Available online at www.scholarsresearchlibrary.com

Scholars research library

Archives of Applied Science Research, 2011, 3 (6):437-442
(http://scholarsresearchlibrary.com/archive.html)

ISSN 0975-508X
CODEN (USA) AASRC9

Nucleotide Sequences Variation of *Osteobrama* (Heckel) Freshwater Fish Species of North-East India Based On Mitochondrial *Cox I* Gene

Konthoujam S. Kumar* and Umesh C. Goswami

Department of Zoology, DST-FIST & UGC-SAP Sponsored Department Centre for Aquaculture and aquatic resources monitoring centre, Biodiversity & Bioinformatics. Gauhati University, Guwahati, Assam, India

---

ABSTRACT

cytochrome c oxidase subunit I (cox1) of 655 bp length is used as a barcode region of Osteobrama, Cyprinids (Teleostei:Cypriniformes:Cyprinidae). Cox1 nucleotide sequences for analysis of species identification comes from the degeneracy nature of the genetic code and also; the nature of the highly variable of the amino acids at the third codon position. The most variable region is dominated by amino acids with four or six codons, while the least variable region is dominated by amino acids with two codons. Therefore, cox1 amino acids sequences diversity is less than the nucleotide sequences diversity. Nucleotide sequence of Osteobrama species of 655 bp for cox1, were sequenced. All the species showed different cox1 nucleotide sequences and hence successfully barcode the region. The average genetic distance within the species ranges from 0% to 22% with the overall mean diversity of 18%

Keywords: cytochrome *c* oxidase subunit I, barcode, genetic code, degeneracy, *Osteobrama* sp.

INTRODUCTION

It has been proposed that a single gene sequence of the cytochrome *c* oxidase subunit I gene could be serve as the basis for a global identification system for animals [1]. The suggestion was that each species would be delineated by a particular sequence or a tight cluster of very similar sequences. The first reading frame region upto 655 base pair of vertebrate cox1 of 1545 total length base pair has been nominated as a barcode region. Support for the barcoding concept ranges from invertebrates (springtails and butterflies) up to birds. Studies in barcoding have now been published for a diverse array of animals like amphibians [2], ants [3], birds [4], collembolans [5], fishes [6], flies [7], moths and butterflies [8; 9] and Spiders [10]. For barcoding to succeed in species identification, the nucleotide sequences within the species need

to be more similar to one another than the sequences in different species. In particular, the mitochondrial gene cytochrome *c* oxidase subunit I can serve as a uniform target gene for a bio identification to complete the taxonomic data and global validation of systemic position, phylogeny and food traceability in trade monitoring.

In this study we examined whether the cox1 nucleotide sequences can discriminate freshwater fish *Osteobrama* species. Here we analyse the DNA sequence and translated protein sequence of the barcode region of *Osteobrama* species. In protein coding, the degeneracy nature of the genetic code means third codon base usually evolves faster than the first base, which in turn evolves faster than the second base [11]. Previous studies have shown the occurance of about 72% changes in third position are synonymous and do not lead to amino acids changes, compared with about 5% in first position and 0% for the second position [12]. Therefore it has been expected that the third base position variability would provide DNA barcoding with its power [1]; here we show this is indeed so by assessing levels of nucleotide sequences variability across the species at the three codon regions of the barcode region of cox1 gene and also at the translated amino acids level.

MATERIALS AND METHODS

Osteobrama is a riverine fish, body short, deep and compressed laterally, scales moderate with complete lateral line. The fish is not an economically important because of its small in size but it is consumed in rural areas and local area where the fish is found. The distribution area of the species varies according to the type of species - *Osteobrama belangeri* in Manipur (Chindwin basin) and Myanmar; *Osteobrama cunma* found in Manipur valley and its tributaries (Chindwin drainage) and Myanmar; *Osteobrama feae* was reported from Manipur (Maklang river) and Myanmar and *Osteobrama cotio* distribution extends from Assam (Brahmaputra drainage), Manipur (Barak-Brahmaputra drainage), West Bengal, Madhya Pradesh, Punjab, Uttar Pradesh, Pakistan and Bangladesh [13]

Tissue samples were isolated from the collected fish samples and preserve in 80-90% alcohol until used. All the fish's species were identified morphologically according to the Fishes of North-East India [14]

DNA Isolation:
Genomics DNA was isolated using a DNeasy tissue kit (Qiagen, Hidden, Germany) following the manufacturer's protocols.

Primer and PCR conditions and sequencing:
We used mitochondrial markers, Cytochrome *c* oxidase subunit 1 (*cox1*) 655 base pairs (bp). All the PCR reactions were carried out with 1X Taq Buffer A, 1.5 mM MgCl2, 0.25 mM dNTP mix (Eppendrof, Hamburg, Germany), 5U Taq DNA Polymerae (Bangalore Genei), 0.25 µM of each primer (Sigma-Aldrich Chemicals, Bangalore, India) and about 1-1.5 µl DNA template. The primers were used in the study

Primer F1 CCTTTATCTTGTATTCGGTGC
Primer R1 CTACTGATGCTCCGGCGTG

PCR amplification was performed at an initial denaturation of 94 ºC for 10 minutes, followed by 25 cycles at annealing temperature of 94 ºC for 30 seconds, 55 ºC for 30 seconds, 72 ºC for 30 seconds and 72 ºC for 7 minutes and stored at 4 ºC.

The PCR product was visualized through 2% agarose gel electrophoresis and it was seen that the fragments of the gene amplified is approximately 655 bp which is similar with the fragments length of the primer.

Each sample was sequenced, bidirectionally, for 655 bp of the *cox1* barcode region. Sequences were aligned using ClustalW software. Translation to amino acids sequences used the vertebrate mitochondrial option of MEGA 4 [15]. MEGA 4 was also used for calculating nucleotide genetic distances within the studied *Osteobrama* species using the Kimura2-parameter (K2P) distance model [16]. Neighbour-joining [17], Minimum Evolution [18] and Maximum Parsimony [19] trees using Kimura2-parameter distances were created to provide a graphical representation of the pattern of divergence between species using *Mastacembelus armatus* (eel) as an outgroup. To verify the robustness of the internal nodes, bootstrap analysis was carried out using 1,000 pseudo replicates [20].

The overall transition/transversion ratio (*R*) for *cox1* sequence data was estimated at 1.8.

RESULTS AND DISCUSSION

A data set of 13 *cox I* gene sequences of the family cyprinidae belonging to genus *Osteobrama* was obtained. The primers amplified the particular target region of all the species. The sequence divergences in all the species based on the pairwise analysis of sequences ranges from 0% to 22% between the *Osteobrama* species with the overall mean diversity of 18% using the Kimura 2-parameter model in MEGA 4. The maximum genetic divergences were occurred between the *Osteobrama belangeri* and *Osteobrama cunma* (22%) whereas the minimum genetic distance was observed between the *Osteobrama feae* and *Osteobrama cotio* (13%). No insertion/deletion or stops codon were found, supporting the view that all of the amplified sequences constitute functional mitochondrial *cox1* sequences.

Considering a total of 623 bp for the present analysis, 416 characters were conservative, 206 were variable and 148 were phylogenetically informative under parsimony. The three different tree building method (NJ, ME and MP) constructed using Kimura2-parameter distance shows same tree topology with similar bootstrap values. The species *Osteobrama cunma* being having longest branch length consider the most ancestral of all species considered while the *Osteobrama belangeri* forms a separate sister clade indicating the species highly diverge from the rest of the *Osteobrama* species of north east India (Fig 1). The different nucleotide composition content at all the codon position ranges from 7.9% guanine at 3rd codon position to 42% of thymine at 2nd codon position. In the first codon position, the proposition of guanine (30.3%) is highest and the thymine proposition (18%) is lowest. In second codon position, the proposition of thymine (42%) is highest and guanine proposition (13.9%) is least and finally in third codon position the proposition of adenine (40.7%) and the proposition of guanine (7.9%) were highest and least respectively. Considering all the codon position, the mean base composition of cox 1 nucleotide of all the taxa were 28.4% thymine, 27.2% Cytosine, 28.4 % adenine and 17.4% guanine. The

average amino acids frequency of all the species ranges from 0.97% (glutamic acid) to 16.50% (leucine), the others concentration of amino acids ranges from 3.40% Asparagine to 9.82% Alanine. The most variable region of amino acids is code by six or four codons, Leucine and Serine is code by six codons each. Valine, Proline, Threonine, Alanine, Arginine and Glycine were code by four codons each whereas the least variable region is caused by amino acids having two codons each; Phenylalanine, Isoleucine, Methionine, Tyrosine, Histidine, Glutamine, Asparagine, Lysine, Glutamic acid, Aspartic acid, Cysteine and Tryptophan.

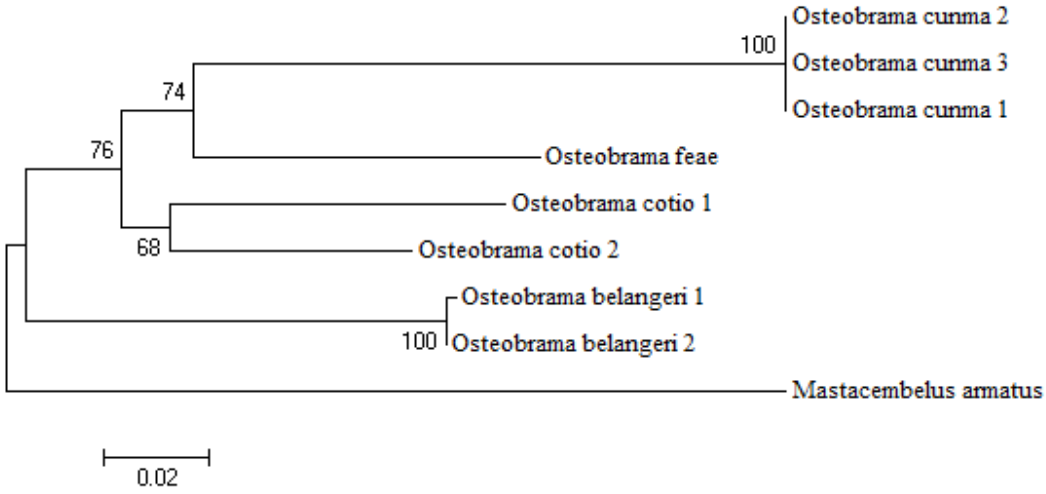

Fig 1: Neighbor-Joining (NJ) bootstrap consensus tree of *COX I* gene of *Osteobrama* species.

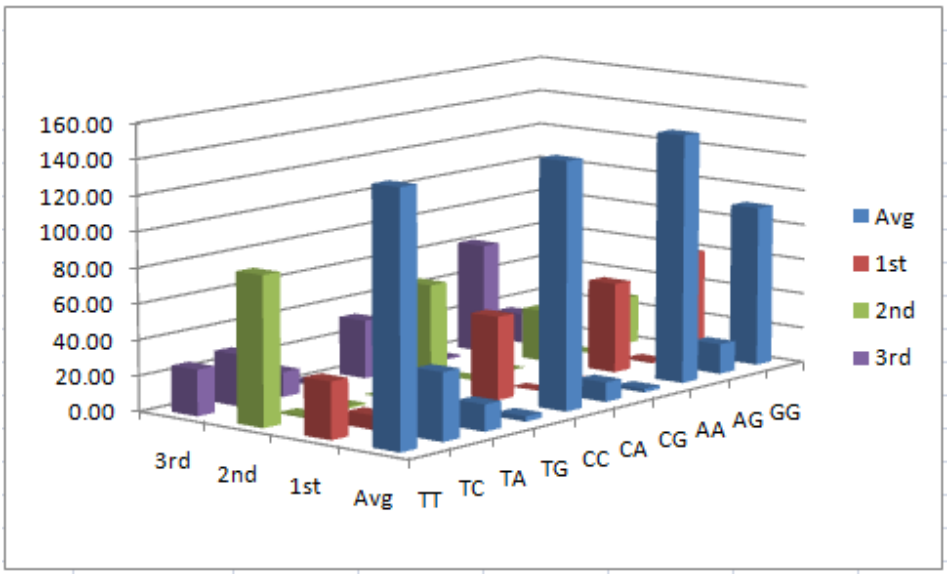

Fig 2: Nucleotide pair frequency of all the taxa considered.

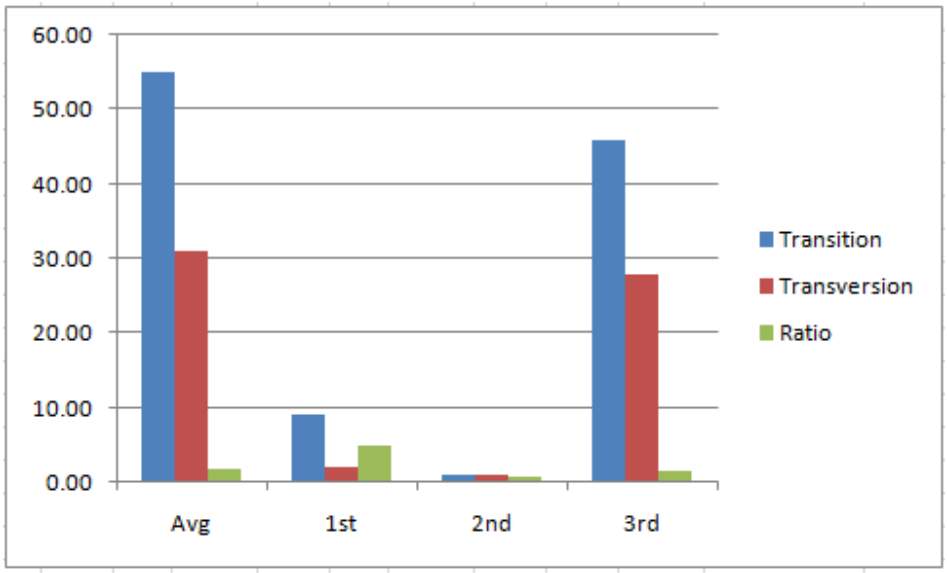

Fig 3: COX I substitution plots. Number of transition and transversion at codon position.

The nucleotide pair frequency within all the taxa were counted site by site for two sequences. The frequencies of these quantities for each sequence as well as an overall average varies to a great extent. GG concentration at first codon constitute 62% whereas the concentration of TT, CA, CG pair is nil. At second codon position, TT makes up 87% while TA, TG, CA, CG, AG pair is absent. Also AA represents the highest concentration of 61% whereas CG represent 2% of the nucleotide pair frequency in third codon position (Fig 2).

CONCLUSION

An analysis of nucleotide variation at each of the 623 bases across the *Osteobrama* species of North East India shows that each third codon position base is highly variable. Almost all the amino acids in the vertebrate genetic mitochondrial code is represented by at least two codons and at the highest of six codons for some amino acids and every amino acids allows some variation in the third base position. Those amino acids encoded by more codons show more flexibility and more variation at the third codon position. Nucleotide substitution at the second codon position is lowest and hence more conserved than the first and third codon position (Fig 3). This highly variable nature at the third codon position reflects the degeneracy nature of the genetic codon and hence successfully delineates the *Osteobrama* species.

REFERENCES

[1] P.D.N. Herbet, A. Cywinska, S.L. Ball, J.R. de Waard, *Proc. R. Soc. Lond.*, **2003**, 270, 313-322.
[2] M. Vences, M. Thomas, R.M. Bonett, D.R. Vieites, *Phil. Trans. R. Soc. Lond.*, **2005**, 360, 1859-1868.
[3] M.A. Smith, B.L. Fisher, P.D.N. Hebert, *Phil. Trans. R. Soc. Lond.*, **2005**, 360, 1825-1834.
[4] P.D.N Hebert, M.Y. Stoeckle, T.S. Zemlak, C.M. Francis, *PLOS Biol.*, **2004**, 2, 1657–1663.
[5] I.D. Hogg, P.D.N. Hebert, *Can. J. Zool.*, **2004**, 82, 749–754.

[6] R.D. Ward, S.Z. Zemlak, B.H. Innes, P.R. Last, P.D.N Hebert, *Phil. Trans. R. Soc. Lond.*, **2005**, 360, 1847–1857.
[7] M.A. Smith, N.E. Woodley, D.H. Janzen, W. Hallwachs, P.D.N. Hebert, *Proc. Natl. Acad. Sci. USA.*, **2006**, 103, 3657–3662.
[8] S.L. Ball, K.F. Armstrong, *Can. J. For. Res.*, **2006**, 36, 337–350.
[9] M. Hajibabaei, D.H. Janzen, J.M. Burns, W. Hallwachs, P.D.N. Hebert, *Proc. Natl. Acad. Sci. USA*, **2006**, 103, 968–971.
[10] R.D.H. Barrett, P.D.N. Hebert, *Can. J. Zool.*, **2005**, 83, 481–491.
[11] M. Kimura, The Neutral Theory of Molecular Evolution, Cambridge University Press, Cambridge, UK, **1983**, 384.
[12] M. Nei, Molecular Evolutionary Genetics, Columbia University Press, New York, USA, **1987**, 512
[13] V. Waikhom, S, M, *Zoos' Print J.*, **2007**, 22 (11), 2881–2884.
[14] V. Waikhom, W.S. Lakra, U.K. Sarkar, Fishes of North East India. National Bureau of Fish Genetic Resources, Lucknow, **2007**, 264.
[15] K. Tamura, J. Dudley, M. Nei, S. Kumar, *Mol. Biol. Evol*., **2004**, 24(8), 1596-1599.
[16] M. Kimura, *J. Mol. Evol*., **1980**, 6, 111–120.
[17] N. Saitou, M. Nei, *Mol. Biol. Evol.*, **1987**, 4, 406-425.
[18] A. Rzhetsky, M. Nei, *Mol. Biol. Evol.*, **1982**, 9, 945–967.
[19] R.V. Eck, M.O. Dayhoff, Atlas of Protein sequence and structure. National Biomedical Research Foundation, Silver Springs, Maryland, **1966**.
[20] J. Felsenstein, *Evol*. **1985**, 39, 783–791