Bangladesh J. Fish. (2018) **30**(1) : 01-09

Identification of small indigenous fish species dhela (*Osteobrama cotio*) and darkina (*Esomus danricus*) using DNA barcoding technique

MD. AZIZUL HAQUE, FOYJUNNESA PROJNA AND MD. SAMSUL ALAM*

Department of Fisheries Biology and Genetics
Bangladesh Agricultural University, Mymensingh 2202, Bangladesh
*Email: samsul.alam@bau.edu.bd

Abstract. Species identification through DNA barcoding has been proved to be an effective method and gained remarkable popularity in taxonomic studies. The present study aimed at developing molecular technique for identification of two SIS species- Dhela (*Osteobrama cotio*) and Darkina (*Esomus danricus)*. For molecular identification of the species, a 650 bp fragment of the mitochondrial COI gene was amplified using FishF1/FishR1 universal primer set from five dhela and five darkina specimens. The PCR fragments were sequenced and were compared with the sequences of NCBI database and BOLD. The results showed a high level of similarities ranging from 98 to 100% with the sequence of known specimens of the two species available in the data base. The identifications have also been supported by the phylogeny tree where dhela samples formed one cluster and the darkina samples formed another cluster. Therefore, DNA bar coding technique can be used as an effective tool in the identification of adult, larvae or even eggs of *O. cotio* and *E. danricus*. In addition to this, the present study detected a very low level of haplotype and nucleotide diversity in *O. cotio* compared to *E. danricus*.

Keywords: DNA barcoding, mtDNA, *Osteobrama cotio, Esomus danricus*

Introduction

Fish account for approximately 60% of animal protein intake and it's frequency of consumption is far more than any other animal-source food in Bangladesh (Belton *et al*. 2014).The species that attain a maximum length of 25 cm are defined as small indigenous fish species (SIS) and in Bangladesh out of estimated 260 indigenous freshwater fish species, 143 have been classified as SIS. These species have been considered as an exceptional source of essential amino acids, vitamins and minerals, which play a significant role in the fulfillment of nutritional deficiency in human body. The existence of SIS in various water bodies including rivers, *beels, khals, haors* and *baors* of Bangladesh is steadily declining due to indiscriminate catching by various baleful fishing gears threatening SIS to a high risk of extinction.

Osteobrama cotio is a SIS and commonly known as dhela, (Bangla) or cotio (English). It is found in Bangladesh, Pakistan and India (Assam, West Bengal, Bihar, Madhya Pradesh and Punjab). It is one of the most important nutrient rich small fishes that contains 1.8 mg iron, 3.7 mg zinc, 1200 mg calcium and 918 μg Vitamin A per 100 g fresh edible tissue (Bogard *et al*. 2015).Once this fish was distributed broadly in the natural waters in the South and South East Asian countries, however, due to habitat destruction and over-exploitation availability of this fish in nature has been severely declined threatening it to be included in the vulnerable group (Chakraborty *et al*. 2003).

Darkina (*Esomus danricus*) is another SIS belonging to the group of flying barbs due to the presence of extremely long pectoral fin. It contains 12 mg iron, 4 mg zinc, 891 mg calcium and 660 μg Vitamin A per 100g of raw edible parts (Bogard *et al.* 2015). Like other SIS, the wild population of this species is declining as a result of habitat destruction and other ecological changes to their surroundings (Hossen *et al.* 2015).

Numerous studies have demonstrated the effectiveness of DNA barcoding in different animal groups. The method has been successfully used to identify birds (Hebert *et al.* 2004a), cryptic species of butterflies (Hebert *et al.* 2004b), mosquitoes (Besansky *et al.* 2003), fish, squids, bivalve molluscs, hoverflies and turtles (Blair *et al.* 2006) and various other species of vertebrates and invertebrates (Hebert *et al,* 2003). The COI barcode is also a powerful tool for identification of individually isolated fish eggs, larvae fillets, fins and seafood products (Ward *et al.* 2005, Hubert *et al.* 2008, Pegg *et al.*2006, Smith *et al.* 2008, Wong and Hanner 2008). Application of DNA barcoding for species authentication in markets and control of commercial landings could contribute to recognition of the real fishing targets and to the conservation and monitoring of fish resources of a specific region (Ardura *et al.* 2010, Zhang *et al.* 2017).

Considering human nutrition, ecological balance, sustainability of SIS and healthy aquatic environment, a comprehensive assessment of SIS is indispensable from every water resources. The present study was aimed at the identification of small indigenous fish species dhela and darkina using DNA barcoding. The molecular identification method would help identify the two SIS in the stomach contents of their predators thus help identify their potential predators.

Materials and Methods

Sample collection and DNA extraction

Five dhela (*Osteobrama cotio*) and five darkina (*Esomus danricus*) specimens were collected from an experimental pond of Faculty of Fisheries, Bangladesh Agricultural University, Mymensingh. Total DNA was extracted from caudal fin tissues using proteinase-K digestion, Phenol:Chloroform:Isoamylalcohol (25:24:1) extraction and isopropanol precipitation method (Islam *et al.* 2004).

PCR amplification

DNA barcoding region of approximately 650bp from the 5′ region of the mitochondrial COI gene was amplified using a pair of primers-

Fish F1: 5′-TCAACCAACCACAAAGACATTGGCAC-3′,

FishR1: 5′-TAGACTTCTGGGTGGCCAAAGAATCA-3′ (Ward *et al.* 2005) in 60-μl reaction volumes, containing 6 μl of 10 x Buffer, 0.25mM of each dNTPs, 2 μl of each primer (10 μM), 1 unit *Taq* polymerase (Takara, Japan and 150 ng of template

DNA. Thermal cycler parameters were set to 94°C for 3 min for an initial denaturation step, followed by 35 cycles each comprising of 94°C for 45 s, annealing at 46°C for 1 min, extension at 72°C for 1 min 30 s. Finally one cycle at 72°C was added for 7 min. PCR products were purified using commercial gel purification kit (Thermo Scientific, USA). DNA sequencing was performed on an ABI Genetic Analysis System using Big Dye Terminator Cycle Sequencing method from Macrogen, South Korea.

Statistical interpretation of the sequence data and identification of species

The raw nucleotide sequences of all the specimens were manually checked and edited (if necessary) using the software, BioEdit (Hall 1999). The nucleotide sequences were then aligned among themselves using the ClustalW software implemented in MEGA 7.0 (Kumar *et al.*2016). Sequence divergence among haplotypes were calculated using the Kimura-2-Parameter model (Kimura 1980).

For species identification, the examined sequences (Fasta file) were submitted to the GenBank and identifications of the specimens were confirmed based on identity with the COI gene sequence of known *O. cotio* and *E. danricus* voucher specimens using Basic Local Alignment Search Tool (BLAST) (Altschul *et al.* 1997). We also submitted the sequence to the BOLD (Barcode of Life Data System) for molecular identification of the species. A phylogeny tree was constructed using neighbor joining method (Saitou and Nei 1987) with 1000 bootstraps using MEGA version 7.0 software (Kumar *et al.* 2016) for revealing taxonomic relationship of the species.

The standard genetic diversity indices, such as mean nucleotide composition, polymorphic sites (Ps), the number of haplotypes (H), haplotype diversity (hd) and nucleotide diversity (Pi) and the mean number of pair-wise differences (k) for each species were calculated using DnaSP 5.10 (Librado and Rozas 2009). A median-joining network of the haplotypes was drawn using POPART (Population Analysis with Reticulate Trees).

Results

Molecular identification of O. cotio and E. danricus

The selected nucleotide sequence length of COI gene of all the 10 specimens, five dhela and five darkina, were 650bp.No insertions, deletions or stop codons were observed in any of the sequences. The lack of stop codons is consistent with all amplified sequences being functional mitochondrial *COI* sequences. Species identification based on the sequence similarity approach was carried out using two databases:GenBank (NCBI) and BOLD (Ratnasingham and Hebert 2007). The identity of *COI* sequences of five specimens of *O. cotio* (DHE-2, DHE-3, DHE-4, DHE-5, DHE-6) with that of voucher specimens available in the genbank ranged from 99% to100%; the identity of COI sequences of five specimens of *E. danricus* (DRK-1, DRK-2, DRK-3, DRK-5, DRK-6) with that of voucher specimens ranged from 99% to 100%(Table I).

Table I. Identity of the five *O. cotio* and five *E. danricus* sequences with the sequence of the voucher specimens of the respective species available in the GenBank and BOLD

SL No.	GenBank (NCBI)				BOLD-IDS	Inference
	Species	Similarity (%)	Total score	Accession No.	Similarity (%)	
DHE2	*O. cotio*	99	1219	KT762359.1	100	*O. cotio*
DHE3	*O. cotio*	99	1208	KT762359.1	100	*O. cotio*
DHE4	*O. cotio*	99	1210	KT762359.1	100	*O. cotio*
DHE5	*O. cotio*	99	1206	KT762359.1	99.84	*O. cotio*
DHE6	*O. cotio*	99	1184	KX550004.1	99.84	*O. cotio*
DRK1	*E. danricus*	99	1195	KJ936756.1	100	*E. danricus*
DRK2	*E. danricus*	99	1164	KJ936756.1	99.38	*E. danricus*
DRK3	*E. danricus*	99	1179	KJ936756.1	99.69	*E. danricus*
DRK5	*E. danricus*	98	1140	KJ936756.1	98.74	*E. danricus*
DRK6	*E. danricus*	99	1190	KJ936756.1	99.84	*E. danricus*

Phylogenetic tree

A phylogentic tree was constructed involving 650 bp sequences of five *O. cotio,* specimens and five *E. danricus* specimens. The sequence of the experimental specimens were compared with the sequence of the voucher specimens of *E. danricus* and *O. cotio* collected from NCBI GenBank. Branch length described the genetic changes through base substitutions. The evolutionary distances were calculated using the Kimura 2-parameter (K2P) method in the units of the number of base substitutions per site.The sequences have been grouped into two clusters, DHE-2, DHE-3, DHE-4, DHE-5, DHE-6 forming one cluster and DRK1, DRK2, DRK3, DRK5 and DRK6 formed another cluster (Fig.1).

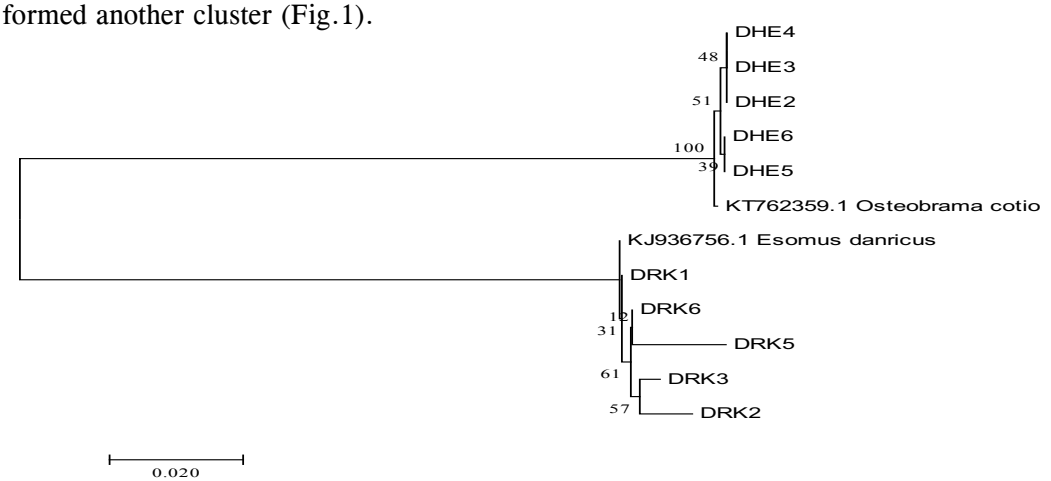

Fig.1. Phylogenetic tree of five *E. danricus* (DRK) and five *O. cotio* (DHE) sequences of the COI gene of mitochondrial genome. The figure shows the percentage of replicate trees in which the associated taxa are clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

Genetic variation in E. danricus and O.cotio samples

The genetic variations in the *E. danricus* and *O. cotio* samples were assessed on the basis of sequence information of COI gene of the mitochondrial genome. Between the two species the genetic variation of *E. danricus* is higher than that of *O. cotio* based on the number and diversity of haplotype, the number of polymorphic sites and nucleotide diversity and the average number of nucleotide difference (Table II).

Table II. Genetic variation in mitochondrial COI gene of *E. danricus* and *O. cotio*

Parameters	*E. danricus*	*O. cotio*
No. of haplotypes (h)	5	2
Haplotype diversity (Hd)	1.00±0.126	0.60±0.175
No. of polymorphic sites (S)	17	1
Nucleotide diversity (Pi)	0.01108±0.00325	0.00092 ± 0.00027
G + C Content (%)	0.418	0.446
Average No. of nucleotide difference (k)	7.20	0.60

Pair-wise genetic distance based on COI gene sequences

The pair-wise genetic distance of the five darkina specimens ranged from 0.002 to 0.022 while the pair-wise genetic distance of the five dhela specimens ranged from 0.000 to 0.002. Genetically, the dhela specimens were more uniform than the darkina specimens. The pair-wise genetic distance between the darkina and dhela specimens ranged from 0.194 to 0.212 (Table III) confirming that they belonged to two different species.

Table III. Individual pair-wise genetic distances (below diagonal) among darkina (DRK) and dhela (DHE) specimen based on Kimura 2-parameter (K2P) model for the 650 bp sequences of COI gene

	DRK6	DRK5	DRK3	DRK2	DRK1	DHE6	DHE5	DHE4	DHE3	DHE2
DRK6	-									
DRK5	0.014	-								
DRK3	0.006	0.020	-							
DRK2	0.008	0.022	0.011	-						
DRK1	0.002	0.016	0.005	0.009	-					
DHE6	0.196	0.210	0.200	0.206	0.194	-				
DHE5	0.196	0.212	0.200	0.208	0.194	0.000	-			
DHE4	0.198	0.212	0.202	0.208	0.196	0.002	0.002	-		
DHE3	0.198	0.212	0.202	0.208	0.196	0.002	0.002	0.000	-	
DHE2	0.198	0.212	0.202	0.208	0.196	0.002	0.002	0.000	0.000	-

The median-joining network of mtDNA COI sequences for *E. danricus and O. cotio* showed genealogical relationships among seven haplotypes based on the least number of substitutions. Out of seven haplotypes, hap_01 to hap_05 belonged to darkina and hap_06 and hap_07 belonged to dhela, the network comprised of two consequently connected sub-networks corresponding to *E. danricus* and *O. cotio* (Fig. 2).

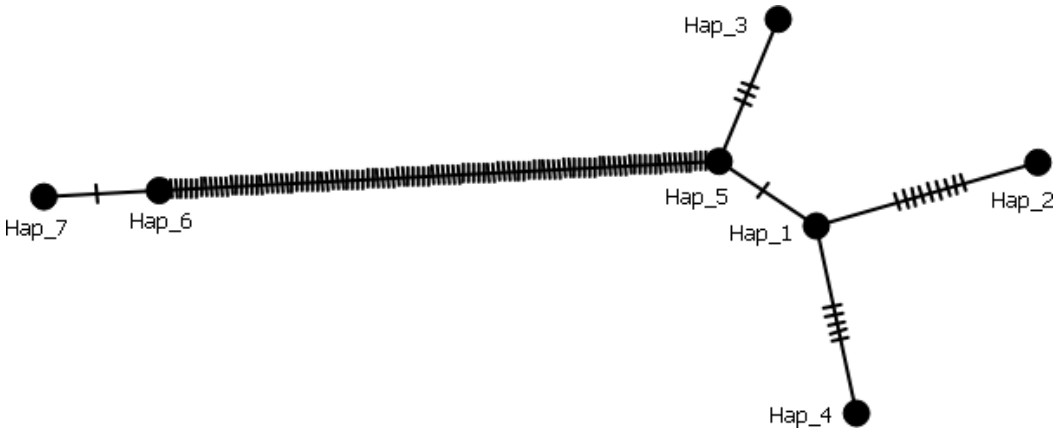

Fig. 2. Median-joining networks of mtDNA COI haplotypes of *E. danricus* and *O. cotio*. Each circle represents one unique haplotype and bars across branches indicates number of substitution separating two haplotypes. Haplotype 1-5 belong to *E. danricus* and 6-7 belong to *O. cotio*.

Discussion

Accurate and unambiguous identification of fish and fish products, from eggs to adults, is important in many areas. It would enable identification of fish products, assist in managing fisheries for long-term sustainability, and improve ecosystem research and conservation. In the past, a wide variety of protein- and DNA-based methods had been used for the genetic identification of fish species (see, for example, Ward and Grewe 1994). Here we have applied *COI* sequence analysis with the goal of determining whether DNA barcoding can achieve unambiguous species recognition. This study has strongly validated the efficacy of COIbarcodes for identifying fish species. We sequenced the barcode region of COI of five specimens of two SIS, *O. cotio* (Dhela) and *E. danricus* (Darkina). With no exceptions, all 10 sequenced specimens could be correctly identified based on information available in GenBank and BOLD-IDS exhibiting high identities (98-100%) (Table I).

The phylogenetic tree of COI sequences also revealed a particular species that formed a distinctive group (Fig. 1), Samples of *O. cotio* species formed a distinct group and the samples of *E. danricus* formed another distinct group. Both species worked as an out group of each other. None of the individuals of the two species crossed the out group barrier and included in the wrong group. Individual sample is clustered in

phylogeny branches grouped by their taxonomic affinity and thus individuals from the same species are clustered closely (Ardura *et al*. 2010).

Ward *et al*. (2008) opined that COI is comparatively a conserved gene within fishes and that's why most of the species show low levels of intra-specific variation. In present study *O. cotio* showed very low level intra-specific variation compared to *E. danricus* in terms of haplotype number, haplotype diversity, polymorphic sites and nucleotide diversities as well as number of pair wise differences (Table II). The very low level of genetic variation in the *O. cotio* samples indicate that the individuals are genetically more uniform. The very low level of pair-wise genetic distance (0.000 to 0.002) also supports the findings (Table III). It might be happened due to significant reductions in the effective breeding number. IUCN has already categorized *O. cotio* as a threatened species. So, attention should be given to conserve the vulnerable species by *in situ* and *ex situ* conservation techniques. On the other hand, *E. danricus* also showed low level of intra-specific variation ranging 0.002 to 0.022, though the value was comparatively higher than *O. cotio,* indicating *E. danricus* possess higher intra-specific genetic variability.

Aquatic resources are declining at a rapid rate due to both man-made and natural causes such as indiscriminate catching, habitat destruction, and environmental changes. SIS are the most significant resources of inland waters in Bangladesh and are suffering from high exploitation rates coupled with inadequate management. The standard DNA barcode sequence obtained in the present study can be an important reference for identifying *O. cotio* and *E. danricus* and also may be beneficial to fisheries management. Once a global COI barcode database has been established for fishes, anyone with direct or indirect access to a DNA sequence will be able to identify, to a high degree of certainty, any fish egg, larvae or carcass fragment. This will be an invaluable tool for fisheries managers, fisheries ecologists and fish retailers, and for those wishing to develop fish identification microarrays. The scientific and practical benefits of fish barcoding are manifold. Thus, it is anticipated that ecologists will increasingly turn to the DNA barcoding approach, because in different circumstances it represents the only easy way to identify species and this trend will be further enriched by the availability of trustworthy databases.

Literature Cited

Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman, 1997.Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.,* 25(17): 3389-3402.

Ardura, A., A.R. Linde, J.C. Moreira and E. Garcia-Vazquez, 2010. DNA barcoding for conservation and management of Amazonian commercial fish. *Biol. Conserv.,* 143: 1438–1443.

Belton, B., I.J.M. Van Asseldonk and S.H. Thilsted, 2014. Faltering fisheries and ascendant aquaculture: implications for food and nutrition security in Bangladesh. *Food Policy,* 44: 77–87.

Besansky, N.J., D.W. Severson and M.T. Ferdig, 2003. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends Parasitol.,* 19: 545–546.

Blair, D., M. Waycott, L. Byrne, G. Dunshea, S.C. Keune and K.M. Neil, 2006. Molecular discrimination of Perna (Mollusca: Bivalvia) species using the polymerase chain reaction and species-specific mitochondrial primers. *Mar. Biotechnol.,* 8: 380-385.

Bogard, J.R., S.H. Thilsted, G.C. Marks, M.A. Wahab, M.A.R. Hossain, J. Jakobsen and J. Stangoulis, 2015.Nutrient composition of important fish species in Bangladesh and potential contribution to recommended nutrient intakes. *J. Food Comp. Anal.,* 42: 120–133.

Chakraborty, B.K., M.I. Miah, M.J.A. Mirza and M.A.B. Habib, 2003. Rearing and nursing of local Sarpunti, *Puntius sarana* (Hamilton) at different stocking densities. *Pakistan J. Biol. Sci.,* 6(9): 797-800.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* 41: 95-98.

Hebert, P.D.N, E.H. Penton, J.M. Burns, D.H. Janzen and W. Hallwachs, 2004b. Ten species in one: DNA barcoding reveals cryptic species in the neo tropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. USA,* 101: 14812–14817.

Hebert, P.D.N., S. Ratnasingham and J.R. deWaard, 2003. Barcoding animal life: cytochrome coxidase subunit 1 divergences among closely related species. *Proc. Royal Soc. London. Series B: Biol. Sci.,* 270: S96–S99.

Hebert, P.D.N., M.Y. Stoekcle, T.S. Zemlak and C.M. Francis, 2004a. Identification of birds through DNA barcodes. *PLoS Biol.,* **2:** 1657–1663.

Hossen, M.A., M.Y. Hossain, K. Yahya and M.N.U. Pramanik, 2015. Threatened fishes of the world: *Labeo bata* (Hamilton, 1822) (Cypriniformes: Cyprinidae). *Croatian J. Fish.,* **73:** 89-91.

Hubert, N., R. Hanner, E. Holm, N.E. Mandrak, E. Taylor, M. Burridge, D. Watkinson, P. Dumt, A. Curry, P. Bentzen, J. Zhang, J. April and L. Bernatchez, 2008. Identifying Canadian freshwater fishes through DNA barcodes. *PLoS One,* 3(6)**:** 24-90.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J.Molec.Evol.,*16: 111-120.

Kumar, S., G. Stecher and K. Tamura, 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molec. Biol. Evol.,* **33:** 1870-1874.

Librado, P. and J. Rozas, 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25(11): 1451–1452.

Pegg, G.G., B. Sinclair, L. Briskey and W.J. Aspden. 2006. MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Sci. Mar.,* 70 (Suppl. 2): 7–12.

Ratnasingham, S. and P.D.N. Hebert, 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molec. Ecol., Notes,* **7**(3): 355–364.

Saitou, N. and M. Nei, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molec. Biol. Evol.,* 4(4): 406-425.

Smith, P.J., M.S. McVeagh and D. Steinke, 2008. Application of DNA barcoding for the identification of smoked fish products. *J. Fish Biol.,* **72:** 464–471.

Ward, R., T. Zemlak, B. Innes, P. Last and P.D.N. Hebert, 2005. DNA barcoding Australia's fish species. *Phil. Trans. Royal Soc. London B: Biol. Sci.,* 360: 1847–1857.

Ward, R.D. and P.M. Grewe, 1994. Appraisal of molecular genetic techniques in fisheries. *Rev. Fish Biol. Fish.,* 4: 300–325.

Ward, R.D., F.O. Costa, B.H. Holmes and D. Steinke, 2008. DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa, but *Zeus faber* and *Lepidopus caudatus* each probably constitute two species. *Aquat. Biol.,* **3:** 71–78.

Wong, E.H.K. andR.H. Hanner, 2008. DNA barcoding detects market substitution in North American seafood. *Food Res. Int.,* 41: 828–837.

Zhang, Y.H., G. Qin, H.X. Zhang, X. Wang and Q. Lin, 2017. DNA barcoding reflects the diversity and variety of brooding traits of fish species in the family Syngnathidae along China's coast. *Fish.Res.,*185: 137–144.

(Manuscript received 15 March 2018)