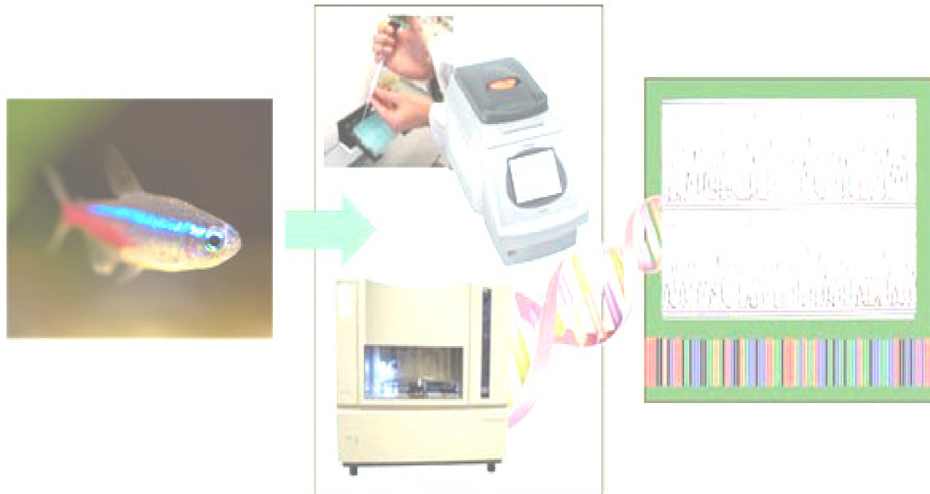

CHAPTER 3

MATERIALS AND METHODS

CHAPTER 3
MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 SAMPLING

The ornamental fishes were collected from different natural water bodies during different seasons of the year from random geographical locations of North East India. Fresh specimens were spot examined for specific morphological characters that define each of the fishes and sampled from different wild habitats as and when caught by the professional fishers and ornamental fish traders. Each of the catch was investigated to ensure correct sampling and labelling. Upon every spate of collection, the tissue samples from each of the specimens were collected aseptically and preserved in 90% ethanol.

The details of sampling that includes sample-ID, collection date and geographical location was given in the following result section. A pictorial diagram of different collection sites is shown in Figure 3.1.

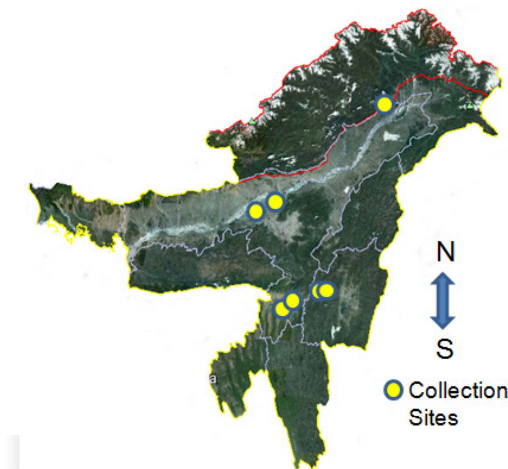

Figure 3.1 Sites of collection of ornamental fishes along different locations of North East India

3.1.2. MORPHOLOGY

For identification, the diagnostic morphology keys are recorded as per the standard literature (Dishma and Vishwanath 2012, Geetakumari and Vishwanath 2011, Hamilton 1822, HEOK HEE 2007, Kottelat 1990, Nebeshwar et al. 2009, Roberts and Ferraris 1998, Shangningam et al. 2013, Talwar and Jhingran 1992, Vishwanath and Kosygin 2000, Vishwanath and Laisram 2004). All the measurements were recorded using a digital slide calliper (0.01mm). Those non quantitative morphological characters, e.g.; Color; Blotch, Bands/Stripes, etc. were also recorded. The details of the taxonomic keys used for the identification purpose are summarized in Appendix 2.

3.1.3 CHEMICALS

The chemicals were mostly of molecular biology grade and this includes- Tris Base, Ethylenediamine Tetrachloro Acetic acid (EDTA), Sodium chloride (NaCl), Sodium Dodecyl Sulfate (Promega, USA), Ethidium Bromide (EtBr), Proteinase-K (Promega, USA), Tris saturated phenol, Chloroform, Glacial Acetic acid, Iso-amyl alcohol and Glycerol (SRL, India), Absolute Ethanol (Bengal Chemical and Pharmaceuticals Works Ltd. Kolkata), *pfu* & *Taq* DNA polymerase, dNTPs mix (Promega), DNA purification and cleanup reagents kit, Magnesium chloride ($MgCl_2$), DMSO, BSA, etc. (Fermentas, USA and Invitrogen, USA), DNA Sequencing reagents (Applied Biosystems, USA).

3.1.3.1 REAGENTS

I. Tris (hydroxymethyl) aminomethane-HCl or Tris-HCl- 10 ml (1M, pH-8.0)

Tris base	1.211 gm
Nuclease free water	upto10 ml

Tris base weighing 1.211 gm was added to 7 ml of nuclease free water; pH was noted and adjusted to pH-8.00 by adding concentrated HCl under continuous stirring and observed. 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 15 minutes.

II. Ethylenediamine-tetra acetic acid or EDTA-10 ml (0.5M, pH-8.00)

EDTA	1.862 gm
Nuclease free water	upto10 ml

1.862 gm of EDTA wad added in 7 ml of nuclease free water; initial pH was noted and adjusted to pH-8.00 by adding NaOH and continuously mixed thoroughly. 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 15 minutes and kept at 22^0C.

III. Sodium chloride or NaCl-10 ml (5M)

NaCl	2.422 gm
Nuclease free water	10 ml

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

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

SDS	2 gm
Nuclease free water	10 ml

2gm of SDS powder was added with 10 ml of nuclease free water at room temperature and kept at 37^0C.

V. Ethidium Bromide or EtBr- (10mg/ml)

Ethidium bromide	10mg
Nuclease free water	1ml

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

VI. Proteinase-K- 200 µg/ml in nuclease free water.

VII. Tris saturated phenol (pH 8.00)

VIII. Chloroform-Isoamyl alcohol (24:1)

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

IX. Alcohol grades- 100%, 90%, 70%

3.1.3.2 BUFFERS

I. TES buffer (10 ml)

Tris-HCl	50mM (added 500 µl from stock 1 M Tris-HCl)
EDTA	25mM (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

II. 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

III. TAE buffer – 10 ml (50X)

Tris-HCl	2.42 gm
Glacial acetic acid	570 µl
EDTA- 500mM	1000 µl
Nuclease free water	8.43 ml

IV. TE buffer

Tris-HCl	10 mM
EDTA	1 mM

IV. Loading dye (6X)

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

3.1.3.3 PRIMERS

Two pairs of primer were used in this study. The first pair was to amplify the full-length barcode segment of mitochondrial *CO1* gene sequence of ornamental fishes while the next pair is the degenerate primer to amplify the mini-barcode segment designed in this study. The first primer pair proposed by Ward et al. (2005) for fishes were used in many fish projects and therefore also used in this study and is given below.

FishF1-$5'$ TCAACCAACCACAAAGACATTGGCAC $3'$ and
FishR1-$5'$ TAGACTTCTGGGTGGCCAAAGAATCA $3'$

For amplification of the mini-barcode fragment, both upstream and downstream sequences (in reverse complement form) spanning the selected mini-barcode fragment was partitioned from the alignment considering all the fundamental primer properties, a degenerate primer pair was designed manually and its properties were checked with the online software OligoCalc (http://simgene.com/OligoCalc) and sequence manipulation suite. The details of the mini-barcode primers are summarized in the chapter 4.4 of the results section.

The designed mini-barcode primers used in the study are-

Fish Com F- $5'$ - GCNTTCCCNCGAATRAANAACAT- $3'$ and
Fish Com R- $5'$ - GATNGTNGTGATGAAGTTNAT - $3'$.
Cypri F- $5'$ - GGCRTTCCCNCGWATAAACAAC - $3'$ and
Cypri R- $5'$ - GGTGGTWGTAATGAARTTAAT - $3'$
Siluri F- $5'$ - GCATTCCCYCGAATRGAYAACA - $3'$ and
Siluri R- $5'$ - GATDGTTGTGATGAAGTTGAT - $3'$

3.2 METHODS

3.2.1 ISOLATION AND PURIFICATION OF DNA FROM ORNAMENTAL FISH SAMPLES

For isolation of genomic DNA both blood sample and tissue sample were preferred. Muscle tissue just above caudal fin or blood was collected aseptically in alcohol and EDTA vial respectively. Further steps of isolation was done

following phenol-chlorofrom-isoamyl method (Green and Sambrook 2012) while few steps differs in case of tissue and blood and may be explained separately

3.2.1.1 ISOLATION FROM TISSUE/FIN

For isolation of genomic DNA tissue samples and fins were preferred. Muscle tissue/Fins were collected aseptically in alcohol. Further steps of isolation were done following phenol-chlorofrom-isoamyl method.

Steps followed:

1. 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.
2. 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.
3. To the above sample, added 50 µl of 10% SDS followed by 5-10 µl of 20mg/ml of Proteinase-K, incubated at 56°C for 1-18 hours until the tissue is totally dissolve.
4. Added equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) and mixed thoroughly for few minutes.
5. Centrifuged the samples for 10 minutes with 12000 rpm.
6. Transferred upper phase to new 1.5 ml tube and added equal volume of chloroform: isoamyl alcohol and centrifuged with 12000 rpm for 10 minutes.
7. Transferred the upper aqueous layer in a fresh sterilized micro-centrifuge tube and added double the volume of chilled Absolute ethanol.
8. The above sample was kept at -20° C for overnight for precipitation.
9. The above sample was centrifuged at 10,000 rpm for 10 minutes.
10. Decanted the supernatant and retained the pellet
11. To the pellet, added 500 µl of 70% ethanol and again centrifuged at 7000 rpm for 10 minutes and decanted the supernatant.
12. The pellet was kept for air dry under laminar air flow.
13. The pellet was re-suspended in Nuclease free water stored either in -20° C or - 86° C for immediate use or long preservation.

3.2.1.2 ISOLATION FROM BLOOD

Blood samples typically were obtained as 1 ml of whole blood stored in EDTA vacutainer tubes frozen at 4° C.

1. The blood sample was pipetted in a sterilized microcentrifuge tube containing 500 µl of Lysis buffer.
2. To the above sample, added 50 µl of 10% SDS and mixed thoroughly for few minutes and added 2-3 µl of Proteinase-K.
3. The samples were mixed thoroughly by inverting the microcentrifuge tube for few minutes and it is then incubated at 56° C for 30 minutes.
4. Added equal volume of phenol: chloroform: Isoamyl alcohol (25:24:1) mixed thoroughly for few minutes.
5. The samples were centrifuged for 10 minutes at 12,000 rpm.
6. The aqueous layer was carefully removed to a new sterilized microcentrifuge tube, added 500 ml of Chloroform: Isoamyl alcohol (24:1) and mixed thoroughly by repeated inverting the microcentrifuge tube for few minutes.
7. It was then centrifuged at 12,000 rpm for 10 minutes in a centrifuge machine.
8. The upper aqueous layer was transferred in a fresh sterilized microcentrifuge tube and added double the volume of chilled absolute ethanol.
9. The above sample was kept at -20° C for overnight for precipitation.
10. The above sample was centrifuged at 10,000 rpm for 10 minutes.
11. Decanted the supernatant and retained the pellet.
12. To the pellet, 500 µl of 70% ethanol was added and again centrifuged at 7000 rpm for 10 minutes.
13. The supernatant was decanted and the pellet was kept for air dry under laminar air flow.
14. The pellet was re-suspended in Nuclease free water stored either in -20° C or - 86° C for immediate use or long preservation.

3.2.2 PURITY CHECK AND YIELD OF THE EXTRACTED DNA

The DNA isolated to be applied for downstream process such as PCR should be pure i.e. free from most of the associated proteins that keeps DNA *CO1* led and should also be in adequate quantity. Therefore, before undergoing for PCR with the extracted DNA, the purity and yield of the DNA was checked by the processes as follows:

3.2.2.1 SPECTROPHOTOMETRIC DETERMINATION

Spectrophotometric determination of purity and yield of the extracted DNA may be explained stepwise as follows:

1. The isolated DNA stock solution was taken and incubated in 56°C for 20 minutes.
2. 50µl of nuclease free water was taken in a cuvette and calibrated the spectrophotometer at 260nm as well as 280nm as blank.
3. 2 µl of nuclease free water was removed and 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 Plus, Eppendrof) against nuclease free water as blank. The yield and purity of DNA samples were estimated as follows:
5. Concentration of DNA stock solution (µg/ml)= OD_{260} X 100 (dilution factor) X 50 µg/ml/1000
6. 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)
7. From the concentration of DNA stock solution, the total yield of DNA was calculated and recorded.

3.2.2.2 AGAROSE GEL ELECTROPHORESIS

The stepwise procedure followed for agarose gel electrophoresis was as follows:

1. Agarose waspoured in electrophoresis buffer (1X TAE) to the desired concentration and mixed properly and then melted in a microwave oven.
2. Ethidium bromide (10 mg/ml) was added to the molten agarose gel (final concentration 0.5 µg/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 gel 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 where samples to be loaded
5. The gel, along with its casting tray, was inserted horizontally into the electrophoresis chamber and immersed with 1X TAE buffer (running buffer).
6. The DNA samples were mixed with 6X gel loading buffer and pipetted into the sample wells, the lid and power cableswere fitted 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 visually monitored by observing the 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 was achieved (around ¾), DNA fragments were visualized in UV-transilluminator stained with Ethidium Bromide
9. For documentation of the visualized DNA, the gel was placed in Gel-Documentation system (BioRad XR) and photograph of the gel electropherogram was taken with controlled exposure of UV under default parameters.
10. 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.

3.2.3 PCR AMPLIFICATION OF BARCODE SEGMENT

As mentioned above published primers of Ward et al. 2005 was used to amplify the barcode segment of *CO1* gene in a Veriti Mastercycler (Applied Biosystems Inc., CA, USA).

***CO1* DNA PCR reaction setting:**

Each 50 µl PCR reaction mixers contain:

10X PCR Buffer	:	5µl
dNTPs mix (10 mM)	:	5µl
Forward prim (20 pmole/µl)	:	1µl
Reverse primer (20 pmole/µl)	:	1µl
High fidelity DNA polymerase (5 Unit/µl)	:	0.2µl
Genomic DNA (100-200 ng)	:	variable
Nuclease free water	:	upto 50 µl

***CO1* DNA-PCR cycling condition:**

The PCR 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 5µl of DNA products from PCR amplification were loaded in 1.5 % agarose gel for electrophoresis in 1X TAE. Gel was stained with ethidium bromide and observed under UV transilluminator and documented with Gel-DOC (BioRad).

3.2.4 GEL-ELECTROPHORESIS AND PURIFICATION OF PCR PRODUCT

The PCR-amplified products were analyzed in 1% low melting agarose gel containing Ethidium Bromide staining (10 mg/ml). The single uniform bandwas excised and purified using GeneJET Gel extraction and DNA Cleanup micro kit (Fermentas, USA), following manufacturer's instructions given below.

Steps:-

- Excised the DNA fragment from the agarose gel with a clean, sharp scalpel.
- Transferred the gel slice to a 1.5 ml or 2.0 ml tube. Added then 650 µl Gel solubilizer.
- Incubated the tube for 10 minutes at 50ºC in a water bath until the gel slice had completely dissolved (3-4 times during incubation).
- Added 50 µl Binding Optimizer to the sample. Mixed by pipetting up and down, or by vortexing.
- Transferred 750 µl of the sample to Spin Column A placed in a 2ml collection tube. Centrifuged at 10,000 x g (12,000 rpm) 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.
- Added 700 µl Wash Buffer A and centrifuged at 10,000 x g (12,000 rpm) for 1 minute. Discarded the filtrate and reused the collection tube by placing the spin column back in the collection Tube.
- The above step was repeated.
- Centrifuged at maximum speed for 2 minutes to remove all traces of ethanol and the collection tube was discarded.
- Placed spin column A into a 1.5 ml Elution Tube and added 30-50 µl Elution buffer directly to the spin column membrane. Incubated at room temperature for 1 minute, centrifuged at 10,000 x g (12,000 rpm) for 1 minute to elute the DNA.
- The DNA was as such purified and stored at -20ºC.

3.2.5 SEQUENCING OF PCR AMPLICONS

The amplified products were purified and sequenced bi-directionally using automated DNA Sequencer (ABI 3500 Genetic Analyzer; Applied Biosystem, Inc. USA). The sequencing reaction was performed using 1 µl of BDTv3.1 Ready Reaction mix (Applied Biosystems), 1.5 µl of 5x Sequencing buffer; a final concentration of 0.5 pmol of each of the primers was maintained in separated reaction with the template DNA of concentration ranging between 70-100 ng/µl. The Chain termination reaction was carried out in a PCR Thermo-cycler at 25 cycles of 1 min at 94°C, 5 sec at 50°C and 4 min at 60°C as per manufacturer protocol. The fragments were then purified by Sodium Acetate/EDTA/Ethanol method prior to run in the 3500 Genetic Analyzer. The Sodium Acetate 3M, pH-5.2 and 125mM EDTA were added maintaining $1/10^{th}$ of the total volume of the product followed by the addition of double the volume of absolute ethanol and consecutive wash with 70% ethanol. The sample were dried properly and resuspended in Hi-Di Formamide (Applied Biosystems) prior to load in the capillary.

3.2.6 SEQUENCE ANALYSIS USING BIOINFORMATICS TOOLS

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 was detailed below

3.2.6.1 INTERPRETATION OF SEQUENCE CHROMATOGRAM

Automated DNA Sequencers generate 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-checking of the interpretation of the primary data is required. 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 Sequence Scanner 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 Genetic analyzer 3500, Applied Biosystems.

For each sample two chromatograms that represent sequences of both the strands of DNA were obtained. Ends of the noisy sequences were trimmed and more. MEGA-BLAST program 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 Sequence Scanner Version 1.0 (Applied Biosystems Inc., CA, USA). The selected fragments of the sequences for all the specimens were aligned using ClustalX software and subjected to MEGA-BLAST searches at the National Centre for Biotechnology Information (NCBI) to see whether the developed sequences align with database mitochondrial *CO1* 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 *CO1* barcode sequences.

3.2.6.2 SEQUENCE FORMATS AND ANALYZING 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 fast 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.

3.2.6.3 SEQUENCE SUBMISSION

All the analyzed 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 IDs.

3.2.6.4 GENBANK DNA BARCODE SEQUENCES RETRIEVED

In addition to the developed sequences, DNA barcode sequences representing same and related taxa of the ornamental species under study were acquired from GenBank as replicate datafor evaluating the taxonomic status of our target species and detailed in the following chapter.

3.2.6.5 ALIGNMENT OF THE SEQUENCES WITH CLUSTALX

ClustalX (Thompson et al. 2002) 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/). In each case, ClustalX provided a graphical user interface with colourful display of alignments. ClustalX performed the progressive alignment and created an output guide tree file and an output alignment file in the default Clustal format. ClustalX indicated the degree of conservation at the bottom of the aligned sequences, which was used to evaluate a given alignment.

3.2.6.6 SIMILARITY SEARCH

The sequence similarity search for species identification was done in two public databases, viz., BOLD and GenBank. The highest percent pair wise identity for each sequence blasted at NCBI was 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.

3.2.6.7 PHYLOGENETIC ANALYSIS

Kimura 2-parameter (K2P) (Kimura 1980) congeneric and conspecific variation, Neighbour Joining (NJ) and Maximum likelihood (ML) tree construction and Model test for ML were done using the computer program MEGA Version 6.2 (Tamura et al. 2011) for calculating the genetic divergence, selecting the best evolutionary model for sequence evolution and for phylogenetic re-construction. 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.

MEGA-6 has the advantage of calculating pair wise 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 ClustalX2) were inputted in MEGA and converted to MEGA format (.meg). The mega formats of the sequences were then used for pair wise genetic divergence calculation, for model test, NJ and ML phylogenetic inference following the MEGA 6.2 usage instruction.