



Molecular identification of the genus *Thryssa* based on DNA barcoding

C.-Y. Ma, H.-Y. Ma, Y. Ni, W. Wang and L.-B. Ma

Key Laboratory of East China Sea and Oceanic Fishery Resources Exploitation,
Ministry of Agriculture, East China Sea Fisheries Research Institute,
Chinese Academy of Fishery Sciences, Shanghai, China

Corresponding author: L.-B. Ma
E-mail: malingbo@vip.sina.com

Genet. Mol. Res. 14 (4): 18580-18586 (2015)
Received August 9, 2015
Accepted October 31, 2015
Published December 28, 2015
DOI <http://dx.doi.org/10.4238/2015.December.28.5>

ABSTRACT. DNA barcoding is an effective method for identifying species by analyzing one or a few short standardized DNA sequences. In this study, we examined the utility of mitochondrial cytochrome oxidase subunit I (*COI*) sequences as a DNA barcode for the identification of six species belonging to the genus *Thryssa*: *T. dussumieri*, *T. hamiltonii*, *T. kammalensis*, *T. mystax*, *T. setirostris*, and *T. vitirostris*. We obtained an intraspecific distance of 0.000 for *T. vitirostris* and *T. hamiltonii*, 0.006 for *T. mystax*, 0.002 for *T. dussumieri*, and 0.005 for *T. kammalensis*. The average intraspecific distance was 0.002, while the average interspecific distance was 0.137. Thus, the interspecific genetic distance was approximately 67-fold larger than the intraspecific genetic distance; the average genetic distance among species was greater than the minimum of 0.020 between species suggested elsewhere. The genetic distance between *T. vitirostris* and *T. mystax* was 0.003. A maximum-likelihood phylogenetic tree constructed using best-fitting tree topology showed distinct clusters corresponding to the species (except for *T. vitirostris* and *T. mystax*). The closest relationship was found between *T. vitirostris* and *T. mystax*. These two species clustered together in the phylogenetic tree. This conclusion contradicts the evolutionary

relationship based on morphological classification.

Key words: *Thryssa*; DNA barcoding; COI; Molecular identification

INTRODUCTION

DNA barcoding is an effective method for identifying species by analyzing one or a few short standardized DNA sequences. The main goals of DNA barcoding are to determine relationship patterns among species and to identify an unknown sample in terms of a preexisting classification (Kress et al., 2005). DNA barcoding can identify a wide variety of species within a known taxonomic framework, and can be used to link different life stages within a species (Schindel and Miller, 2005; Puillandre et al., 2009; Feng et al., 2011). Barcoding has also been used to determine whether species should be combined or separated (Koch, 2010). One of the most widely used barcoding sequences is based on the mitochondrial cytochrome oxidase subunit I gene (*COI*) (Hebert et al., 2003). The *COI* gene sequence is short enough to be sequenced quickly and cheaply, but is also sufficiently long to identify variations among species in a range of animal groups (Wakabayashi et al., 2006; Persis et al., 2009; Glover et al., 2010; Tang et al., 2010; Yang et al., 2010). An abundant amount of barcode data is available in the public domain, for example, the Fish Barcode of Life Initiative (FISH-BOL) (Ward et al., 2009) is a global effort to coordinate the assembly of a standardized DNA barcode library for all fish species and is based on voucher specimens with authoritative taxonomic identification (Steinke and Hanner, 2011).

The genus *Thryssa* includes marine pelagic fish species that are found throughout the Indo-Pacific region. Six species of this genus have been identified: *T. dussumieri*, *T. hamiltonii*, *T. kammalensis*, *T. mystax*, *T. setirostris*, and *T. vitrirostris* (Zhu, 1984). Previous studies on the taxonomy of the genus used classical comparative morphological approaches (Zhu, 1984; Ma et al., 2010b). However, these methods are not effective in discriminating closely related species, especially if they lack characteristic features or possess misleading morphologies (Teletchea, 2009). Mitochondrial 16S rRNA gene sequences have also been used to evaluate phylogenetic relationships among *Thryssa* species (Ma et al., 2010b). Apart from this study, there is little information available from molecular genetic approaches to *Thryssa* taxonomy. In this study, we performed the first investigation on the utility of the mtDNA *COI* gene as a DNA barcode for the identification of different *Thryssa* species; this candidate barcode may provide valuable phylogenetic and population genetic information for this commercially important genus.

MATERIAL AND METHODS

Sample collection and DNA extraction

Specimens of six *Thryssa* species were collected from the seas off Fujian and Zhejiang provinces of China: *T. dussumieri* (Fujian, N = 3), *T. hamiltonii* (Fujian, N = 3), *T. setirostris* (Fujian, N = 1), *T. kammalensis* (Zhejiang, N = 3), *T. mystax* (Zhejiang, N = 2), and *T. vitrirostris* (Zhejiang, N = 2) (Table 1). Genomic DNAs were extracted from the muscle tissue using a standard phenol-chloroform extraction protocol (Ma et al., 2010a).

Table 1. List of *COI* sequences, GenBank accession Nos., and the geographic sources of the samples.

Species	N	Number of haplotypes	GenBank accession No.	Collection location
<i>T. vitrirostris</i>	2	1	KP403942	Zhejiang, China
<i>T. mystax</i>	2	2	KP403943, KP403944	Zhejiang, China
<i>T. dussumieri</i>	3	3	KP403945, KP403946, KP403947	Fujian, China
<i>T. hamiltonii</i>	3	1	KP403948	Fujian, China
<i>T. setirostris</i>	1	1	KP403949	Fujian, China
<i>T. kammalensis</i>	3	3	KP403950, KP403951, KP403952	Zhejiang, China

'N' indicates the number of sequences analyzed.

Mitochondrial COI gene sequences collection

The primer pair CFF (5'-TCR ACY AAY CAY AAA GAY ATY GGC AC-3') and CFR (5'-ACT TCW GGG TGR CCR AAG AAT CA-3') were used for PCR amplification of *COI* sequences. PCR was performed in a 25 μ L reaction volume containing 2.0 mM MgCl₂, 0.2 mM dNTP mix, 0.2 μ M each primer, 1U Taq DNA polymerase (TaKaRa, Dalian, China), 1X PCR buffer, approximately 100 ng template DNA, and deionized water. The amplification conditions were: initial denaturation of 5 min at 94°C; 35 cycles of 45 s at 94°C, 50 s at 54°C, and 45 s at 72°C; and a final extension at 72°C for 5 min. The PCR products were separated on 1.5% agarose gels (TaKaRa) and then sequenced in both directions using an ABI3730XL sequencer (Applied Biosystems).

Data analysis

The obtained sequences were edited and aligned using DNASTAR version 7.1 software. Variable sites were determined using MEGA 4.0 software (Tamura et al., 2007). Haplotypes were identified using DNA SP version 4.1 software (Rozas et al., 2003). Genetic divergence between species and between haplotypes was estimated using the Kimura 2-parameter (K2P) model (Kimura, 1980) by MEGA 4.0 software (Tamura et al., 2007). Maximum-likelihood analyses were performed on the *COI* sequences using PAUP v4.0 PAUP* version 4.0 (Swofford, 2003) with a heuristic search option, stepwise addition, 1000 replications and tree bisection reconnection (TBR) branch swapping. Likelihood-ratio tests to choose the best-fit model were performed with Modeltest v3.7 software (Posada and Crandall, 1998). Two *COI* sequences from *T. setirostris* (accession No. EF607597 and JF494687) and one sequence from *Anchoa lucida* (accession No. KF614704) were downloaded from the GenBank database. *A. lucida* was defined as the outgroup.

RESULTS

COI sequence divergence

We identified 14 *COI* homologous sequences in the six *Thryssa* species. As the sequences were protein encoding, no insertions or deletions were found after alignment and compilation. A consensus sequence of 543 bp was used for the barcoding analysis. The 14 *COI* sequences had an average of 29.2% T, 26.5% C, 26.2% A, and 18.1% G. In total, 11 haplotypes were found for *Thryssa*: 1 in *T. vitrirostris*, 2 in *T. mystax*, 3 in *T. dussumieri*, 1 in *T. hamiltonii*, 1 in *T. setirostris*, and 3 in *T. kammalensis*. The haplotypes varied at 161 sites, with the majority of these (146 sites) being located at the 3rd position of codons (Figure 1). Based on mitochondrial vertebrate codon

usage, a total of 181 amino acids were predicted with no premature stop codons; only 10 out of the 181 amino acids showed variation.

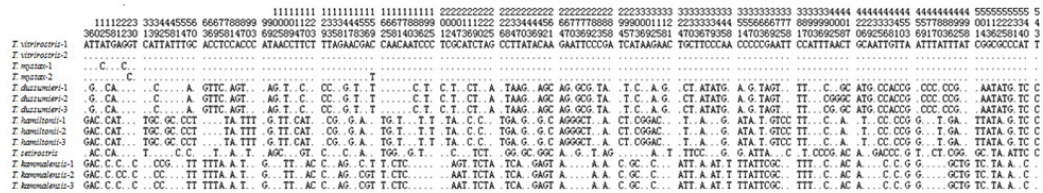


Figure 1. Variable sites in the mitochondrial cytochrome *c* oxidase I (*COI*) gene of six *Thryssa* species. Nucleotides identical to the sequence at the top are indicated by dots.

Intraspecific and interspecific divergence

Intraspecific and interspecific genetic distances were calculated using the Kimura-2-parameter (K2P) model and MEGA 4.1 software. We obtained an intraspecific distance of 0.000 for *T. vitrirostris* and *T. hamiltonii*, 0.006 for *T. mystax*, 0.002 for *T. dussumieri*, and 0.005 for *T. kammalensis*; the average intraspecific distance was 0.002 (Table 2). In contrast, interspecific genetic distances ranged from 0.003 to 0.160 (Table 3), with an average of 0.137. The average interspecific distance was approximately 67-fold larger than the intraspecific genetic distance; it exceeded the minimum genetic distance (0.020) between species recommended by Hebert et al. (2003). However, the genetic distance between *T. vitrirostris* and *T. mystax* was 0.003.

Table 2. Average and ranges of Kimura 2-parameter (K2P) distance values for the six *Thryssa* species.

Species	N	Minimum	Maximum	Average
<i>T. vitrirostris</i>	2	0.000	0.000	0.000
<i>T. mystax</i>	2	0.006	0.006	0.006
<i>T. dussumieri</i>	3	0.001	0.003	0.002
<i>T. hamiltonii</i>	3	0.000	0.000	0.000
<i>T. setirostris</i>	1	-	-	-
<i>T. kammalensis</i>	3	0.001	0.006	0.005

'N' indicates the number of sequences analyzed.

Table 3. Interspecific Kimura 2-parameter (K2P) distance values for the six *Thryssa* species.

Species	<i>T. vitrirostris</i>	<i>T. mystax</i>	<i>T. dussumieri</i>	<i>T. hamiltonii</i>	<i>T. setirostris</i>
<i>T. vitrirostris</i>	-				
<i>T. mystax</i>	0.003	-			
<i>T. dussumieri</i>	0.149	0.151	-		
<i>T. hamiltonii</i>	0.156	0.160	0.139	-	
<i>T. setirostris</i>	0.135	0.139	0.138	0.132	-
<i>T. kammalensis</i>	0.151	0.152	0.147	0.160	0.149

Phylogenetic relationship

Likelihood ratio tests to identify the best-fit model selected the transversion model (with a proportion of invariable sites and with mixed chi-square distribution HKY+G) (Hasegawa et al.,

1985). The *COI* data set had 148 parsimony-informative characters. The phylogenetic tree in Figure 2 shows the best-fitting tree topology based on the maximum likelihood analysis. Individuals of the same species formed distinct clusters, with the exception of *T. vitrirostris* and *T. mystax*. The latter two species showed a very close relationship and clustered together. This outcome does not agree with the taxonomic classification based on morphological characteristics.

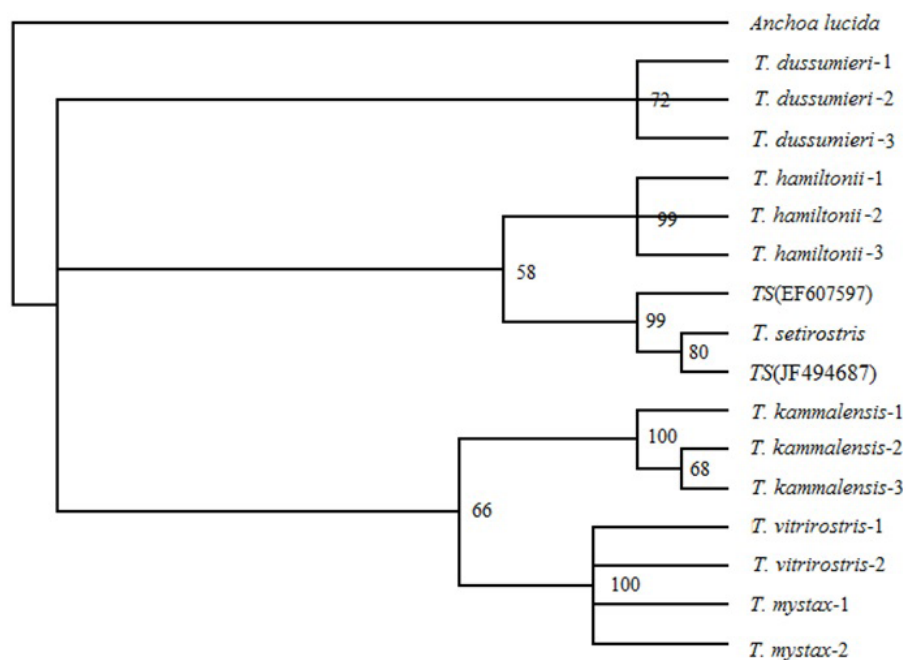


Figure 2. The maximum-likelihood (ML) phylogenetic tree of the haplotypes of the 14 specimens collected in six *Thyryssa* species and of the *T. setirostris* and *A. lucida* *COI* sequences obtained from GenBank.

DISCUSSION

Methods for rapid and reliable identification of species are important for fish research; sequence differences between species are a prerequisite for an effective identification method. Hebert et al. (2003) reported that interspecific genetic distances were greater than 0.02 in more than 98% of closely related species pairs, and that intraspecific genetic distances were less than 0.01. In the present study, we confirmed the utility of DNA barcoding for the identification of *Thyryssa* species. We found an average intraspecific distance of 0.002 and average interspecific distance of 0.137 for the six *Thyryssa* species. With the exception of *T. vitrirostris* and *T. mystax*, interspecific distances exceeded 0.02, with a range from 0.132 to 0.160. The interspecific genetic distances were much larger than intraspecific distances suggesting that the *COI* gene sequence was an appropriate marker for use in species identification. Morphological classification of the genus *Thyryssa* is based on the length of the maxilla. In *T. mystax* and *T. vitrirostris*, the maxilla stretches up to the pectoral fin base; other morphological characteristic, such as fin ray number, number of longitudinal scales, and rows of scales, similarly do not show clear-cut differences between the

species. In our *COI* gene sequence analysis, we obtained intraspecific K2P distances of 0.00 and 0.6% and an interspecific distance of 0.3%. Phylogeny of the genus *Thryssa* was analyzed using maximum likelihood methods, which indicated that four species formed monophyletic clusters, while *T. mystax* and *T. vitrostris* clustered together. This result was in accordance with a previous study (Ma et al., 2010b). Thus, we infer that *T. mystax* and *T. vitrostris* do not diverge at the species level. In conclusion, we confirmed that the *COI* gene is suitable for DNA barcoding in the genus *Thryssa*. This sequence will be of great value for future studies on the developmental characteristics of the various species of the genus *Thryssa*.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the National Basic Research Special Foundation of China (#2013FY110700), and National Non-Profit Institutes (East China Sea Fisheries Research Institute) (#2012T04), Special Fund for Agroscientific Research in the Public Interest (#201303047), and the National Infrastructure of Fishery Germplasm Resources.

REFERENCES

- Feng Y, Li Q, Kong L and Zheng X (2011). DNA barcoding and phylogenetic analysis of Pectinidae (Mollusca: Bivalvia) based on mitochondrial *COI* and 16S rRNA genes. *Mol. Biol. Rep.* 38: 291-299.
- Glover RH, Collins DW, Walsh K and Boonham N (2010). Assessment of loci for DNA barcoding in the genus *Thrips* (Thysanoptera: Thripidae). *Mol. Ecol. Resour.* 10: 51-59.
- Hasegawa M, Kishino H and Yano T (1985). Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22: 160-174.
- Hebert PDN, Ratnasingham S and deWaard JR (2003). Barcoding animal life: cytochrome *c* oxidase subunit I divergences among closely related species. *Proc. Biol. Sci.* 270: S96-S99.
- Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.
- Koch H (2010). Combining morphology and DNA barcoding resolves the taxonomy of Western Malagasy *Liotrigona* Moure, 1961 (Hymenoptera: Apidae: Meliponini). *Afr. Invertebr.* 51: 413-421.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, et al. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102: 8369-8374.
- Ma CY, Chen QQ, Zhang QY, Zhuang P, et al. (2010a). Genetic variation of *Coilia ectenes* (Clupeiformes: Engraulidae) revealed by the complete cytochrome *b* sequences of mitochondrial DNA. *J. Exp. Mar. Biol. Ecol.* 385: 14-19.
- Ma C Y, Ma LB, Ni Y, Shen AL, et al. (2010b). Phylogenetic relationships of *Thryssa* inferred from morphologic characteristic and mitochondrial 16S rRNA gene sequences. *J. Fish. Sci. China* 17: 79-85.
- Persis M, Chandra Sekhar Reddy A, Rao LM, Khedkar GD, et al. (2009). *COI* (cytochrome oxidase-I) sequence based studies of Carangid fishes from Kakinada coast, India. *Mol. Biol. Rep.* 36: 1733-1740.
- Posada D and Crandall KA (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Puillandre N, Strong EE, Bouchet P, Boisselier MC, et al. (2009). Identifying gastropod spawn from DNA barcodes: possible but not yet practicable. *Mol. Ecol. Resour.* 9: 1311-1321.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
- Schindel DE and Miller SE (2005). DNA barcoding a useful tool for taxonomists. *Nature* 435: 17.
- Steinke D and Hanner R (2011). The FISH-BOL collaborators' protocol. *Mitochondrial DNA* 22 (Suppl 1): 10-14.
- Swofford DL (2003). PAUP*. Phylogenetic analysis using parsimony (and other methods). Version 4. Sinauer Associates, Sunderland.

- Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA 4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
- Tang RWK, Yau C and Ng WC (2010). Identification of stomatopod larvae (Crustacea: Stomatopoda) from Hong Kong waters using DNA barcodes. *Mol. Ecol. Resour.* 10: 439-448.
- Teletchea F (2009). Molecular identification methods of fish species: reassessment and possible applications. *Rev. Fish Biol. Fish.* 19: 265-293.
- Wakabayashi T, Suzuki, N, Sakai M, Ichii T, et al. (2006). Identification of ommastrephid squid paralarvae collected in northern Hawaiian waters and phylogenetic implications for the family Ommastrephidae using mtDNA analysis. *Fish. Sci.* 72: 494-502.
- Ward RD, Hanner R and Hebert PDN (2009). The campaign to DNA barcode all fishes, FISH-BOL. *J. Fish Biol.* 74: 329-356.
- Yang R, Wu X, Yan P and Li X (2010). Using DNA barcodes to identify a bird involved in a birdstrike at a Chinese airport. *Mol. Biol. Rep.* 37: 3517-3523.
- Zhu YD (1984). The fishes of Fujian. Fujian Scientific and Technical Publishers, 137-146.