



Department of Biotechnology  
Hargobind Khurana School of Life Sciences  
ASSAM UNIVERSITY, SILCHAR  
(A central university constituted under act XIII of 1989)  
Silchar- 788011, Assam, India

## *DECLARATION*

*I, Mohua Chakraborty, bearing Registration Number, Ph.D. 1711/2011 dated 15.09.11, and Assam University Registration Number 16-100010544 of 2010-11, do hereby declare that the subject matter of the thesis entitled "DNA Barcode based calibration of Indian freshwater fishes for analyzing species diversity and shared nucleotide traits" is the record of work done by me and that the contents of this thesis did not form the basis for award of any degree to me or to anybody else to the best of my knowledge. The thesis has not been submitted in any other University or Institute.*

*Place: Silchar*

*Mohua Chakraborty*

*Date:*

## ***Acknowledgement***

*I am using this opportunity to express my gratitude to everyone who supported me throughout the course of this program.*

*Foremost, I would like to express my sincere gratitude to my supervisor Dr. Sankar Kumar Ghosh, Professor, Department of Biotechnology, for the continuous support of my Ph.D. study and research. I was constantly motivated by his patience, motivation, enthusiasm, and immense knowledge. I am thankful to him for encouraging my research and for allowing me to grow as a research scientist. His advice on both research as well as on my career have been invaluable. I could not have imagined having a better advisor and mentor for my Ph.D. study.*

*I thank Honourable Vice Chancellor, Dean School of Life Sciences, Head of the Department of Biotechnology and Board of Research Studies, Assam University for their support for various phases of the programme.*

*I gratefully acknowledge the funding sources that made my Ph.D. work possible. In this regard I extend my gratitude to DBT for providing me financial aid during the first two years of my research.*

*I am extremely thankful to Dr. Amalesh Choudhury for helping me in sample collection and identification. Working with him was a great learning experience.*

*All the faculty members and non-teaching staffs of the Department of Biotechnology, Assam University have been very kind to extend their help at various phases of this research. I would especially like to thank Dr. Yashmin Choudhury and Dr. Mohua Sengupta for their support and encouragement.*

*I thank my fellow labmates and seniors Boni Amin Laskar, M. Joyraj Bhattacharjee, Bishal Dhar, Ruhina Shirin Laskar, Fazlur Rehman, Shantanu Kundu, Ksh Miranda Devi, Monika Ahanthem, Shilpi Srivastava, Pradosh Mahadani, Jagadish Hansa, Rosy Mondal, Sharbadeb Kundu, Moonmi Chetia, Anil Seram, Joyobrata Nath, Shailendra Yadav, Susweta Mitra, Mridul Das, Dr. Bishwadeep Choudhury and my friends Saradia Kar, Aditi Bhowmik, Sarmistha Dey. It was their wonderful*

*companionship that helped me enjoy and survive through these four years of my life. The discussions in the lab, sharing of knowledge, cumulative effort in solving various technical and theoretical problems have enriched me in numerous ways.*

*I cannot but give full credit to my family members for having encouraged me and driving me forward to this point. I am thankful to my parents Haran Chakraborty and Uttara Chakraborty and my in-laws Bijon Kanti Choudhury and Nabanita Choudhury for their moral support and guidance throughout this period. I am grateful to my husband Budhaditya Choudhury for his unparalleled sacrifice and understanding. Without his constant support in all my pursuits, nonstop motivation and constructive criticism, it was impossible for me to complete my journey. I am extremely thankful to my brother Hillol Chakraborty for his much needed input in the entire coding related work of my study. Finally, I extend my gratitude to all other family members and friends especially to my brother in law Bappaditya Choudhury for their good wishes and encouragement.*

# Contents

Chapter 1. INTRODUCTION .....	1
1.1 Freshwater fishes: biodiversity and taxonomy .....	1
1.1.1 Ichthyology .....	1
1.1.2 Global geographical distribution of freshwater fishes.....	2
1.1.3 Taxonomy of fishes .....	3
1.1.4 Impediments in freshwater fish taxonomy.....	4
1.2 Indian freshwater fish diversity and problem in taxonomy .....	5
1.2.1 Fish biodiversity in India .....	5
1.2.2 Taxonomic impediments in Indian freshwater fishes.....	7
1.3 Molecular methods to resolve taxonomic impediments.....	8
1.3.1 Molecular vs. morphological methods .....	8
1.3.2 Mitochondrial DNA in species identification .....	8
1.3.3 The advent of DNA barcoding.....	9
1.3.4 Fish DNA barcoding.....	11
1.4 DNA barcode for higher taxon assignment .....	13
1.5 Factors effecting efficacy of DNA barcodes .....	15
1.5.1 Reference barcode library .....	15
1.5.2 Sample condition .....	16
1.5.3 Algorithms for DNA barcode based identification.....	18
1.6 Prospect of the study .....	21
1.7 Objective of the study.....	22
Chapter 2. REVIEW OF LITERATURE.....	23
2.1 Cataloging and classification of global fish species .....	23
2.2 Cataloging and classification of Indian freshwater fishes.....	25
2.3 Molecular approaches and use of DNA barcoding in fish systematics.....	27
2.3.1 Early molecular studies.....	27
2.3.2 Mitochondrial DNA in taxonomy studies.....	28

2.3.3	Introduction of DNA Barcoding.....	29
2.3.4	Launch of iBOL and next phase of DNA Barcoding .....	31
2.4	Applications of fish DNA barcoding .....	35
2.5	Evolution of DNA barcoding as a technique.....	40
2.5.1	Distance and monophyly based methods.....	40
2.5.2	Advent of Character based methods.....	41
2.5.3	Comparison of distance based and character based methods .....	42
2.5.4	New methods of barcoding .....	43
Chapter 3.	MATERIALS AND METHODOLOGY.....	45
3.1	Materials.....	45
3.1.1	DNA barcode sequence .....	45
3.1.1.1	Retrieved DNA barcode sequences .....	45
3.1.1.2	Generated barcode sequences from eastern and Northeastern India. .	46
3.1.2	Chemicals.....	47
3.1.2.1	Reagents.....	47
3.1.2.2	Buffers .....	48
3.1.2.3	Primers .....	49
3.2	Methodology .....	50
3.2.1	Generating <i>COI</i> DNA barcodes from fish samples.....	50
3.2.1.1	Isolation from tissue .....	50
3.2.1.2	Purity check and yield of the extracted DNA .....	51
3.2.1.3	Spectrophotometric determination .....	51
3.2.1.4	Agarose gel electrophoresis for genomic DNA quantification and quality analysis.....	52
3.2.1.5	PCR amplification of barcode segment of <i>COI</i> gene .....	53
3.2.1.6	Gel-electrophoresis and purification of PCR product .....	53
3.2.1.7	Sequencing of PCR amplicons.....	55
3.2.1.8	Sequence submission.....	55
3.2.2	<i>In silico</i> sequence analysis .....	55
3.2.2.1	Sequence formats used for sequence analysis.....	55

3.2.2.2	Data analysis .....	56
3.2.2.3	Rseq calculation for sequence information analysis .....	58
3.2.2.4	Selecting high-informative segment within the <i>COI</i> barcode.....	58
3.2.2.5	<i>In silico</i> primer designing of minibarcode segment .....	59
3.2.2.6	Neighbor Joining tree construction .....	59
3.2.2.7	Developing consensus barcode motifs using MOTIF-BUILD .....	59
3.2.2.8	Specificity check of the barcode motifs using MEGA .....	60
3.2.2.9	Specificity check using MOTIF-MATCH.....	60
Chapter 4.	RESULTS .....	63
Chapter 4.1	Assessment of freshwater fishes of India.....	64
4.1.1	Assessment of the recorded status .....	64
4.1.2	Assessment of the barcoded status .....	66
4.1.3	Species authentication using DNA barcoding.....	70
4.1.3.1	Detecting erroneous identification of fillet. ....	70
4.1.3.2	Detecting erroneous identification of juvenile market specimen.....	75
Chapter 4.2	Compositional analysis of <i>COI</i> DNA barcodes. ....	77
4.2.1	Nucleotide composition analysis.....	77
4.2.2	Nucleotide pair frequencies.....	84
4.2.3	Codon usage .....	85
4.2.4	Sequence information in the <i>COI</i> barcode using $R_{seq}$ value.....	87
Chapter 4.3	DNA barcode based taxonomic rank assignment using distance method.....	92
4.3.1	Species level identification .....	92
4.3.2	Special case of the unusual genetic diversity in <i>Clarias batrachus</i> . ....	103
4.3.3	Genus level identification .....	106
4.3.4	Family level identification .....	108
4.3.5	Order level identification .....	109
Chapter 4.4	Development of character profiles for different hierarchical levels of a taxon.....	110
4.4.1	Character profiles for species.....	110
4.4.2	Character profiles for genus .....	115

4.4.3	Character profiles for family .....	118
Chapter 4.5	Development of species specific barcode motif.....	121
4.5.1	Retrieval of highly informative region .....	121
4.5.2	Primer design and validation.....	121
4.5.3	Validation of the proposed segment for use in species delimitation .....	124
4.5.4	Design of barcode-motifs.....	124
4.5.5	Motif specificity validation .....	125
4.5.6	Validation of the barcode motifs with global Cypriniformes data .....	128
Chapter 5.	DISCUSSION .....	131
5.1	Status of DNA barcoding of freshwater fishes in India.....	131
5.2	Sequence composition of <i>COI</i> barcode and taxonomic rank assignment.....	132
5.3	Species level identification using distance based DNA barcode method.....	135
5.4	Character based assessment and development of Barcode Motif .....	140
5.5	Conclusion.....	143
Chapter 6.	SUMMARY .....	144
BIBLIOGRAPHY.....		146
APPENDICES .....		166
PUBLICATIONS .....		190

## List of Tables

No	Table Title	Page No
4.1	Summary of occurrence status of species belonging to different orders of Indian freshwater fishes.	65
4.2	Species recorded versus Species barcoded in different orders of Indian freshwater fishes.	69
4.3	Similarity search table of specimens barcoded in the study using BLAST search and BOLD identification engine.	73
4.4	Details of specimen sequenced as submitted in GenBank.	74
4.5	Nucleotide compositional analysis of 1307 species belonging to 10 orders of Indian freshwater fishes.	78
4.6	Nucleotide pair frequencies of 1383 sequences of 10 orders of Indian freshwater fishes across 510bp region of <i>COI</i> gene.	84
4.7	Codon Usage for 1383 sequences of Indian freshwater fishes across 510bp region of <i>COI</i> gene.	86
4.8	Frequency distribution of $R_{seq}$ value of variable nucleotides and amino acids in <i>COI</i> gene of different taxonomic ranks of Indian freshwater fishes.	91
4.9	Summary of genetic divergences (K2P model) for each taxonomic level of comparison.	92
4.10	Same named species with divergence above minimum congeneric value and formed two or more distinct sub-cluster.	95
4.11	Mean divergence within and between clusters of <i>Barilius</i> genus.	100
4.12	Different named species that clustered cohesively and showed divergence below maximum conspecific divergence.	102
4.13	Genetic divergence of different species of <i>Clarias</i> genus across various geographical location calculated using K2P model.	105
4.14	Summary of genera forming distinct cluster in NJ tree.	107
4.15	Summary of families forming distinct cluster in NJ tree.	108
4.16	Summary of orders forming distinct cluster in NJ tree.	109
4.17	Character profiles of <i>COI</i> barcode for species of Cyprinidae family at 30 nucleotide positions	113
4.18	Character Profiles of <i>COI</i> barcode for genera of Cyprinidae family at 20 nucleotide positions	117
4.19	Character profiles of <i>COI</i> barcode for families of Cypriniformes order at 22 nucleotide positions.	120
4.20	Motif verification of Cypriniformes species developed from Indian sequences against representative barcodes from other parts of the world present in BOLD.	129
5.1	Correlation of Nucleotide and Amino acid Rseq values between different orders of Indian freshwater fishes.	133
5.2	Number of indistinguishable species and the number of species with UCS.	137



## List of Figures

No.	Figure Title	Page No
1.1	Map showing drainage and river system of India.	6
3.1	Flowchart describing the categorization of the species into three groups based on clustering pattern in NJ tree.	57
3.2	Flow chart describing the algorithm of MOTIF-MATCH Program.	61
3.3	Flowchart describing the Match-String Module of the MOTIF-MATCH Program.	62
4.1	Diagram showing a) species recorded b) species barcoded c) species recorded vs species barcoded in 20 recorded orders of Indian freshwater fishes.	68
4.2	Representative image of DNA sequencing process showing: a) Gel image of PCR amplicons of COI barcode region of fish samples. b) DNA sequencing Chromatogram data.	71
4.3	Representative image of collected specimen of <i>Labeo bata</i> .	43
4.4	Nucleotide compositions in 10 orders of Indian freshwater fishes.	79
4.5	AT-GC Bias in 10 orders of Indian freshwater fishes.	79
4.6	Correlation of average GC content with GC content at each codon position, in 10 orders of Indian freshwater fishes.	80
4.7	Correlation between GC content at each codon position and average GC content in 1307 species of the orders Cypriniformes, Siluriformes and Perciformes.	80
4.8	Plot of AT-GC SKEW of 1307 species of Indian freshwater fishes a) Overall b) 1st codon position c) 2nd codon position d) 3rd codon position.	82
4.9	Correlation between GC skew at each codon position and average GC skew in species of different orders of Indian freshwater fishes.	83
4.10	Correlation between AT skew at each codon position and average AT skew in species of different orders of Indian freshwater fishes.	83
4.11	Pattern of distribution of sequence conservation across various taxonomic ranks of Indian freshwater fishes.	88

4.12	Distribution on conspecific and congeneric K2P mean divergence among 175 species of Indian freshwater fishes (arranged in ascending order).	94
4.13	(i) Sections of Neighbor-Joining (from Appendix 3) tree showing the problematic groups of (a) <i>Clarias batrachus</i> (b) Heteropneustes genus.	97
	(ii) Sections of Neighbor-Joining tree (from Appendix 3) showing the problematic groups of (c) Labeo genus (d) Channa genus (e) Macrognathus genus.	98
4.14	Neighbor joining tree showing clustering of different species of Barilius genus.	101
4.15	Neighbor joining tree showing genetic divergence of <i>Clarias batrachus</i> from different geographic location.	105
4.16	Pairwise interspecies difference in character attributes between 137 species.	111
4.17	Average intraspecies differences in character attributes within 116 species.	111
4.18	Pairwise intergenus difference in character attributes between 64 genera.	116
4.19	Average congeneric differences in character attributes within 64 genera.	116
4.20	Distribution of number of diagnostic characters within families.	119
4.21	Distribution patterns of transitions and transversion across full-length barcode (654bp) of 1307 sequences.	122
4.22	Comparison of interspecies distance between full length <i>COI</i> barcode (513bp) and minibarcode (154bp).	126
4.23	Comparison of intraspecies distance between full length <i>COI</i> barcode (513bp) and minibarcode (154bp).	127
5.1	Divergence summary (1:1, computed with K2P model) of 90 <i>COI</i> barcode sequences of the genus Barilius and Opsarius.	139

## List of Appendices

<b>NO</b>	<b>APPENDIX</b>	<b>PAGE No.</b>
Appendix 1	List of recorded freshwater fish species of India along with representative number of barcoded sequence present in the database.	166
Appendix 2	List of Indian freshwater fish <i>COI</i> barcode sequences used in the study.	175
Appendix 3	Neighbor joining tree using full length <i>COI</i> barcode of 175 species of Indian freshwater fishes.	180
Appendix 4	Neighbor joining tree of 160 species of Indian freshwater fishes using 154bp length <i>COI</i> barcode fragment.	183
Appendix 5	154bp length <i>COI</i> Barcode motifs designed in this study for 109 species.	186

## Abbreviations

µl	microliter
°C	Degree Celsius
A	Adenine
BOLD	Barcode of Life Data System
bp	base pair
C	Cytosine
CBOL	Consortium of Barcode of Life
COI	Cytochrome c oxidase subunit I
DNA	Deoxy Ribonucleic acid
EMBL	European Molecular Biology Laboratory
G	Guanine
gm	gram
IUPAC	International Union of Pure and Applied Chemistry
IUB	International Union of Biochemistry
K2P	Kimura 2 Parameter
lb	pound
MEGA	Molecular Evolutionary Genomic Analysis
mg	milligram
ml	millilitre
mM	millimolar
mtDNA	Mitochondrial DNA
NCBI	National Centre for Biotechnological Information
NJ	Neighbor Joining
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
pmole	picomole
rpm	Revolution per minute
RSCU	Relative Synonymous Codon Usage
T	Thymine