CHAPTER-3:

MATERIALS AND METHODS

3.1 Materials

3.1.1 Selection of sampling site

The state of Manipur harbours a wide variety of endemic ornamental fishes constituting 139 species (Mahapatra *et al*., 2007) most of which remain unexplored especially in molecular level. Therefore this region is selected as the study site. The ornamental fishes especially the endemic species were collected from different natural water bodies during different seasons of the year from random geographical locations in Manipur.

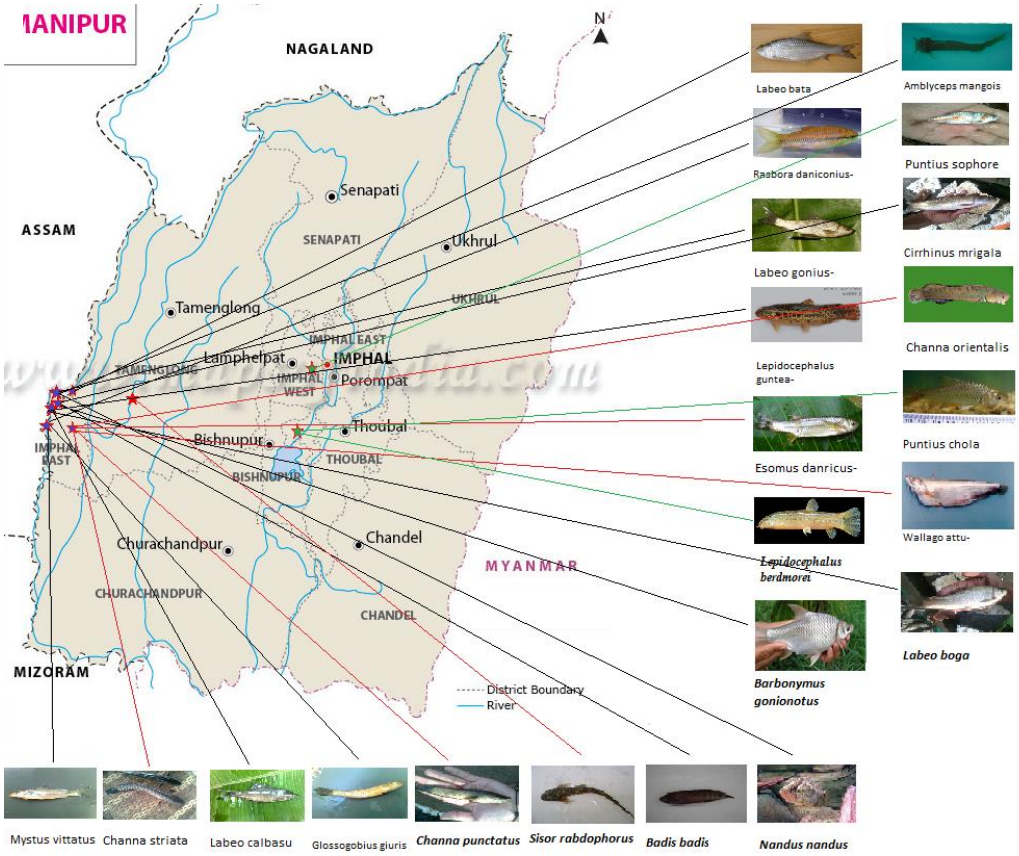

Figure 3. Map of Manipur, India showing the different locations of sample collection

3.1.2 Sampling, documentation of experimental taxa and accessing the database

Sampling operation was conducted for the first two years of the research program sequentially in winter to summer. A comprehensive data for all studied species were accumulated from www.fishbase.org and http://www.iucnredlist.org/ as a preliminary information for the present research work. Fresh specimens were spot examined for specific morphological characters that define the ornamental fishes and sampled from different wild habitats as and when caught by the professional fishers. Each of the catch was investigated to ensure correct sampling and labelling. The tissue samples from each of the specimens were collected aseptically and preserved in 90% ethanol. Major taxonomic keys of each of the fish specimens were noted and measurements were taken using a digital caliper. Species level identification of the specimens was confirmed by comparing with the described characters and the taxonomic keys available in the leading taxonomic guides of the fishes in India. However, the nomenclature of species follows the Catalogue of Fishes (Eschmeyer, 2012).

Possible ornamental fish samples (Blood, tissue) were collected from wild habitats as well as from different markets of Manipur, India after communicating with local fishermen. A small amount of muscle tissue (in case of market samples), blood from near caudal fin (in case of whole body specimens), was collected aseptically from each sample and stored in 500 µL of TES buffer (50 mM TrisHCl, 25 mM EDTA and 150 mM NaCl) or Lysis buffer (10 mM Tris-HCl, 10 mM EDTA and 50 mM NaCl) (Rohilla and Tiwari, 2008; Ghosh *et al.*, 2012).

3.1.3 Chemicals

The chemicals used in this study mainly includes those required for DNA isolation, PCR amplification, DNA sequencing and other associated techniques such as gel electrophoresis and gel purification. The chemicals were mostly of molecular biology grade and include,

(i) Tris (hydroxymethyl) aminomethane also called as Tris Base

(ii) Ethylene Diamine Tetrachloro Acetic acid (EDTA)

(iii) Sodium Chloride (NaCl)

(iv) Sodium Dodecyl Sulfate (SDS)

(v) Ethidium Bromide (EtBr)

(vi) Proteinase-K (Promega, USA)

(vii) Tris saturated phenol

(viii) Chloroform

(ix) Glacial Acetic acid

(x) Iso-amyl alcohol

(xi) Glycerol (SRL, India)

(xii) Absolute Ethanol (Bengal Chemical and Pharmaceuticals Works Ltd. Kolkata)

(xiii) *Taq* DNA polymerase

(xiv) dNTPs Master mix

(xv) Reagents kit

(xvi) Magnesium chloride ($MgCl_2$)

(xvii) PCR Buffer (Applied Biosystems USA).

The buffers and reagents were mostly molecular biology grade and includes,

(i) Tris (hydroxymethyl) aminomethane-HCl or Tris-Hcl-10 ml (1M, pH-8.00)

Tris base	1.211 gm
Nuclease free water	10 ml

Tris base 1.211 gm was added first to 7 ml of nuclease free water; initial pH is noted and adjusted to pH-8.00 by adding concentrated HCl. Once pH becomes stable at pH-8.00 remaining amount of nuclease free water was added to make the final volume 10 ml and then autoclaved at 15 lb for 10 min.

(ii) Ethylene Diamine Tetra Acetic acid or EDTA-10 ml (0.5M, pH-8.00)

EDTA	1.862 gm
Nuclease free water	10 ml

EDTA 1.862 gm was added first in 7 ml of nuclease free water; initial pH is noted and adjusted to pH-8.00 by adding NaOH. Once pH becomes stable at pH-8.00 remaining amount of nuclease free water was added to make the final volume 10 ml and then autoclaved at 15 lb for 10 min and kept at 22^0C.

(iii) Sodium Chloride or NaCl-10 ml (5M)

NaCl	2.422 gm
Nuclease free water	10 ml

NaCl 2.422 gm was added in 10 ml of nuclease free water and dissolved properly with an autoclaved stirrer and then autoclaved at 15 lb for 10 min and kept at 22^0C.

(iv) Sodium Dodecyl Sulfate or SDS- 10 ml (20%)

SDS	2 gm
Nuclease free water	9 ml

1 ml SDS powder is added with 9 ml of nuclease free water at room temperature and kept at 37^0C.

(v) Ethidium Bromide or EtBr-(10 mg/ml)

Ethidium bromide	10 mg
Nuclease free water	1ml

Mixed properly and wrapped the container in Aluminium foil and store at room temperature.

(vi) Proteinase -K- 200 µg/ml in nuclease free water.

(vii) Tris saturated phenol (pH 8.00)

Solid phenol was liquefied at 68°C in water bath and equal volume of 1 M TrisHCl (pH 8.00) was added. Discarded the upper aqueous phase and repeated the extraction procedure until the pH of the aqueous phase became 8.00. After final extraction an equal volume of TE was added to it and stored in a bottle wrapped with aluminium foil at 4°C.

(viii) Chloroform-Isoamyl alcohol (24:1)

24 parts of chloroform was added with 1 part of isoamyl alcohol and mixed properly and stored at 4°C.

(ix) Alcohol grades- 100%, 70%

(x) Glacial acetic acid

(xi) TES buffer (10 ml)

Tris-HCl	50 mM (added 500 µl from stock 1 M Tris-HCl)
EDTA	25 mM (added 500 µl from stock 0.5 M EDTA)
NaCl	150 mM (added 300 µl from stock 5 M NaCl)
Nuclease free water	8.7 ml

(xii) Lysis buffer (10 ml)

Tris-HCl	10 mM (added 100 µl from stock 1 M Tris-HCl)
EDTA	10 mM (added 200 µl from stock 0.5 M EDTA)
NaCl	50 mM (added 100 µl from stock 5 M NaCl)
SDS	20% (100 µl)
Nuclease free water	8.6 ml.

(xiii) TAE buffer – 10 ml (50X)

Tris base	9.68 g.
Glacial acetic acid	2.284 ml.
0.5 M EDTA (pH 8.0)	4 ml.
Distilled water up to	500 ml

Mixed properly and stored at room temperature**.**

(xiv) TE buffer

Tris-HCl	10 mM
EDTA	1 mM

(xv) Loading dye (6X)

Bromophenol Blue	25 mg
Glycerol	9 ml
Nuclease free water	7 ml

3.1.4 Primers

One set of forward and reverse primer was used to amplify the *COI* gene of the ornamental fishes' mtDNA. The primers selected for the study were published Universal fish *COI* primers (Ward *et al*., 2005).

The primer sequences were:

FishF1- 5'TCAACCAACCACAAAGACATTGGCAC3'

FishR1- 5'TAGACTTCTGGGTGGCCAAAGAATCA3'

3.1.5 Additional barcode data acquired from databases

In addition to the developed sequences, 46 DNA barcode sequences representing same and related taxa of the ornamental fish species under study were acquired from GenBank as replicate data for evaluating the taxonomic status of target species.

Table 1. Details of GenBank DNA barcode sequences retrieved in this study for analysis.

Sl no.	Species name	Accession no.	Geographical position
1	*Labeo bata*	HMI147890.1	Not Available
2	*Labeo bata*	HMI147889.1	Not Available
3	*Labeo bata*	HMI147888.1	Not Available
4	*Labeo bata*	HMI147886.1	Not Available
5	*Labeo bata*	HMI147887.1	Not Available
6	*Barbonymus gonionotus*	EU924634.1	Not Available
7	*Barbonymus gonionotus*	EU924633.1	Not Available
8	*Barbonymus gonionotus*	EU924632.1	Not Available
9	*Barbonymus gonionotus*	EU924635.1	Not Available
10	*Barbonymus gonionotus*	EU924631.1	Not Available
11	*Mystus vittatus*	JN628888.1	24.47 N 93.10 E
12	*Mystus vittatus*	DQ508093.1	Not Available
13	*Puntius sophore*	FJ459407.1	26.52 N 80.54 E

14	*Puntius sophore*	FJ459405.1	26.52 N 80.54 E
15	*Rasbora daniconius*	JN673954.1	27.30 N 94.58 E
16	*Rasbora daniconius*	EF452872.1	Not Available
17	*Labeo gonius*	EU417800.1	Not Available
18	*Labeo gonius*	EU417801.1	Not Available
19	*Labeo gonius*	EU417802.1	Not Available
20	*Labeo gonius*	HQ645092.2	Not Available
21	*Labeo gonius*	HQ645094.2	Not Available
22	*Wallago attu*	FJ170771.1	26.51 N 80.55 E
23	*Wallago attu*	JN628895.1	24.51 N 92.58 E
24	*Labeo calbasu*	GU195094.1	Not Available
25	*Labeo calbasu*	GU195093.1	Not Available
26	*Labeo calbasu*	GU195089.1	Not Available
27	*Labeo calbasu*	GU195092.1	Not Available
28	*Labeo calbasu*	GU195095.1	Not Available
29	*Labeo calbasu*	GU195091.1	Not Available
30	*Labeo calbasu*	GU195096.1	Not Available
31	*Labeo calbasu*	GU195090.1	Not Available
32	*Channa punctatus*	EU417795.1	Not Available
33	*Channa punctatus*	EU417796.1	Not Available
34	*Lepidocephalus guntea*	FJ459503.1	26.11 N 91.44 E
35	*Lepidocephalus guntea*	FJ459507.1	26.11 N 91.44 E
36	*Glossogobius giuris*	HQ945956.1	30.058 S 30.883 E
37	*Glossogobius giuris*	EF607394.1	20.45 N 109.89 E
38	*Glossogobius giuris*	EF609360.1	Not Available
39	*Esomus danricus*	HM224168.1	Not Available
40	*Sisor rabdophorus*	JN628884.1	24.51 N 92.58 E
41	*Sisor rabdophorus*	JN628916.1	24.50 N 92.45 E
42	*Sisor rabdophorus*	DQ508090.1	Not Available
43	*Amblyceps mangois*	DQ508067.1	Not Available
44	*Cirrhinus mrigala*	GU195083.1	Not Available
45	*Cirrhinus mrigala*	GU195085.1	Not Available
46	*Channa orientalis*	FJ459481.1	26.11 N 91.44 E

3.2 Methods

3.2.1 Isolation and quantification of DNA

Isolation of genomic DNA is started with cell lysis or cell disruption which is the phenomenon of rupturing the cell and the nuclear membrane to bring out the inner content of the cell which is considered to be first step towards isolation of genomic DNA. The other major contaminant, protein, is removed by shaking the solution gently with water-saturated phenol or with a phenol/chloroform mixture, either of which will denature proteins but not nucleic acids. Phenol-chloroform extraction is a liquid-liquid extraction technique in Biochemistry. Centrifugation of the emulsion formed by the mixing produces a lower organic phase, separated from the upper aqueous phase by an inter phase of denatured proteins. The aqueous solution is recovered and de-proteinase repeatedly, until no more material is seen at the interphase. Once the nucleic acid complex has been purified, precipitation can be accomplished and stored (Sambrook and Russell, 2001; Ghosh *et al*., 2012).

3.2.1.1 Isolation of DNA from different samples

- **Method-I (Tissue sample)**

(i) The tissue sample was removed from alcohol and chopped with a sterilized blade as small as possible and then kept dry in $(-80)^0$C for 30 minutes.

(ii) Added the tissue to a pre-cooled (dry ice) mortar, Homogenized gently in 2 volume (w/v) cold TES buffer, homogenizer was kept in ice previously. Adjusted the volume (500µl) with TES buffer.

(iii) To the above sample, added 50 µl of 10% SDS followed by 3-5 µl of 20 mg/ml of Proteinase K, incubated at 56°C for 1-18 hours until the tissue is totally dissolved.

(iv) Added equal volume of phenol: chloroform: isoamylalcohol (25:24:1) and mixed thoroughly for few minutes.

(v) Centrifuged the samples for 10 minutes with 12000 rpm.

(vi) Transferred upper phase to new 1.5 ml tube and added equal volume of chloroform: isoamylalcohol and centrifuged with 12000 rpm for 10 minutes.

(vii) Transferred the upper aqueous layer in a fresh sterilized microcentrifuge tube and added double the volume of chilled absolute ethanol.

(viii) The above sample was kept at $(-20)^0$ C for overnight for precipitation.

(ix) The above sample was centrifuged at 10,000 rpm for 10 minutes.

(x) Decanted the supernatant and retained the pellet.

(xi) To the pellet, added 500 µl of 70% ethanol and again centrifuged at 7000 rpm for 10 minutes and decanted the supernatant.

(xii) The pellet was kept for air dry under laminar air flow.

(xiii) The pellet was re-suspended in Nuclease free water stored either in $(-20)^0$ C or $(-86)^0$ C for immediate use or long preservation.

- **Method-II (Blood sample)**

(i) Blood samples typically were obtained as 1 ml of whole blood stored in EDTA vacutainer tubes frozen at 4^0 C.

(ii) The blood sample was pipetted in a sterilized microcentrifuge tube containing 500 µl of Lysis buffer.

(iii) To the above sample, added 50 µl of 10% SDS and mixed thoroughly for few minutes and added 1-2 µl of Proteinase-K.

(iv) Mixed the sample thoroughly by inverting the microcentrifuge tube for few minutes and it was then incubated at 56^0 C for 30 minutes.

(v) Added equal volume of phenol: chloroform: Isoamyl alcohol (25:24:1) mixed thoroughly for few minutes.

(vi) Centrifuged the sample for 10 minutes at 12,000 rpm cooling centrifuge.

(vii) Carefully removed the aqueous layer to a new sterilized microcentrifuge tube, added 500 ml of Chloroform: Isoamyl alcohol (24:1), mixed thoroughly by repeatedly inverting the microcentrifuge tube for few minutes.

(viii) Centrifuged at 12,000 rpm for 10 minutes in a cooling centrifuge.

(ix) Transferred the upper aqueous layer in a fresh sterilized microcentrifuge tube and added double the volume of chilled absolute ethanol.

(x) The above sample was kept at -20° C for overnight for precipitation.

(xi) The above sample was centrifuged at 10,000 rpm for 10 minutes.

(xii) Decanted the supernatant and retained the pellet.

(xiii) To the pellet, added 500 µl of 70% ethanol and again centrifuged at 7000 rpm for 10 minutes and decanted the supernatant.

(xiv) The pellet was kept for air dry under laminar air flow.

(xv) The pellet was re-suspended in nuclease free water stored either in $(-20)^0$ C or $(-86)^0$ C for immediate use or long preservation.

3.2.1.2 Spectrophotometric determination

Spectrophotometric analysis for checking purity and yield of the extracted DNA may be explained stepwise as follows:

1. The isolated DNA stock solution was taken at different dilution rate viz. 200, 50 and 10 times by adding nuclease free water.
2. Taken 50µl of nuclease free water in a cuvette and calibrated the spectrophotometer at 260nm as well as 280nm.
3. Added 2 µl of each DNA sample to 48 µl of nuclease free water in a cuvette and mixed well.
4. Optical densities (OD) were measured at 260 (OD_{260}) and 280 (OD_{280}) in UV spectrophotometer (Biophotometer, Eppendrof) against nuclease free water as blank. The yield and purity of DNA samples were estimated as follows:

Concentration of DNA stock solution (µg/ml) = OD_{260} X 100(dilution factor) X 50 µg/ml/1000

Purity of DNA stock solution = OD_{260}/OD_{280} (for pure DNA sample this ratio must be in the range of (1.75 – 1.80)

From the concentration of DNA stock solution, the total yield of DNA was calculated and recorded.

3.2.1.3 Agarose gel electrophoresis

The stepwise procedure followed was as follows:

1. Agarose powder was poured in electrophoresis buffer (IX TAE) to the desired concentration and mixed properly and then heated in a microwave oven until completely melted
2. Ethidium bromide (10 mg/ml) was added to the gel (final concentration 0.5 ug/ml) at this point to facilitate visualization of DNA after electrophoresis.

3. After cooling the solution to about 60^0C, it was poured into a casting tray containing a sample comb and allowed to solidify at room temperature

4. After the gel solidified, the comb was removed carefully without disturbing the wells formed

5. The gel, still in its plastic tray, was inserted horizontally into the electrophoresis chamber and covered with buffer

6. Samples containing DNA mixed with loading buffer were then pipetted into the sample wells, the lid and power leads are placed on the apparatus and current was applied and flow confirmed through observing bubbles coming off the electrodes.

7. The distance DNA had migrated in the gel was judged by visually monitoring migration of the tracking dyes, Bromophenol blue that migrate through agarose gels at roughly the same rate as double-stranded DNA fragments of 300 and 4000 bp.

8. When adequate migration occurred, DNA fragments were visualized by staining with Ethidium bromide

9. To visualize DNA, the gel was placed on an ultraviolet transilluminator and photograph was taken in Gel-DOC (BioRad).

The concentration of the DNA extracted was measured by calibrating the marker that was run parallel with the samples. The illumination intensity of the sample DNA in comparison to the marker DNA (whose value was used for calibration) gives rough idea regarding the concentration of the extracted DNA.

3.2.2 PCR amplification of *COI* barcode segment

As mentioned above published primers of Ward *et al*., 2005 was used to amplify the barcode segment of COI gene in a Veriti Mastercycler (Applied Biosystems Inc.,CA,USA).

3.2.2.1 PCR reaction settings

Each 50 µl PCR reaction mixers contain:

10X PCR Buffer	:	5µl
dNTPs Mastermix(10 mM)	:	5µl
Forward prim (20 pmole/µl	:	1µl

Reverse primer (20 pmole/µl)	:	1µl
High fidelity DNA polymerase (5 Unit/µl)	:	1µl
Genomic DNA (100-200 ng)	:	variable
Nuclease free water	:	upto 50 µl

3.2.2.2 PCR cycling condition

The PCR reaction was set with an initial denaturation temperature of 94^0C host start for 3 min and subsequently, 94^0C for 1 min for denaturation, 50 ^{0}C for 45 sec 72^0C for 45 sec for extension primer annealing for 30 cycles followed by 72^0C for 10 min for final extension using gradient thermal cycler (Applied Biosystem, Inc. USA). Aliquots for 10µl of DNA products from PCR amplification were loaded in 0.8-1.2 % agarose gel for electrophoresis in 1X TAE. Gel was stained with ethidium bromide and observed under UV transilluminator and documented with Gel-DOC (BioRad).

3.2.2.3 Purification of PCR products

The PCR-amplified products were analyzed in 1% low melting agarose gel containing ethidium bromide staining (10 mg/ml).The single uniform band was excised and purified using QIA quick^R Gel extraction kit (QIAGEN, USA), following manufacturer's instructions given below.

1. Excised the DNA fragment from the agarose gel with a clean, sharp scalpel.

2. Transferred the gel slice to a 1.5ml or 2.0ml tube. Added then 650µl Gel solubilizer.

3. Incubated the tube for 10 minutes at 50ºC in a water bath until the gel slice had completely dissolved (3-4 times during incubation).

4. Added 50µl Binding Optimizer to the sample. Mixed by pipetting up and down, or by vortexing.

5. Transferred 750µl of the sample to Spin Column A placed in a 2ml collection tube. Centrifuged at 10,000 x g (12,000rpm) for 1 minute. Discarded the filtrate and reused the collection tube by placing the Spin Column back in the collection Tube. Loaded residual solution and repeated the centrifugation step. Reused the collection tube.

6. Added 700µl Wash Buffer A and centrifuged at 10,000 x g (12,000rpm) for1 minute. Discarded the filtrate and reused the collection tube by placing the spin column back in the collection Tube.

7. Repeated step 6.

8. Centrifuged at maximum speed for 2 minutes to remove all traces of ethanol and the collection tube was discarded.

9. Placed spin column A into a 1.5ml Elution Tube and added 30-50µl Elution buffer directly to the spin column membrane. Incubated at room temperature for 1 minute, centrifuged at 10,000 x g (12,000rpm) for 1 minute to elute the DNA.

10. The DNA was as such purified and stored at -20ºC.

3.2.3 Sequencing of PCR amplicons

The purified amplicons of the COI were bidirectionally sequenced in an automated DNA sequencer (ABI 3500, Applied Biosystems Inc., CA, USA), through GCC Biotech India Pvt. Ltd. (Kolkata, India).

3.2.4 Sequence checking and formatting

The raw sequences generated after sequencing was annotated and brought in proper format in order to make the sequences applicable for phylogenetic analysis. The techniques and procedures followed for the annotation and analysis of the sequences were detailed below.

3.2.4.1 Interpretation of sequencing chromatogram

Automated DNA Sequencers generated a four-color chromatogram showing the results of the sequencing run, as well as a computer program's best guess at interpreting that data - a text file of sequence data. That computer program, however, does make mistakes and thus manually double-checked the interpretation of the primary data. Predictable errors occurred near the beginning and again at the end of sequencing run. Other errors crop up in the middle, invalidating individual base calls or entire swaths of data. The sequence chromatogram was viewed using software **SeqScanner Version 1.0** (Applied Biosystems Inc., CA, USA). The software was used in this study to display,

edit, and trim, print, generate and export reports for sequencing sample files from Applied Biosystems genetic analysis instruments. Detail of the software and downloading link is available at https://products.appliedbiosystems.com/ab/en/US/adirect/ab?cmd=catNavigate2&catID=600583&tab=DetailInfo.

For each sample two chromatograms that represent sequences of both the strands of DNA were obtained. Ends of the noisy sequences were trimmed. BLASTN program (Altschul *et al*., 1990) was used to compare the sequences from the two chromatograms, and the fragment of the two sequences showing 100% alignment with no gap or indels (insertion/deletions) was selected. In case of any discrepancy, both the sequences were reviewed and quality value of the sequence was considered to determine the most likely nucleotide using SeqScanner Version 1.0 (Applied Biosystems Inc., CA, USA). The selected fragments of the sequences for all the specimens were aligned using ClustalX software (Thompson *et al*., 2002) and subjected to BLASTN searches at the National Centre for Biotechnology Information to see whether the developed sequences align with database mitochondrial COI gene without having any indels. The sequences were translated using the online software ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and aligned through BLASTP to further check matching of the array of amino acid in the developed sequences in comparison to mitochondrial COI barcode sequences.

3.2.4.2 Different formats of sequences to be compatible with analysing software

The sequence formats used in this study were as follows:

FASTA is the simplest text-based format for representing either nucleic acid sequence or protein sequence. The format begins with single line sequence description proceed by ">".On the next line, the actual sequence is represented in the standard IUB/IUPAC amino acid or nucleic acid code.

CLUSTAL is the standard output format of popular alignment software and is also widely supported as input format in phylogenetic programs. The format is recognized by the word CLUSTAL at the beginning of the file. The sequence alignment output from CLUSTAL software is usually given the default extension "aln".

In **MEGA** format, the "#Mega" keyword indicates that the data is prepared for analysis using MEGA. It must be present on the very first line in the data file. On the second line, the word "Title" must be written, which can be followed by some description of data on the same line. Each taxon label must be written on a new line starting with #.

GenBank format consists of a mixture of compressed and uncompressed ASCII text files, containing sequence data and indices that cross reference author names, journal citations, gene names and keywords to individual GenBank records.

3.2.5 Sequence submission in GenBank and BOLD

All the analysed sequences were deposited in GenBank through the Banklt sequence submission tool (http://www.ncbi.nlm.nih.gov/WebSub/?tool=genbank) of GenBank and received valid accession numbers. The sequences were also submitted in BOLD following BOLD sequence submission protocol (BOLD management paper) and received valid process ID's.

3.2.6 Similarity search

The sequence similarity search for species identification was done in two public databases, viz., BOLD and GenBank. The highest percent pairwise identity for each sequence blasted at NCBI were compared with the percent similarity scores of the same sequence within the BOLD-IDS (BOLD Identification System). The query species that matched either with the same or different species in the databases has been termed as 'specific' or 'non-specific' respectively. The similarity range of 97%-100%, 92%-96% and ≤91% between the query and the database sequence have been expressed as significant, moderate and insignificant respectively.

3.2.7 Sequence alignment

ClustalX was used for multiple alignments of the sequences. It has taken a set of input sequences in fasta format and carried out the entire progressive alignment procedure automatically. The sequences were aligned in pairs in order to generate a distance matrix that can be used to make a simple initial tree of the sequence. ClustalX is freely available and can be downloaded from EMBL/EBI file server

(ftp://ftp.ebi.ac.uk/pub/software/) or from ICGEB in Strasbourg, France (ftp://ftp-igbmc.ustrasbg.fr/pub/ClustalW/ and ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/). In each case, ClustalX (X stands for windows) provided a graphical user interface with colourful display of alignments. Clustal X performed the progressive alignment and created an output guide tree file and an output alignment file in the default Clustal format. Clustal X indicated the degree of conservation at the bottom of the aligned sequences, which was used to evaluate a given alignment.

3.2.8 Model selection and sequence analysis

For the construction of Maximum likelihood tree, the first thing which is a must to develop a perfect tree is to choose the best model. The optimal model that best explain the evolution of the sequences were strained through goodness-of-fit test of each model measured through Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). Non-uniformity of evolutionary rates among sites may be modelled by using a discrete Gamma distribution (+*G*) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+*I*).This result is an evaluation of 24 models for nucleotide substitutions. The model with the lowest Bayesian Information Criterion (BIC) value are considered to describe the substitution pattern the best.

Maximum conspecific and minimum congeneric divergences have been determined considering the sequences showing cohesive NJ clustering within a species and remained distinct from other species. The number of times the minimum congeneric divergence differs from the maximum conspecific divergence is the lowest divergence between congeners and has been assumed to be the threshold level of species delineation and thereby considered as a barcoding gap in this study.

Kimura 2-parameter (K2P) (Kimura, 1980) congeneric and conspecific variation, Neighbor Joining (NJ), Maximum Likelihood (ML) tree, maximum Parsimony (MP) tree and Model test for ML were done using the computer program MEGA version 5.0 (Tamura *et al*., 2011) for calculating the genetic divergence, selecting the best model for sequence evolution and for phylogenetic assessment.

MEGA-5.0 has the advantage of calculating pairwise genetic divergence between sequences using various models and for phylogenetic interpretation using the distance based method like NJ and character state method like MP and ML. In this study the aligned file of the sequences (done by ClustalX) were inputted in MEGA and converted to MEGA format (.meg). The mega format of the sequences were then used for pairwise genetic divergence calculation, for model test, NJ and ML phylogenetic inference following the MEGA 5 usage instruction of Tamura *et al.*, (2011).