prokofiev, 2001; He, 2006; Hou et al., 2012; Li et al., 2013; Li et al., 2014). Therefore, we also assessed the power of DNA barcoding for identifying *Triplophysa* species in this study.

2. Materials and methods

2.1. Sample collection

Triplophysa fish were collected from March 2003 to August 2015 and a total of 234 specimens were sampled (Table 1). Most of the species were obtained from Sichuan Province, and some were collected from Tibet, Gansu, Qinghai, Yunnan and Xinjiang, China. Specimens were identified by more than two people, respectively, based on morphological characters (characters of the shape of trunk, caudal peduncle and head, eye diameter and its position in head, shape and relative position of fins, structures of air-bladder and intestine) according to Zhu (1989) and Wu and Wu (1992). Several specimens that have the same morphological characteristics and could not be reliably identified to species were showed as *Triplophysa* sp. and controversial samples were not adopted. The whole fish or fin clips were immediately preserved in 100% ethanol after capture and stored under room temperature in laboratory. All experiments were in line with the Regulation of Animal Experimentation of Sichuan province of China.

2.2. DNA extraction, PCR and sequencing

Total genomic DNA was extracted using the standard proteinase K/phenol method with minor modifications (Sambrook and Russell, 2002). DNA quality was verified by 0.8% agarose gel electrophoresis. For all the samples, a fragment of the COI gene was amplified using two forward (FishF2, VF2) and two reverse primers (FR1d, FishR2) tagged with M13 tails at 5′ ends (Ivanova et al., 2007). Each PCR mixture, in a total volume of 25 μL, contained 1 μL DNA template, 2.5 μL 10 × PCR buffer, 2.5 μL $MgCl_2$ (2.5 mM), 1.0 μL dNTPs (2.5 mM), 1.0 μL each primer (0.1 μg/μL), 0.2 μL Taq polymerase (5 U/μL; Takara) and 16.8 μL MilliQ. The PCR reactions were performed as follows: initial 5 min denaturation at 95 °C, followed by 34 cycles of 95 °C (30 s), 54 °C (45 s), and 72 °C (50 s), a final elongation at 72 °C (10 min). The amplified products were electrophoresed on 0.8% agarose gel and purified with the DNA Agarose Gel Extraction Kit (Omega, USA). The purified fragments were sequenced on an ABI PRISM 3730 by Invitrogen Biotechnologies Company.

2.3. Data analysis

All sequence chromatograms of forward and reverse sequencing were edited and assembled using DNAstar 5.0 (Madison, WI, USA). Alignments were created using MUSCLE (Edgar, 2004) algorithm with default settings implemented in MEGA version 5.2 (Tamura et al., 2011). All the sequences have been submitted to GenBank (species name and GenBank accession numbers were shown in Table 1) and ten DNA barcoding (COI) sequences were obtained from GenBank (JQ686729, NC018774, NC024597, NC024611, NC027517, NC017890, NC025632, NC027517, KM396312, NC019587). Sequences statistics (variance in sequence length, base composition and number of invariable sites) were calculated with MEGA 5.2. All COI sequences were translated into amino acids in order to exclude sequencing errors and to avoid the inclusion of pseudogene sequences in the dataset. The effectiveness of the barcodes was evaluated in the following methods.

Pairwise genetic distances were computed with MEGA 5.2 using Kimura 2-parameter (K2P) distance model (Kimura, 1980) as it is the standard model for DNA barcoding data sets. Frequency histograms of the distribution of interspecific and intraspecific pairwise distances were created to detect whether a "barcoding gap" exists (Wiemers and Fiedler, 2007).

To investigate species boundaries, the threshold-based approach modified after Meyer and Paulay (2005) was employed (Nicolas et al., 2012). In this framework, the assumption (H_0) indicates that "two specimens belong to different species", when the interspecific distance is greater than a threshold "t" or intraspecific distance is less than "t", then "H_0" is accepted. Otherwise, "false-negatives" mean specimens coming from two different species that are classified within the same species and "false-positives" are specimens belonging to the same species that are classified in two different species. By varying the threshold "t" from 0 to the maximal interspecific distances, we drew the cumulative distribution functions of "false-positives" and "false-negatives" as a function of the interspecific K2P distances. The optimal "t" was obtained when the cumulative number of errors was at its minimum value.

Neighbour-joining (NJ) tree (Saitou and Nei, 1987) and Bayesian (BI) tree (Huelsenbeck and Ronquist, 2001) were constructed to gauge the accuracy of our results. NJ tree was created based on K2P genetic distances in MEGA 5.2 with bootstrap tests of 1000 replicates to verify the robustness of the tree (Felsenstein, 1985). Gaps and missing data were treated as pairwise deletions. Substitutions included transitions and transversions, and uniform substitution rate were used. BI tree was constructed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The best-fitting nucleotide substitution model was determined for Bayesian inferences analysis by the Akaike Information Criterion (AIC) in Modeltest 3.7 (Posada and Crandall, 1998). The TIM + I + G model was selected to generate the BI tree. Analysis was carried out using Markov Chain Monte Carlo (MCMC) analysis with one cold chain and three heated chains for 2,000,000 generations, with every 100th sample being retained. The first 5000 were discarded as a conservative burn-in and the remaining samples were used to generate a majority rule consensus tree. All MCMC runs were repeated to confirm