Short Communication

Length variation and sequence divergence in mitochondrial control region of Schizothoracine (Teleostei: Cyperinidae) species

Mudasir Ahmad Syed¹, Farooz Ahmad Bhat², Masood-ul Hassan Balkhi², and Bilal Ahmad Bhat²

¹Division of Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, Shuhama, Sher-e-Kashmir University of Agricultural Sciences and Technology, Kashmir, India and ²Faculty of Fisheries, Rangil, Sher-e-Kashmir University of Agricultural Sciences and Technology – Kashmir, Shuhama Srinagar, India

Abstract

Schizothoracine fish commonly called snow trouts inhibit the entire network of snow and spring fed cool waters of Kashmir, India. Over 10 species reported earlier, only five species have been found, these include Schizothorax niger, Schizothorax esocinus, Schizothorax plagiostomus, Schizothorax curvifrons and Schizothorax labiatus. The relationship between these species is contradicting. To understand the evolutionary relation of these species, we examined the sequence information of mitochondrial D-loop of 25 individuals representing five species. Sequence alignment showed D-loop region highly variable and length variation was observed in di-nucleotide (TA), microsatellite between and within species. Interestingly, all these species have (TA)n, microsatellite not associated with longer tandem repeats at the 3’ end of the mitochondrial control region and do not show heteroplasmy. Our analysis also indicates the presence of four conserved sequence blocks (CSB), CSB-D, CSB-1, CSB-II and CSB-III, four (Termination Associated Sequence) TAS motifs and 15bp pyrimidine block within the mitochondrial control region, that are highly conserved within genus Schizothorax when compared with other species. The phylogenetic analysis carried by Maximum likelihood (ML), Neighbor Joining (NJ) and Bayesian inference (BI) generated almost identical results. The resultant BI tree showed a close genetic relationship of all the five species and supports two distinct grouping of S. esocinus species. Besides the species relation, the presence of length variation in tandem repeats is attributed to differences in predicting the stability of secondary structures. The role of CSBs and TASs, reported so far as main regulatory signals, would explain the conservation of these elements in evolution.

Keywords

CSB, D-loop, length variation, microsatellites, Schizothoracine fishes, species variation, TAS

Introduction

The schizothoracine fishes belong to schizothoracinae, cyprinidae and cypriniformes, comprise 90 species (subspecies) and 12 genera. According to Cao et al. (1981), the subfamily Schizothoracinae originated from primitive barbine fish distributed in Tibet during the late Tertiary, and subsequently evolved into more specialized groups owing to the uplift of Tibet. The fish are believed to have migrated into the water of the Kashmir region from Central Asiatic south slopes of the Himalayas and Sulaiman range (Das & Subla, 1964). However, no palaeogeographical hypothesis has been proposed to account for the distribution pattern of Schizothorax, based on dispersal and vicariance. There is great confusion in the number of species present. Heckel (1838) for the first time reported 10 species of Schizothoracines from Kashmir. Silas (1960), however, framed the list of 12 species.

The morphometric characteristics to classify fishes of this region are highly contradicting, it is often difficult to decide whether these are different species, different phenotypes of single species or an intermediate situation between the two extremes (Raina & Petr, 1999). Based on recent survey by Talwar & Jhingran (1991), Kullander et al. (1999), Yousuf et al. (1996) and Balkhi (2005), there are only five recognized species within the genus Schizothorax from the Kashmir, India; these include Schizothorax esocinus (Heckel, 1838), Schizothorax plagiostomus (Heckel, 1838), Schizothorax labiatus (McClelland, 1842), Schizothorax niger (Heckel, 1838) and Schizothorax curvifrons (Heckel, 1838). The uncertainty surrounding the phylogenetic relationship and biogeography of this group continues to the present day. During the last few years, the mitochondrial control region has been used to specifically study population structure, gene flow, migration, phylogeography and female nesting behavior assessment (Norman et al., 1994).

In this study, besides molecular phylogeography of the Schizothorax species, we report the special structural features of a microsatellite sequence within the mitochondrial control region of these species. To the best of our knowledge, this study is the first approach to investigate the role of mitochondrial microsatellite as an important marker to study inter- and intraspecific variation within schizothoracines.
Materials and methods

Taking into account potential problems with the identification of species (Talwar & Jhingran, 1991), all the five *Schizothorax* species were properly identified using the taxonomies of Jhingran (1991). Five individuals per species collected from different localities were subjected for DNA isolation (Table S1). DNA was isolated from muscle tissue following standard proteinase K, phenol/chloroform extraction protocol (Sambrook et al., 1989) D-loop was amplified using primers and PCR conditions of Thai et al. (2004).

Sequences were aligned using MUSCLE v3.5 (Edgar, 2004) and edited with MacClade v4.06 (Maddison & Maddison, 2003). The unreadable sequences were removed from the analysis. Maximum-likelihood (ML), neighbor-joining (NJ) and maximum parsimony (MP) methods were implemented using PAUP* version 4.0b10 (Swofford, 2002); and the Bayesian method was carried out using MrBayes 3.0 (Huelsenbeck & Ronquist, 2001). Taking into account the results of He & Chen (2006), *Barbus barbus* (NC_008654) was added as an outgroup for phylogenetic analyses.

We considered nodes with bootstrap values ≥70% (Hillis & Bull, 1993) and Bayesian posterior probabilities ≥95% to be well-supported. Heuristic searches used for ML analyses utilized 100 replicates of random sequence additions followed by non-parametric bootstrapping comprising 100 replications with 10 random sequence additions. The NJ tree was constructed with distances calculated under the same model of evolution as the ML analysis with bootstrapping performed using 1000 replicates. Bayesian analyses were performed with random starting trees and run for 1.0 × 10^6 generations and sampling the four Markov chains every 100 generations resulting in 10,000 trees. The likelihood scores of the sampled trees were plotted against generation time to ensure that stationarity was reached; trees generated prior to stationarity were discarded as ‘burn-in’ (2000 trees in this case). Bayesian posterior probabilities of each bipartition, representing the percentage of times each node was recovered, were calculated from a 50% majority rule consensus of the remaining trees.

Result and discussion

Sequence alignment of control region showed length variation ranged from 785 to 793 bp in all the five species under study. When compared with sequences deposited in the GenBank, the length of completed-loop gene in *Schizothorax* genus was found to be 933 bp (*Schizothorax macropogon*, KC020113). The sequence composition and arrangement of the repeats varied considerably between and within species. Interestingly, the length variation was observed in d-di-nucleotide (T)n microsatellite with a variable number of repeat units (n = 7–14). These species had a (T)n microsatellite not associated with longer tandem repeats in the 3' end light (L)-strand in of the mitochondrial control region. Comparison of repeat array lengths among five species of schizothracines reveals that the model number of microsatellite repeats varied from seven in *S. esocinus* to 14 in *S. curvifrons* species (Table S2). It is unusual to find a microsatellite not associated tandem repeats in the mitochondrial control region with different inter- and intraspecific variations. Hoelzel et al. (1993), for the first time, reported (AC)n GT microsatellite not associated with longer tandem repeats in the 3' end of the mitochondrial control region showing extensive heteroplasy with up to three length variants present in single individuals of two elephant seal species. However, in our case, we did not find heteroplasy in D-loop of the species. Heteroplasy for length variation has been frequently observed (reviewed in Rand, 1993). The exact molecular mechanisms that create variations are not completely understood, but length changes in microsatellite DNA are much recurrent criticism of replication slippage – that is, transient dissociation of the replication DNA strands followed by misaligned reassociation (Rochej et al., 1990). Mitochondrial DNA length variations, caused by tandem repeats, have previously been identified in a number of fish species: sturgeon (e.g. *Acipenser transitor montanus*, *Acipenser oxyrhynchus desotoi*; Brown et al., 1992) cod (*Gadus mor populor hau*; Johansen et al., 1990), European sea bass (*Dicentrarchus labrax*; Cecconi et al., 1995), American shad (*Alosa sapidissima*; Bentzen et al., 1988), and Leuciscininae (*Pisces: Cyprinid; Sasaki et al., 2007*).

The nucleotide sequence alignment also allowed the identification of four conserved sequence blocks (CBS-D, CBS-I, CBS-II, and CBS-III). The sequence of conserved block D (CBS-D) is highly conserved within the family Cyprinidae and showed more than 85% of the total identity belonged to other teleost (Table 1). According to Southern et al. (1988), CBS-D is most universally conserved segment among fish families, suggesting that it contains functions critical for mitochondrial metabolism. Although a number of different approaches have been tried (Mignotte et al., 1987), the function of this central conserved region is not understood. Next to CBS-D, GTGGG-box, which is common to euteleosts, has also been identified (Figure S1). Compared to other teleost, all the three CBSs (CBS-I, CBS-II and CBS-III) were found similar within the genus *Schizothorax*. Comparing these to the available CBS sequences from representative teleost, we identified the following key features. CBS-1 has an A/T rich region followed by a similar TCAAGTGCAATA motif. The motif of high similarity (CBS-1) seemed to be commonly distributed among fish species. Unlike other studies, CBS-3 is found more conserved than CBS-2 within the family Cyprinidae. CBS-2 is characterized by a poly-C stretch separated by TA and CBS-3 containing a sequence of three ‘A’s followed by a poly-C stretch. These CSBs are thought to be involved in the formation of proper RNA primer for mtDNA replication and play a role in the switch from RNA to DNA synthesis that commences at O_H (Clayton, 1991).

In addition, we found three TAS (Termination Associated Sequence) motif – TACAT – in 5' part of control region and one in the central domain (Figure S1), Brzuzab & Ciesielski (2002), however, reported all the four similar TAS mitofs in 5' D-loop of three coregonine species. Up to three termination associated sequences (TASs) involved in the premature termination of the H-strand replication (Doda et al., 1981) as well as several copies of the conserved palindromic motif, 5'-TACAT-3' (Saccone et al., 1991) were found at the 5' end of the control region. However, no secondary structures were identified in the left region of D-loop. We also identified a pyrimidine block of 15 bp, between CBS-D and CBS-I. A 15 bp pyrimidine block was identified in the Atlantic cod (Johansen et al., 1990) and 26 blocks in the rainbow trout (Digby et al., 1992). This site may provide a point of interaction with mtDNA single stranded-binding protein (Digby et al., 1992).

The sequence divergence and mutation analysis between the species was calculated using PAUP. The D-loop was found AT rich and 5' end was found conserved than 3' end. The mean uncorrected pairwise genetic distance between species was 20.22%. Of 789 total characters, 638 (80%) characters were constant, 89 (11%) variable characters were parsimony uninformative, and 62 (7.8%) variable characters were parsimony informative. The proportion of invariable sites was estimated to be zero and the shape of the gamma parameter was estimated to be 0.573. The estimated transition/transversion bias (R) was 4.44. The ML, NJ and BI methods of analysis generated almost identical relationships. The likelihoods of trees, optimized under the topologies obtained by the ML, MP and BI methods did
not differ significantly. Because of this congruence, only the BI tree is presented with posterior probabilities following maximum likelihood bootstraps (Figure 1). The best-fit model for the D-loop data was GTR + G + I. The model contained the nucleotide substitution estimated nucleotide frequencies $A = 0.32$, $C = 0.19$, $G = 0.14$, and $T = 0.33$; nucleotide; nucleotide substitution rate matrix $A$–$C = 0.40$, $A$–$G = 4.6$, $A$–$T = 0.44$, $C$–$G = 0.21$, and $C$–$T = 2.6$.

The resulted BI tree (Figure 1) clearly indicates two different groups of *S. esocinus* and *S. plagiostomus* populations. One group of population nests with *S. labiatus* and shows a close relationship with *S. niger* and *S. curvifrons*. Other groups of *S. esocinus* and *S. plagiostomus* populations clustered separately. The result based on the D-loop also supports the high genetic similarity between *S. niger* and *S. curvifrons* species. The finding that some populations of *S. plagiostomus* and *S. esocinus* cluster together may indicate hybridization between the species. The lack of difference in the mitochondrial sequence data of *Schizothorax* species may be explained in terms of introgressive hybridization, incomplete lineage sorting, rapid radiation in lineages and multiple hits (homoplasy) (He & Chen, 2006). Silas (1960) reported that the inter-specific hybridization in nature takes place to a greater extent among the Schizothoracinae and easily recognizable combinations of hybrids resulting from *Schizothorax esocinus* × *Schizothorax* spp. and *Oreinus plagiostomus* × *Schizothorax* spp. Since Heckel (1838) erected the genus *Schizothorax* (species with four barbels and with or without labial papillae described from the Kashmir Valley), the taxonomy of the group has been confused and disputed. Silas (1960) further reported that the differences between *S. curvifrons* and *S. niger* are not sufficiently distinct to justify their treatment as separate species; therefore, the latter is treated as a subspecies of *S. curvifrons*. The BI phylogenetic tree (Figure 1) depicts that *S. niger* species might have evolved from the *S. curvifrons* relatively recent. Our results based on D-loop analysis support the complexity of relationship within schizothoracine species found India.

This study suggests that the length variation in microsatellite di-nucleotide (TA)$_n$ repeats at the 3′ end of D-loop may be a potentially informative molecular marker for studying population structure and genetic diversity as well as conservation practices of fish species. Further, the presence of genus-specific conserved sequence blocks (CSB) might be usual for identification and explain different conversation mechanisms of these elements in evolution. Investigating more schizothoracine populations might therefore further elucidate the mutation mechanisms and

| Table 1. Comparison of conserved sequence blocks (CSB) of Schizothoracine species with other teleost. |
|---------------------------------------------------------------|
| **Element** | **Species** | **Sequence** | **Position** |
| CSB D | Coregonus sp. | TACTTGCACTCTGGTTCC | 421–438 |
| | Oncorhynchus mykiss$^2$ | TCCTTGGCAATTTGGTTCC | 447–463 |
| | Cyprinus carpio | TACTGCGACATCTGGTTCC | 349–366 |
| | Schizothorax macropogon* | TACTGCGACATCTGGTTCC | 348–356 |
| | S. niger | TACTGCGACATCTGGTTCC | 348–356 |
| | S. plagiostomus | TACTGCGACATCTGGTTCC | 348–356 |
| | S. curvifrons | TACTGCGACATCTGGTTCC | 348–356 |
| | S. esocinus | TACTGCGACATCTGGTTCC | 348–356 |
| | S. labiatus | TACTGCGACATCTGGTTCC | 348–356 |
| CSB1 | Coregonus sp. | ATATTTATATCAGTCGATA | 726–742 |
| | Oncorhynchus mykiss$^2$ | ATACTTGGAATCAAGTCGATA | 727–749 |
| | Cyprinus carpio | ATACTTTATATCAGTCGATA | 612–634 |
| | Schizothorax macropogon* | ATATTTTTAAGTCAAGTCGATA | 613–636 |
| | S. niger | ATATTTTTAAGTCAAGTCGATA | 613–636 |
| | S. plagiostomus | ATATTTTTAAGTCAAGTCGATA | 613–636 |
| | S. curvifrons | ATATTTTTAAGTCAAGTCGATA | 613–636 |
| | S. esocinus | ATATTTTTAAGTCAAGTCGATA | 613–636 |
| | S. labiatus | ATATTTTTAAGTCAAGTCGATA | 613–636 |
| CSB2 | Coregonus sp. | TTAACCCCTATCCCCC | 750–768 |
| | Oncorhynchus mykiss$^2$ | TTAACCCCTATCCCCC | 811–828 |
| | Cyprinus carpio | CAAACCCCTATCCCCC | 693–710 |
| | Schizothorax macropogon* | CAAACCCCTATCCCCC | 697–715 |
| | S. niger | CAAACCCCTATCCCCC | 697–715 |
| | S. plagiostomus | CAAACCCCTATCCCCC | 697–715 |
| | S. curvifrons | CAAACCCCTATCCCCC | 697–715 |
| | S. esocinus | CAAACCCCTATCCCCC | 697–715 |
| | S. labiatus | CAAACCCCTATCCCCC | 697–715 |
| CSB3 | Coregonus sp. | GTCAAAACCCGAAACC | 795–810 |
| | Oncorhynchus mykiss$^2$ | GTTAAACCCCTAAAC | 855–870 |
| | Cyprinus carpio | GTCAAAACCCGAAACC | 736–752 |
| | Schizothorax macropogon* | GTCAAAACCCGAAACC | 741–757 |
| | S. niger | GTCAAAACCCGAAACC | 741–757 |
| | S. plagiostomus | GTCAAAACCCGAAACC | 741–757 |
| | S. curvifrons | GTCAAAACCCGAAACC | 741–757 |
| | S. esocinus | GTCAAAACCCGAAACC | 741–757 |
| | S. labiatus | GTCAAAACCCGAAACC | 741–757 |

The CSB sequences for superscript 1 and 2 were reported by Brzuzab & Ciesielski (2002) and asterisk represents the Genbank sequences (*Schizothorax* esocinus, accession no. KC020113: *Cyprinus carpio carpio*, accession no. JN105352).
complexity of relationship for this geographically isolated area to lead more conclusive results. The use of more mitochondrial and nuclear markers for each the species could help resolve the species relation.

**Declaration of interest**

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. This work was supported by the Department of Science and Technology, New Delhi, India under Project No. SR/SO/AS-15/2010.

**References**

Balkhi MH. (2005). Fish diversity of Jammu and Kashmir and conservation. In: Patloo RA, editor. Kashmir speaks. Kashmir, India: G. M. Publishers. p 104–18.

Bentzen P, Leggett WC, Brown GG. (1988). Length and restriction site heteroplasmy in the mitochondrial DNA of American Shad (Alosa sapidissima). Genetics 118:509–18.

Brown JR, Beckenbach AT, Smith MJ. (1992). Mitochondrial DNA length variation and heteroplasmy in populations of white sturgeon (Acipenser transmontanus). Genetics 132:221–8.

Brzuzan P, Ciesiels S. (2002). Sequence and structural characteristics of mtDNA control regions of three coregonine species (Coregonus albula, C. lavaretus and C. peled). Adv Limnol 57:11–20.

Cao WX, Chen YY, Wu YF, Zhu SQ. (1981). Origin and evolution of schizothoracine fishes in relation to the upheaval of the Xizang Plateau. In: Tibetan Expedition Team of the Chinese Academy of Sciences, editors. Studies on the period, amplitude and type of the uplift of the Qinghai-Xizang Plateau. Beijing: Science Press. p 118–30.

Cecconi F, Giorgi M, Mariottini P. (1995). Unique features in the mitochondrial D-loop region of the European sea bass Dicentrarchus labrax. Gene 160:149–55.

Clayton DA. (1991). Nuclear gadgets in mitochondrial DNA replication and transcription. Trends Biochem Sci 16:107–11.

Das SM, Subla BA. (1964). The ichthyofauna of Kashmir: Part II. The speciation of Kashmir fishes. Ichthyologica 3:57–62.

Digby JT, Gray MW, Lazier CB. (1992). Rainbow trout mitochondrial DNA: Sequence and structural characteristics of the non-coding control region and flanking IRNA genes. Gene 113:191–204.

Doda JN, Wright CT, Clayton DA. (1981). Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc Natl Acad Sci USA 78:6116–20.

Edgar RC. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res 32:1792–7.

He D, Chen Y. (2006). Biogeography and molecular phylogeny of the genus Schizothorax (Teleostei: Cyprinidae) in China inferred from cytochrome b sequences. J Biogeogr 33:1448–60.

Heckel JJ. (1838). Fische aus Caschmir, gesammelt und herausgegeben von Carl Freiherrn von Hügel, beschrieben von Joh. Vienna: Mechita Risten.
Hillis DM, Bull JJ. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 42:182–92.

Hoelzel AR, Hancock JM, Dover GA. (1993). Generation of VNTRs and heteroplasmy by sequence turnover in the Valdes mitochondrial control region of two elephant seal species. J Mol Evol 37:190–7.

Huelsenbeck JP, Ronquist F. (2001). MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–5.

Johansen S, Guddal PH, Johansen T. (1990). Organization of the mitochondrial genome of Atlantic cod, Gadus morhua. Nucleic Acids Res 18:411–19.

Kullander SO, Fang F, Delling B, Ahlander E. (1999). The fishes of the Kashmir valley. In: Nyman L, editor. River Jhelum, Kashmir valley, impact on the aquatic environment. Sweden: Swedmar. p 99–162.

Maddison DR, Maddison WP. (2003). MacClade 4.06: Analysis of phylogeny and character evolution. Sunderland, MA: Sinauer Associates.

McClelland J. (1842). On the fresh water fishes collected by William Griffith, Esq. E.L.S. Madras Medical Service, during his travel under the orders of supreme Governor of India, from 1835 to 1842. Calcutta J Nat Hist 2:560–89.

Mignotte B, Dunon-Blutean D, Reiss C, Mounolou JC. (1987). Sequence deduced physical properties in the D-loop region common to five vertebrate mitochondrial DNAs. J Theor Biol 124:57–69.

Norman JA, Moritz C, Limpus CJ. (1994). Mitochondrial DNA control region polymorphisms: Genetic markers for ecological studies for marine turtles. Mol Ecol 3:363–73.

Raina HS, Petr T. (1999). Coldwater fish and fisheries in the Indian Himalayas: Lakes and reservoirs. In: Petr T, editor. Fish and fisheries at higher altitude: Asia. FAO Fish. Tech. Pap. No. 385. Rome: FAO. p 64–88.

Rand DM. (1993). Endotherms, ectotherms, and mitochondrial genome-size variation. J Mol Evol 37:281–95.

Rochej LA, Snyder M, Cook I, Fuller K, Zouros E. (1990). Molecular characterization of a repeat element causing large-scale variation in the mitochondrial DNA of the sea scallop, Placopecten magellanicus. Mol Biol Evol 7:45–64.

Saccone C, Pesole G, Sbisfi E. (1991). The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. J Mol Evol 33:83–91.

Sambrook J, Frisch EF, Maniatis T. (1989). Molecular cloning: A laboratory manual. 2nd ed. New York: Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press.

Sasaki T, Kartavtsev YP, Chiba SN, Uematsu T, Sviridov VV, Hanzawa N. (2007). Genetic divergence and phylogenetic independence of Far Eastern species in subfamily Leuciscinae (Pisces; Cyprinidae) inferred from mitochondrial DNA analyses. Genes Genet Syst 82:329–40.

Silas EG. (1960). Fishes from the Kashmir valley. J Bombay Nat Hist Soc 57:66–77.

Southern SO, Southern PJ, Dizon AE. (1988). Molecular characterization of a cloned dolphin mitochondrial genome. J Mol Evol 28:32–42.

Swofford DL. (2002). PAUP*: Phylogenetic analysis using parsimony. Ver. 4.0b10. Sunderland, MA: Sinauer Associates.

Thai BT, Burridge CP, Pham TA, Austin CM. (2004). Using mitochondrial nucleotide sequences to investigate diversity and genealogical relationships within common carp (Cyprinus carpio). Animal Genet 36:23–8.

Yousuf AR. (1996). Fishery resource of Kashmir. In: Khan AH, Pandit AK, editors. Ecology, environment and energy. Kashmir, India: Kashmir University. p 75–120.

Supplementary material available online

Supplementary Figure S1, Tables S1 and S2