

# CHAPTER-III

---

## Materials and Methods



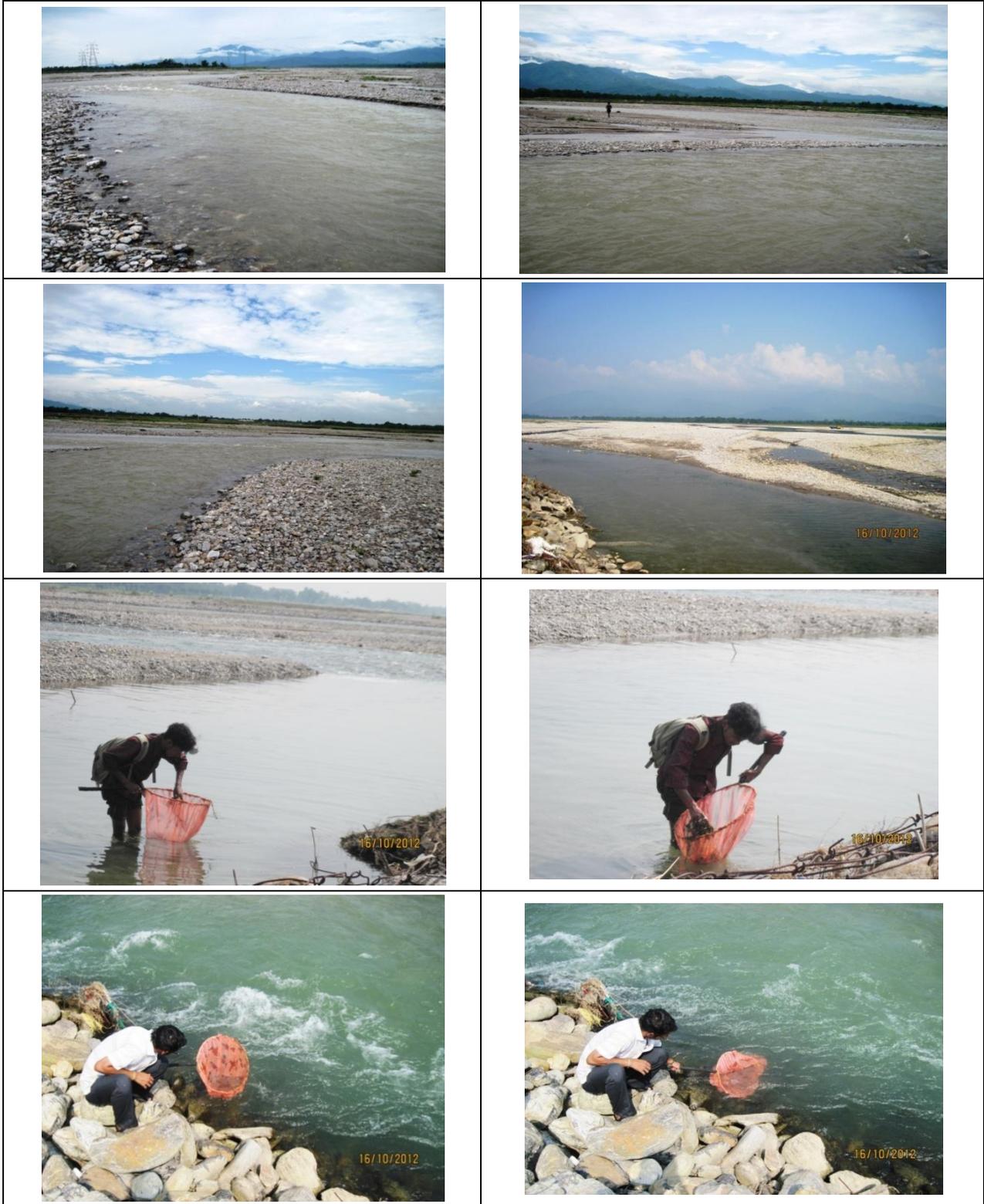
### 3. MATARIALS & METHODS

#### 3.1 Survey of Major River Systems of Terai and Dooars Region

A detailed survey has been carried out in the sub-Himalayan region of the West Bengal, India especially in the Terai and Dooars region. I have investigated the rivers, flowing through this region and the nearby landing centres to get the detailed picture of the habit and habitats of the studied fishes viz., *Badis badis* and *Amblyceps mangois*. After getting enough information about the fishes the major rivers of the sub-Himalayan Terai and Dooars region of the West Bengal India were considered for surveys and sample collection. The rivers under consideration were Mahananda river, Balason river, Teesta river and Jaldhaka river. The main tributaries of the rivers were also targeted for sample collection and to study the natural hierarchical distribution of the fishes. The collection sites with their geographic coordinates were recorded by hand-held Garmin GPS (eTrex Vista HCX, Garmin, USA) and are depicted in Figure 7 and Table 2.



A. Balasan River, Palpara, Matigara.



B. Balasan River, Tarabari.



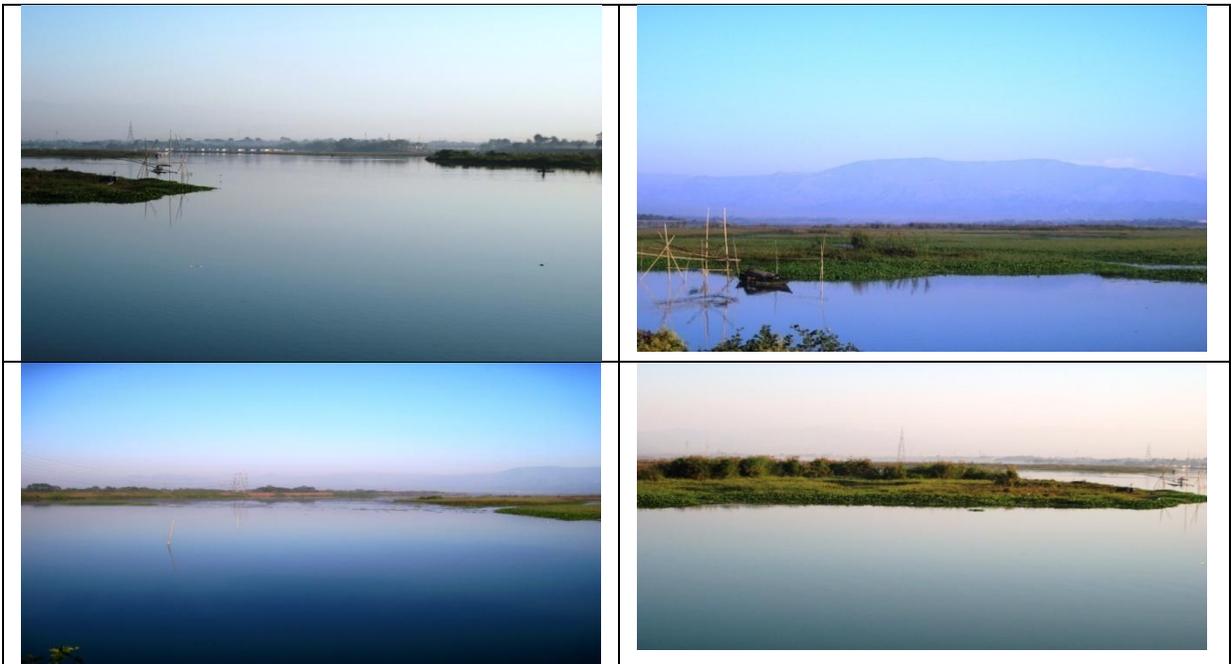
C. Mahananda River, Champasari, Siliguri.



D. Mahananda-Panchnoi River Junction, Siliguri.



**E. Panchnoi River, Siliguri.**



**F. Mahananda Barrage, Fulbari, Siliguri.**



**G. Teesta River, Sevoke, Near Siliguri.**



**H.** Ghish River, Jalpaiguri.



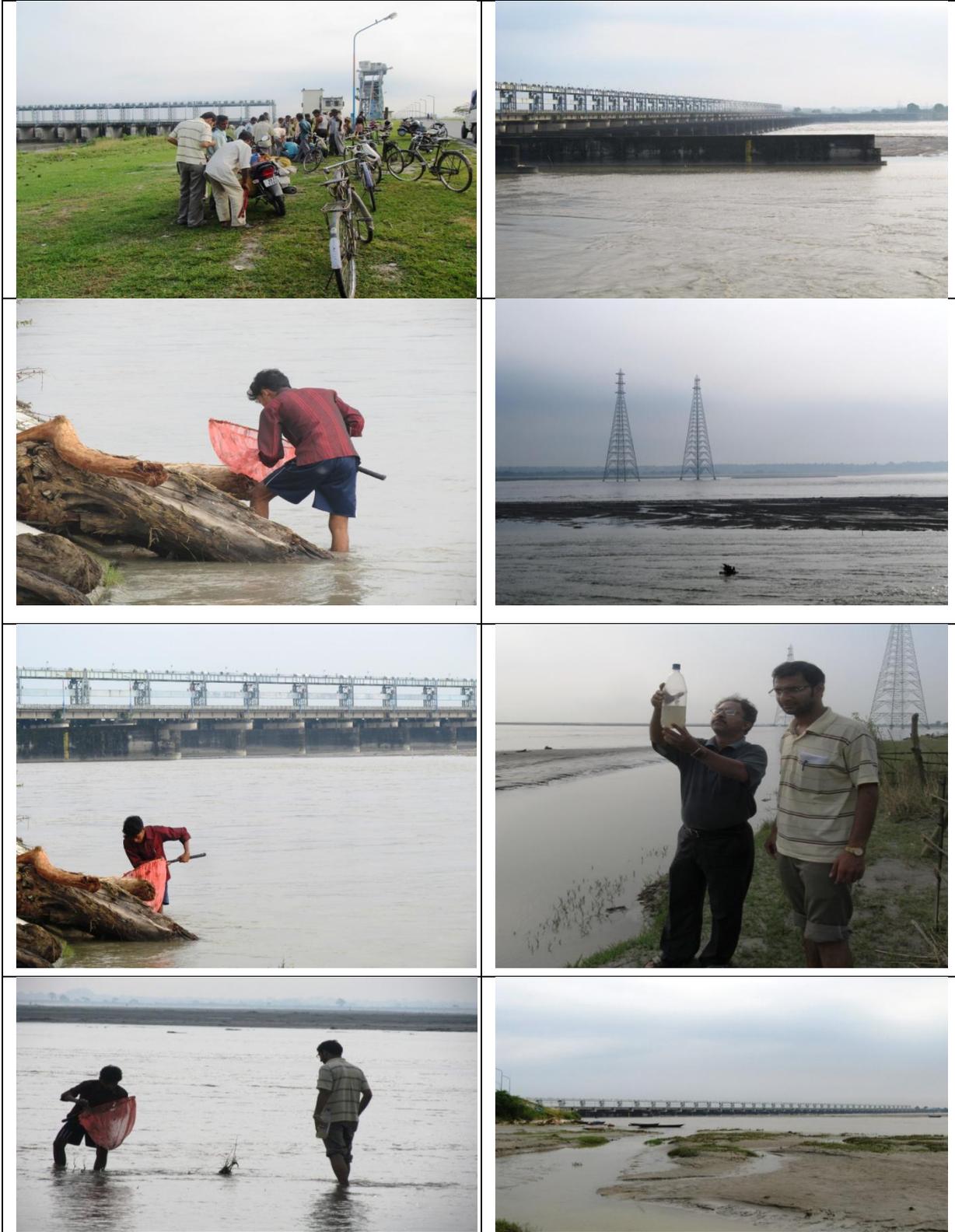
**I.** Chel River.



**J. Neora River.**



**K. Teesta River, Jalpaiguri.**



L. Teesta Barrage, Gajoldoba.



**M. Jaldhaka River.**



**N. Murti River.**



**O. Ghotia River.**



**P.** Diana River.

**Figure 7.** Photographic records of the sample collection sites (A to P).

<b>Table 2.</b> Geographical coordinates recorded with the help of hand-held GPS device			
<b>Sl/No.</b>	<b>Population Code</b>	<b>Name of the River and adjacent place</b>	<b>Geographic position and GPS Record</b>
1	<b>TR-1</b>	Mahananda Barrage, Fulbari.	26°38.884'N; 88°24.125'E; Elevation 319 AMSL
2	<b>TR-2</b>	Mahananda-Panchanoi River Junction, Siliguri.	26°42.125'N; 88°24.717'E; Elevation 620 AMSL
3	<b>TR-3</b>	Balason River, Palpara, Matigara.	26°43.177'N; 88°22.825'E; Elevation 646 AMSL
4	<b>TR-4</b>	Panchanoi River, Siliguri.	26°43.007'N; 88°24.396'E; Elevation 666 AMSL
5	<b>TR-5</b>	Mahananda River, Champasari, Siliguri.	26°44.452'N; 88°25.497'E; Elevation 717 AMSL
6	<b>TR-6</b>	Balason River, Tarabari.	26°45.632'N; 88°18.912'E; Elevation 731 AMSL
7	<b>DR-1</b>	Sevok (Teesta River)	26°53'043 N, 88°28'367E Elevation 480 AMSL
8	<b>DR-2</b>	Ghish River	26°52'327N, 88°36'355E Elevation 536 AMSL
9	<b>DR-3</b>	Gajoldoba (Teesta River Barrage)	26°44'584N, 88°35'314E Elevation 354 AMSL
10	<b>DR-4</b>	Chel River	26°51'499N, 88°38'048E Elevation 522 AMSL
11	<b>DR-5</b>	Neora River	26°52'486N, 88°46'205E Elevation 527 AMSL
12	<b>DR-6</b>	Dharla River	26°40'496N, 88°44'126E Elevation 299 AMSL
13	<b>DR-7</b>	Jalpaiguri (Teesta River)	26°33'499N, 88°45'369E Elevation 274 AMSL
14	<b>DR-8</b>	Jaldhaka River	26°34'13.17 N, 88°56'14.26 E Elevation 267 AMSL
15	<b>DR-9</b>	Murti River	26°52'57.73 N, 88°49'44.98 E Elevation 578 AMSL
16	<b>DR-10</b>	Ghotia River	26°52'14.89 N, 88°53'37.98 E Elevation 540 AMSL
17	<b>DR-11</b>	Diana River	26°51'37.96 N, 89°00'07.40 E Elevation 647 AMSL

\*AMSL= Above Mean Sea Level

### *3.2 Habit and Habitat Identification of the Fishes*

The fishes are mainly found in the river streams having moderate current in water. They are found mainly in shallow water near the river banks, among vegetation and under the stony rubbles. They are found mainly after the monsoon season. The *Badis badis* are mainly found hiding underneath the boulders, gravels, bushes or wooden logs distributed along the river banks. The *Amblyceps mangois* are mainly found hiding in the bushes, garbage and debris distributed along the river banks. The fishes are small in size (1-2inch in length). They are also very fast swimming fishes and difficult to net.

### *3.3 Sample collection and Preservation*

#### 3.3.1 Collection Procedure

I have used scoop net for collection of the samples. The scoop net was constructed by cutting the mosquito net and then tied it with a round shaped metal rod. A long bamboo-cane handle was attached with the metal rod for proper gripping purpose. The scoop net was waved against the water current or placed underneath/or beside the boulders and gently flapped the water around the boulder or bushes to trap the fishes (Figure 8).



**Figure 8.** Sample collection procedure of fishes from rivers with scoop net.

### 3.3.2. Preservation

Some of the fishes were collected to the laboratory within water bottle. It was ensured that the water inside the bottle get aerated all the time. Therefore, the mouth of the bottle was loosely tightened to get enough air for respiration of the fishes. Fishes were brought to the laboratory and proper identified. After identification, tiny amount of fin clips were taken noninvasively from the fishes and kept within the aquarium with proper aeration and adequate amount fish food. This was carried out to check whether cutting-out of tiny portion of fin clips affect/harm the fishes any way or not. But the fins were regenerated fully within 30-40 days and the fishes were healthy. Some type specimen of the fishes were kept in plastic pouches and stored at  $-20^{\circ}\text{C}$  for preservation and record.

### 3.4 Identification of Samples

3.4.1 *Badis badis* (Hamilton-Buchanan, 1822) [Shaw and Shebbeare, 1934, Talwar and Jhingran, 1991; Jayaram, K.C., 1994]

Etymology: *Badis*: Latin, badius = the colour of swift (Romero, 2002)

#### 3.4.1.1 Morphometric Measurement

1. Total number of Dorsal spines 16-20
2. Total number of soft rays 8-10 and Anal soft rays 6-9
3. Total vertebrae 26-28.
4. Body depth 30.7-38.9% SL.
5. Interorbital width 6,5-8,3% SL
6. Scales in lateral row 26-28
7. Pectoral rays 12-14
8. Contain a series of dark blotches along the dorsal fin base and/or a series of dark blotches along middle of dorsal fin and has an indistinct bars on side.

#### 3.4.1.2 Taxonomic Categorization

Kingdom- Animalia

Phylum-Chordata

Class-Actinopterygii (Ray-finned fishes)

Order-Perciformes (Perch-like)

Family-Badidae (Chameleon fishes)

Genus-*Badis*

Species- *B. badis*

### 3.4.1.3 Photographs



**Figure 9.** Samples of *Badis badis*

### 3.4.2 *Amblyceps mangois* (Hamilton-Buchanan, 1822)

[Shaw and Shebbeare, 1934, Talwar and Jhingran, 1991; Jayaram, K.C., 1994]

Etymology: *Amblyceps*: Greek, amblys = soft, darkness + Latin, ceps = head (Romero, 2002)

#### 3.4.2.1 Morphometric Measurement

1. Short body with 34-36 vertebrae.
2. Caudal fin with upper and lower lobes of distinctly different shapes.
3. Pectoral spine smooth.
4. Proximal lepidotrichia of the median caudal-fin rays has strongly-developed projections.

### 3.4.2.2 Taxonomic Categorization

Kingdom- Animalia

Phylum-Chordata

Class-Actinopterygii (Ray-finned fishes)

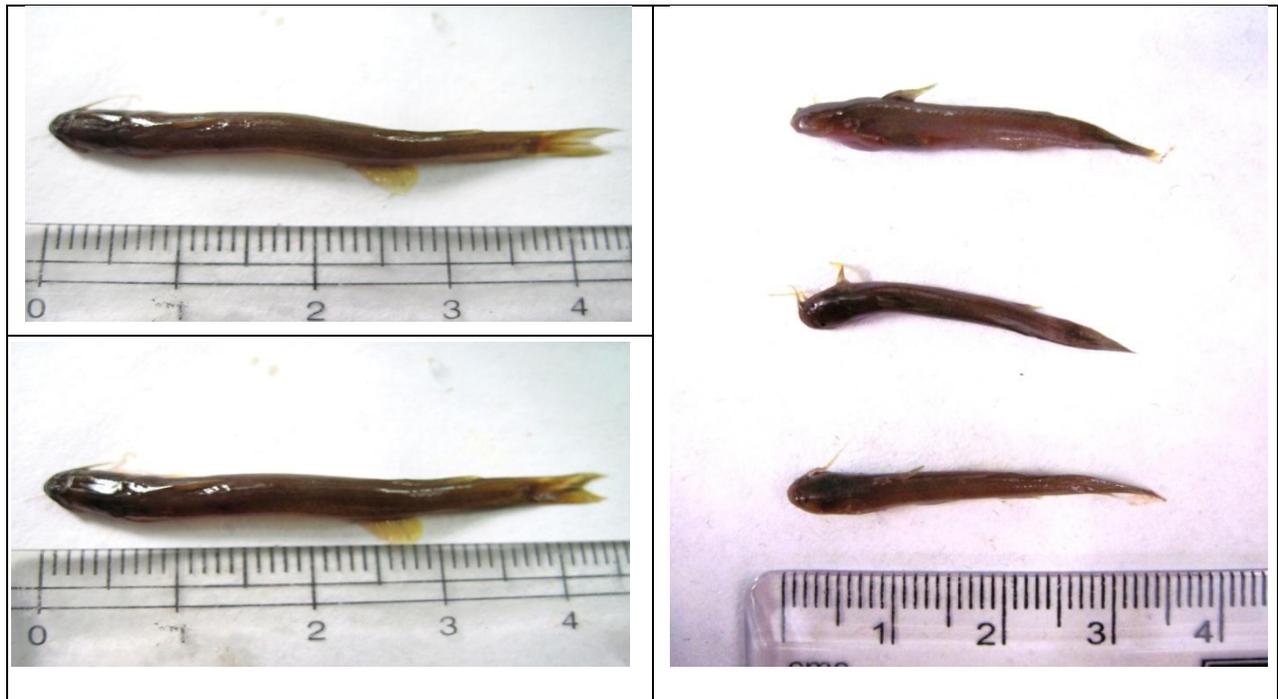
Order-Siluriformes (Catfish)

Family-Amblycipidae (Torrent catfishes)

Genus-*Amblyceps*

Species- *A. mangois*

### 3.4.2.3 Photographs



**Figure 10.** Samples of *Amblyceps mangois*

### 3.5 Genomic DNA Extraction

#### 3.5.1 Materials

##### 3.5.1.1 Phenol-Chloroform Procedure:

1. Fish Tissue (Fin/Scale)
2. Filter paper
3. Micro-centrifuge tube (1.5ml and 2.0 ml)
4. Centrifuge Machine
5. Water Bath
6. Lysis buffer

a. **Composition of Lysis buffer:**

- i. 200 mM Tris HCl, pH 8.0
- ii. 100 mM EDTA, pH 8.0
- iii. 250 mM NaCl

b. **Preparation of Lysis buffer:**

- i. 200 mM Tris HCl (pH 8.0) was prepared by dissolving 24.22gm of Tris base in 800 ml of distilled H<sub>2</sub>O. The pH was adjusted to the desired value of 8.0 by adding concentrated HCL. The solution was allowed to cool to room temperature before making final adjustment to the pH. The volume of the solution was adjusted to 1 litre with distilled H<sub>2</sub>O. The solution was dispensed into aliquots and sterilised by autoclaving. The pH of the solution is temperature dependent and decreases ~0.03 pH units for each 1°C increase in temperature. The solution was stored at 4°C.
- ii. 100 mM EDTA was prepared by dissolving 37.22gm of disodium EDTA.2 H<sub>2</sub>O to 800ml of distilled H<sub>2</sub>O and stirred vigorously on the magnetic stirrer. The pH of the solution was adjusted to 8.0 with NaOH pellets and sterilised by autoclaving. The solution was stored at 4°C. [Note: The disodium salt of EDTA will not go into solution until the pH of the solution is adjusted to ~8.0 by the addition of NaOH.]
- iii. 250 mM NaCl was prepared by adding 14.6gm of NaCl into 800ml of distilled H<sub>2</sub>O. The volume was adjusted to 1lt with distilled H<sub>2</sub>O. Then the solution was sterilised by autoclaving and stored at room at 4°C.

7. Proteinase K (20 mg/ml)

**Preparation of Proteinase K:** Proteinase K was initially centrifuged at 12000rpm for 30sec. The concentration of the solution adjusted to 20mg/ml by adding required amount of sterilized Milli-Q water.. The solution divided into small aliquots of 1 ml and stored at -20°C and store at -20°C.

8. SDS (20% w/v)

**Preparation of SDS:** SDS is also called sodium lauryl sulphate. 200gm of electrophoresis grade SDS was dissolved in 900 ml distilled H<sub>2</sub>O. The solution was heated to 68°C and stirred with a magnetic stirrer to assist dissolution. The pH of the solution was adjusted to 7.2 by adding few drops of concentrated HCl. The volume was adjusted to 1lt with distilled H<sub>2</sub>O and stored at room temperature. [Note: Sterilization and autoclaving is not required]

9. Water Saturated Phenol

10. Chloroform

11. Iso-amyl alcohol

12. Ammonium acetate (10 M)

**Preparation of Ammonium acetate:** 77 gm of ammonium acetate was added to 70 ml of distilled H<sub>2</sub>O and the volume was adjusted to 100 ml with distilled H<sub>2</sub>O at room temperature. The solution was sterilized by filtration by passing through a 0.22 µm filter and stored the solution in tightly sealed bottles at 4°C or at room temperature. [Note: Ammonium acetate decomposes in hot H<sub>2</sub>O and solution containing it should not be autoclaved.]

13. Ethanol (70%)

**Preparation of Ethanol:** 70 ml of dehydrated ethanol was added to 30 ml of distilled H<sub>2</sub>O and stored at room temperature. Before The ethanol was stored at -20°C to make it chilled before gDNA washing and precipitation.

14. Milli-Q water

Type I water was obtained from Elix pre-filtration and Milli-Q filtration unit.

### 3.5.1.2 DNA extraction Kit (DNase Easy Blood and Tissue DNA extraction kit, Qiagen)

1. ATL Buffer
2. Proteinase K
3. ALBuffer
4. Ethanol
5. DNeasy mini spin Column
6. Collection tube
7. AW1Buffer
8. AW2Buffer
9. AE Buffer

### 3.5.2 Methods

#### 3.5.2.1 Phenol-Chloroform Procedure:

A little portions of the fin (20-30 mg) were Clipped off (from the Caudal and ventral portion of the body) and the scales (30-50 pieces) were noninvasively (by gentle scraping from the dorsal portion of the body) from the live and frozen specimens (fish). The samples were dried on a filter paper before further processing. Genomic DNA extraction was carried out following Kumar et al. (2007) with minor modifications. A lysis buffer containing 200 mM Tris HCl, pH 8.0; 100 mM EDTA, pH 8.0; 250 mM NaCl was prepared and warmed at 50°C for 5–10 min. The tissues were then chopped and placed in a 2 ml micro-centrifuge tube containing 955 µl pre-warmed lysis buffer, 15 µl Proteinase K (20 mg/ml) (NEB, USA) and 30 µl 20 % SDS. Then the mixture was incubated at 52°C for 60–70 min in a water bath and. the tubes were gently stirred occasionally during incubation period. After the incubation equal volume of phenol: chloroform: isoamyl alcohol (25:24:1, v/v/v) was added to the tubes and mixed by gently inverting the tubes for about 15 min. The tubes were then centrifuged for 12 min at 9,500 rpm at 4°C in an Eppendorf refrigerated Centrifuge (Model-5424R, Germany). After centrifugation clear supernatants were taken in fresh tubes and equal volume of isopropyl alcohol and 0.2 volume of 10 M ammonium acetate were added. The mixture was incubated for 30–40 min at -20°C to precipitate the gDNAs. The gDNAs were pelleted by centrifugation at 13,500 rpm for 12 min at 4°C. The supernatants were decanted gently and pellets were washed once in 500 µl chilled 70% ethanol, air dried and re-suspended in 100 µl sterile Milli-Q water.

### 3.5.2.2 DNA extraction Kit (DNase Easy Blood and Tissue DNA extraction kit, Qiagen)

A little portion of the fin (20-30 mg) were clipped-off (from the caudal and ventral portion of the body) and the scales (30-50 pieces) were taken non-invasively (by gentle scraping from the dorsal portion of the body). Then the tissues were placed in a 1.5 ml micro-centrifuge tube and 180 µl of buffer ATL was added followed by 20 µl of Proteinase K. The samples were mixed thoroughly by vortexing and incubate at 56°C until the tissue was completely lysed. Then 200 µl of buffer AL was added to the sample and mixed thoroughly by vortexing. The tube was incubate at 56°C for 10 min. 200 µl of ethanol (96-100%) was added and mixed again thoroughly by vortexing. In the next step the mixture was pipetted out (including any precipitate) into a DNeasy mini spin column placed in a 2 ml collection tube. The tube was centrifuged at 8000 rpm for 1 min at 4°C. The collection tube was replaced and 500 µl buffer AW1 was added. Then the sample was centrifuged at 8000 rpm for 1min at 4°C followed by another replacement of the collection tube. In the same tube 500 µl buffer AW2 was added and centrifuged at 14000 rpm for 3 min at 4°C (to dry DNeasy membrane). Finally the collection tube is replaced by a clean 1.5ml micro-centrifuge tube and 200 µl of buffer AE was added directly onto the DNeasy membrane. Incubation was done at room temperature for 1 min followed by centrifugation at 8000 rpm for 1 min at 4°C (to elute the DNA). Then the extracted gDNAs samples were stored in 1.5ml micro-centrifuge tube in -20°C refrigerator.

## 3.6 Characterization of genomic DNA

### 3.6.1 Integrity checking by electrophoresis

#### 3.6.1.1 Materials

##### 1. Ethidium Bromide/EtBr (10mg/ml) [Stock Solution]

**Preparation of EtBr:** 1gm of Ethidium bromide was added to 100 ml of Milli-Q H<sub>2</sub>O. The solution was stirred on a magnetic stirrer for several hours to ensure that the dye has completely dissolved. The container was wrapped in aluminium foil and then the solution was transferred to a dark bottle and stored at room temperature.

##### 2. Agarose Gel (0.7%) (Lonza, Basel, Switzerland)

**Preparation of gel:** 0.7gm agarose was added in 98 ml distilled H<sub>2</sub>O. Then 2ml TAE buffer was added from 50X TAE stock solution. The solution was heated in microwave oven to

dissolve the agarose. Required amount of ethidium bromide was added from stock solution to ensure the final concentration of 0.5 µg/ml in the final gel solution.

### 3. TAE Buffer (50X) [Stock Solution]

#### Composition for 1 liter:

- i. 242gm Tris base
- ii. 57.1 ml glacial acetic acid
- iii. 100ml of 0.5M EDTA: 186.1 gm of disodium EDTA.2 was dissolved in 800ml of distilled H<sub>2</sub>O. The solution was stirred vigorously on the magnetic stirrer. The pH was adjusted to 8.0 with NaOH pellets and sterilised by autoclaving. The disodium salt of EDTA was not dissolved into solution until the pH of the solution was adjusted to ~8.0 by the addition of NaOH. The solution was stored at 4°C).

4. Gel casting tray (BenchTop Labsystems BT-MS- 300, Taiwan)

5. Electrophoretic Apparatus with power pack (BenchTop Labsystems BT-MS- 300, Taiwan)

6. 6X DNA loading Dye

#### Composition:

Buffer Type	Composition	Storage
I	0.25% (w/v) bromophenol blue 0.25% (w/v) xylene cyanol FF 40% (w/v) sucrose in H <sub>2</sub> O	4°C
III	0.25% (w/v) bromophenol blue 0.25% (w/v) xylene cyanol FF 30% (v/v) glycerol in H <sub>2</sub> O	4°C

7. Lambda DNA/Hind III Digest DNA size marker (New England Biolabs, USA)

8. UV-transilluminator (Spectroline BI-O-Vision, NY, USA)

10. Digital Camera [Nikon DSLR Camera (Model-D3100 with WF-S-18-55 VR Kit)]

### 3.6.1.2 Methods

The integrity of the gDNAs was checked by loading 10 µl extracted DNA with 1X DNA loading dye in ethidium bromide (0.5 µg/ml) pre-stained 0.7 % agarose gel (Lonza, Basel,

Switzerland) in 1X TAE tank buffer. Lambda DNA/Hind III double digest DNA size marker were used to estimate the expected size of the extracted genomic DNA. The gel was subjected to electrophoresis in constant current (90 mA) and voltage (90 V). The gels were checked on the UV-transilluminator (Spectroline BI-O-Vision, NY, USA) and then photographed using Nikon D3100 digital Camera.

### 3.6.2 Purity Checking

The gDNA samples were subjected to spectrophotometric assessment for purity checking at A260/A280 in Rayleigh UV2601 Spectrophotometer, Beijing, China. The ratio between ~1.7-~1.8 means the DNA is pure. If the ratio is lower, it indicates the presence of protein or phenol or other contaminants.

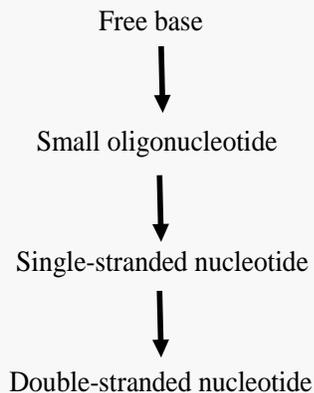
### 3.6.3 Quantification of PCR amplification

The gDNAs isolated from fish samples were treated with RNase A (USB, USA) at 37°C followed by phenol–chloroform purification and were quantified spectrophotometrically. Concentration of the extracted gDNA was estimated by the formula Concentration of DNA ( $\mu\text{g/ml}$ ) = OD at 260nm x dilution factor x 50/1000 and adjusted to ~50 ng/ $\mu\text{l}$  for PCR amplifications.

[\*Note: Spectrophotometric Analyses of DNA (Sambrook and Russell, 2001)

For quantifying the amount of DNA, reading was taken at wavelength of 260nm and 280nm. The reading at 260nm allows calculation of the concentration of nucleic acid in the sample. The absorption spectra of DNA and RNA are maximal at 260nm; absorbance data for nucleic acids were almost always expressed in  $A_{260}$  or  $OD_{260}$ . An OD of 1 corresponds to  $\sim 50\mu\text{g/ml}$  for double stranded DNA,  $\sim 40\mu\text{g/ml}$  for single stranded DNA and RNA, and  $\sim 33\mu\text{g/ml}$  for single stranded oligonucleotides.

The molar extinction co-efficient or molar absorption co-efficient ( $\epsilon$ ) decreases as the ring system of adjacent purines and pyrimidines become stacked in a polynucleotide chain. The value of  $\epsilon$  decreases in the following series



The concentration of stock DNA was determined by the following formula; Concentration of DNA ( $\mu\text{g/ml}$ ) =  $OD_{260} \times \text{dilution factor} \times 50 / 1000$ .

The ratio between the reading at 260nm and 280nm ( $OD_{260} : OD_{280}$ ) provides the purity of the nucleic acid. Pure preparations of DNA and RNA have  $OD_{260} : OD_{280}$  values of 1.8 and 2.0 respectively. If there was a significant contamination of phenol or protein, the  $OD_{260} : OD_{280}$  will be less than the values given above.]

### 3.7 RAPD-PCR amplification

#### 3.7.1 Selection of primers

Forty arbitrary decamer RAPD primers of random sequences (Kit-A and Kit-B, twenty primers from each kit) were purchased from Imperial Life Science Pvt. Ltd., India. Firstly, all the populations were screened with the forty primers and finally twenty (ten primers from Kit-A and ten primers from Kit-B) were selected for further analyses of *Badis badis* and *Amblyceps mangois* population on the basis of the variability and reproducibility of the bands obtained. The GC content of the primers was between 60-70%. The selected primer name and sequence for the study of *Badis badis* and *Amblyceps mangois* population were given in Tables 3 and 4.

Sl/No.	Primer	Sequence (5'→ 3')	G+ C Content (%)	Annealing temperature
1	OPA-01	CAGGCCCTTC	70	36°C
2	OPA-02	TGCCGAGCTG	70	
3	OPA-04	AATCGGGCTG	60	
4	OPA-07	GAAACGGGTG	60	
5	OPA-09	GGTAACGCC	70	
6	OPA-10	GTGATCGCAG	60	
7	OPA-13	CAGCACCCAC	70	
8	OPA-16	AGCCAGCGAA	60	
9	OPA-19	CAAACGTCCG	60	
10	OPA-20	GTTGCGATCC	60	
11	OPB-01	GTTTCGCTCC	60	
12	OPB-03	CATCCCCCTG	70	
13	OPB-04	GGACTGGAGT	60	
14	OPB-06	TGCTCTGCCC	70	
15	OPB-07	GGTGACGCAG	70	
16	OPB-11	GTAGACCCGT	60	
17	OPB-12	CCTTGACGCA	60	
18	OPB-15	GGAGGGTGTT	70	
19	OPB-17	AGGGAACGAG	60	
20	OPB-18	CCACAGCAGT	60	

**Table 4.** RAPD Primers for *Amblyceps mangois*

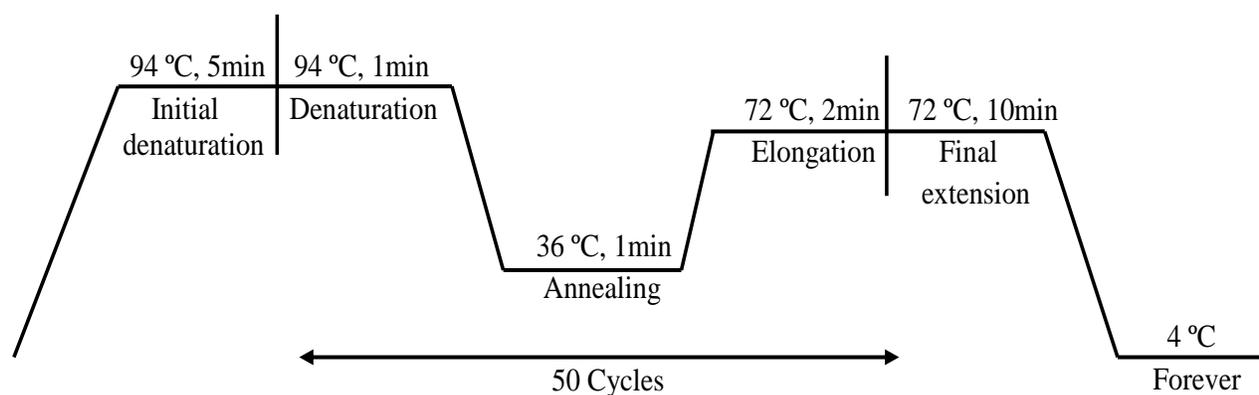
Sl/No.	Primer	Sequence (5'→ 3')	G+ C Content (%)	Annealing temperature
1	OPA-01	CAGGCCCTTC	70	36°C
2	OPA-02	TGCCGAGCTG	70	
3	OPA-04	AATCGGGCTG	60	
4	OPA-06	GGTCCCTGAC	60	
5	OPA-08	GTGACGTAGG	60	
6	OPA-10	GTGATCGCAG	60	
7	OPA-12	TCGGCGATAG	60	
8	OPA-13	CAGCACCCAC	70	
9	OPA-14	TCTGTGCTGG	60	
10	OPA-16	AGCCAGCGAA	60	
11	OPB-01	GTTTCGCTCC	60	
12	OPB-04	GGACTGGAGT	60	
13	OPB-06	TGCTCTGCCC	70	
14	OPB-07	GGTGACGCAG	70	
15	OPB-08	GTCCACACGG	70	
16	OPB-09	TGGGGGACTC	60	
17	OPB-10	CTGCTGGGAC	60	
18	OPB-11	GTAGACCCGT	60	
19	OPB-12	CCTTGACGCA	60	
20	OPB-17	AGGGAACGAG	60	

### 3.7.2 PCR reaction mixture

RAPD-PCR amplifications were performed in a 96 well Eppendorf® thermal cycler (USA Scientific, USA) and 96 well Peltier Thermal Cycler (Applied Biosystems 2720, Life Technologies, USA) in a final reaction volume of 25 µl, each containing a final concentrations of ~100-150 ng of isolated gDNA, 1.6 pM of oligonucleotide primers (RAPD), standard Taq polymerase buffer (10mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>) (NEB, USA), 200 µM of each dNTPs (dATP, dTTP, dCTP, dGTP) (NEB, USA), and one unit of Taq DNA Polymerase (NEB, USA). After the standardization of each reaction regime, all PCR amplifications were replicated to verify reproducibility and authenticity of the DNA bands.

### 3.7.3 PCR program

PCR cycling programs were as follows: initial denaturation at 94° C for 5 min followed by 50 cycles of 94°C, 1 min for denaturation; 36°C, 1 min for annealing; 72 °C, 2 min for elongation and finally an extension at 72 °C for 10 min.



### 3.8 ISSR PCR amplification

#### 3.8.1 Selection of primers

Twenty-one ISSR primers, purchased from Xcelris Genomics, Ahmadabad, Gujarat, India, were used to screen all the populations; and finally, fifteen ISSR primers for *Badis badis* and twelve ISSR primer for *Amblyceps mangois* population study were selected for further analyses on the basis of the variability and reproducibility. The primer names, sequences and GC contents are depicted in Tables 5 and 6.

Sl/No.	Primer	Sequence (5'→ 3')	G+ C Content (%)	Annealing temperature
1	ISSR-1	(CT) <sub>8</sub> TG	50	45°C
2	ISSR-2	(CT) <sub>8</sub> AC	50	
3	ISSR-3	(CT) <sub>8</sub> GC	55.6	
4	ISSR-4	(CA) <sub>6</sub> AG	50	33°C
5	ISSR-5	(CA) <sub>6</sub> AC	50	
6	ISSR-7	(CAC) <sub>3</sub> GC	72.7	
7	ISSR-8	(GAG) <sub>3</sub> GC	72.7	
8	ISSR-9	(GTG) <sub>3</sub> GC	72.7	42°C
9	ISSR-10	(GGAC) <sub>3</sub> T	69.2	
10	ISSR-11	(GA) <sub>6</sub> CC	57.1	37°C
11	ISSR-13	(GT) <sub>6</sub> CC	57.1	
12	ISSR-15	(ACC) <sub>6</sub>	66.66	33°C
13	ISSR-19	(GGAC) <sub>3</sub> A	69.2	42°C
14	ISSR-20	(GGAC) <sub>3</sub> C	76.9	
15	ISSR-21	(GGAC) <sub>4</sub>	75	45°C

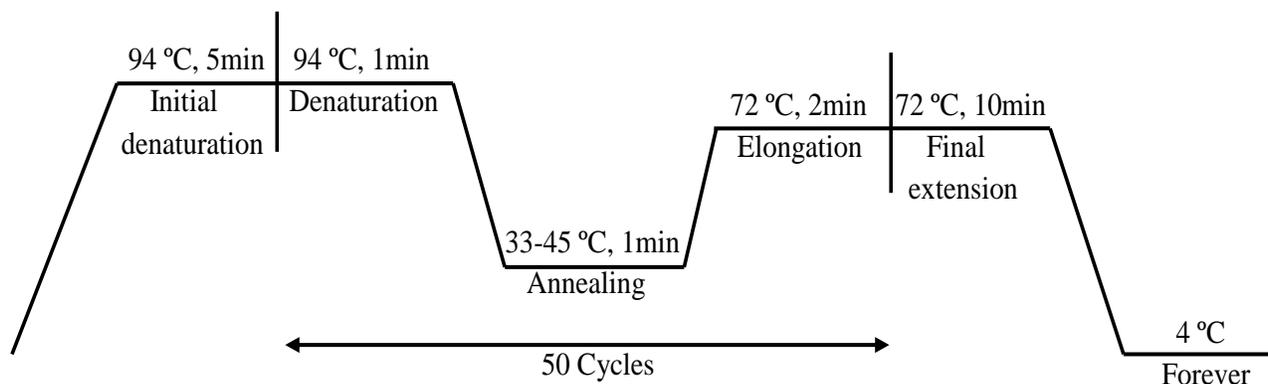
Sl/No.	Primer	Sequence (5'→ 3')	G+ C Content (%)	Annealing temperature
1	ISSR-1	(CT) <sub>8</sub> TG	50	45°C
2	ISSR-2	(CT) <sub>8</sub> AC	50	
3	ISSR-3	(CT) <sub>8</sub> GC	55.6	
4	ISSR-4	(CA) <sub>6</sub> AG	50	33°C
5	ISSR-5	(CA) <sub>6</sub> AC	50	
6	ISSR-7	(CAC) <sub>3</sub> GC	72.7	
7	ISSR-8	(GAG) <sub>3</sub> GC	72.7	
8	ISSR-9	(GTG) <sub>3</sub> GC	72.7	
9	ISSR-13	(GT) <sub>6</sub> CC	57.1	37°C
10	ISSR-19	(GGAC) <sub>3</sub> A	69.2	42°C
11	ISSR-20	(GGAC) <sub>3</sub> C	76.9	
12	ISSR-21	(GGAC) <sub>4</sub>	75	45°C

### 3.8.2 PCR reaction mixture

ISSR-PCR amplifications were performed in a 96 well Eppendorf® Thermal Cycler (USA Scientific, USA) and 96 well Peltier Thermal Cycler (Applied Biosystems 2720, Life Technologies, USA) in a final reaction volume of 25µl, each containing a final concentration of ~100-150 ng of isolated gDNA, 1.6 pM of oligonucleotide primers, standard Taq polymerase buffer (10mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>) (NEB, USA), 200 µM of each dNTPs (dATP, dTTP, dCTP, dGTP) (NEB, USA), and one unit of Taq DNA Polymerase (NEB, USA). PCR amplifications were replicated to verify reproducibility and authenticity of the DNA bands.

### 3.8.3 PCR program

PCR cycling programs were as follows: initial denaturation at 94°C for 5 min followed by 40 cycles of 94°C, 1 min for denaturation; 38°C - 47°C (specific and optimal annealing temperature for each primer), 1 min for annealing; 72 °C, 2 min for elongation and finally an extension at 72 °C for 10 min.



### 3.9 Electrophoresis and Checking of amplified products

#### 3.9.1 Materials

1. Ethidium bromide (10mg/ml)
2. Agarose (1.4 % for RAPD analyses and 1.6% for ISSR analyses) (Lonza, Basel, Switzerland)

**Preparation:** 1.6 gm agarose was added in 98 ml distilled H<sub>2</sub>O. 2ml TAE buffer was added from 50X stock solution. The solution was heated in microwave oven or electric heater to dissolve the agarose. Required amount of ethidium bromide was added from the stock solution to ensure the final concentration of 0.5 µg/ml in the final gel solution.

3. TAE Buffer (50X )
4. Electrophoretic apparatus with power pack (BenchTop Labsystems BT-MS- 300, Taiwan).
5. Standard 100bp DNA size marker (100-1500 bp, New England Biolabs, USA ).
6. DNA loading dye (6X)

## Composition:

Buffer Type	Composition	Storage
I	0.25% (w/v) bromophenol blue 0.25% (w/v) xylene cyanol FF 40% (w/v) sucrose in H <sub>2</sub> O	4°C
III	0.25% (w/v) bromophenol blue 0.25% (w/v) xylene cyanol FF 30% (v/v) glycerol in H <sub>2</sub> O	4°C

### 3.9.2 Methods

- I. The amplified products were electrophoresed in an ethidium bromide (0.5µg/ml) pre-stained 1.4 % (for RAPD) and 1.6% (for ISSR) (w/v) agarose gel (Lonza, Basel, Switzerland) at a Constant voltage 100 V and Current 100 mAmp in 1X TAE buffer.. The molecular weight of each band was estimated using a standard 100 base pair ladder. Reproducibility of the amplification pattern in each primer was checked by repeating each reaction thrice without deliberately altering the protocol.

### 3.9.3 Photographic records

The gels were visualized on the UV-transilluminator (Spectroline BI-O-Vision®NY, USA) and photographed using a Nikon D3100 Digital Camera.

### 3.10 *Statistical and Analytical Procedures to study genetic diversity*

Polymorphism data can be scored in presence/absence matrices manually or with the aid of specific software. The banding patterns produced, largely depended on the nature of the methods and whether they generate dominant or co-dominant markers. There are two features fingerprinting markers that considerably constrain how they can be statistically analyzed. Firstly, polymorphic marker loci are generally scored as two alleles, the “band-presence” allele and the “band absence” allele. Secondly, the banding patterns produced by each technique are treated as dominant markers during the evaluation process, and the methods are described accordingly. Thus, it is difficult to distinguish heterozygous individuals from homozygous ones with respect to the band-presence allele.

### 3.10.1 Band-based approaches

The easiest way to analyze banding patterns and to measure the diversity represented by them is to focus on band presence or absence and to compare it between samples. This method is routinely used at the level of the individual, and produces distances generated from the bands rather than taking into account the diversity of the population. The disadvantage is that it treats the data as polymorphism produced by the markers rather than describing the genetic diversity of the organism.

### 3.10.2 Polymorphism or polymorphic loci

A locus is considered polymorphic when the band is present at a frequency of between 5% and 95%. The polymorphism represented in the banding patterns can be extracted simply by observing the total number of polymorphic bands (*PB*) and then calculating the percentage of polymorphic bands (*PP*) present in any individual based on this number. This “band informativeness” (*I<sub>b</sub>*) can be represented on a scale ranging from 0 to 1 according to the formula:  $I_b = 1 - (2 \times |0.5 - p|)$ , where *p* is the portion of genotypes containing the band.

### 3.10.3 Hardy-Weinberg equilibrium Test (HWT)

Hardy-Weinberg equilibrium holds that the both gene and genotype frequencies will be constant from generation to subsequent next generations (Lebate, 2000). Chi-square test is useful for determining whether the allelic frequencies are in HW equilibrium. The statistical test follows this formula (Guo and Thompson, 1992):  $HWT = \sum \frac{(O_i - E_i)^2}{E_i}$ ; when: HWT = Statistical test, *O<sub>i</sub>* = Observed frequencies, *E<sub>i</sub>* = Expected frequencies, *df* = Degree of freedom. If  $X^2_{cal} \leq X^2_{tab}$  then *H<sub>0</sub>* hypothesis is accepted, it means that allele frequencies for loci in a given population are HWT equilibrium, if  $X^2_{cal} \geq X^2_{tab}$  then *H<sub>0</sub>* hypothesis is rejected (Hartl and Clark, 1997).

#### 3.10.4 Average number of Alleles per locus

These measures provide complementary information to that polymorphism (Kimura and Crow, 1964):  $N = \frac{1}{k} \sum_{i=1}^k n_i$ , where  $k$  = Number of loci,  $n_i$  = Number of alleles detected by locus. This parameter has the best application in co-dominant markers.

#### 3.10.5 Effective Number of Alleles

The measure explains about the number of alleles that would be expected in a locus in each population (Kimura and Crow, 1964):  $A_e = \left( \frac{1}{\sum_{a=1}^k p_a^2} \right)$ , where  $p_a^2$  is the frequency of the  $a^{\text{th}}$  of  $k$  alleles. By taking allele frequencies into account, this descriptor of allelic richness is less sensitive to rare alleles.

#### 3.10.6 Average Expected Heterozygosity ( $H_e$ )

Average expected heterozygosity is the probability that at a single locus of a diploid organism any two alleles, chosen at random, are different from each other. It is an indicator of genetic diversity in a population (Kimura and Crow, 1964):

$$H_e = 1 - \frac{1}{m} \sum_{i=1}^m \sum_{a=1}^k p_a^2$$

where,  $m$  is the number of analyzed loci. The range of this parameter from 0-1 and it is maximized when there are many alleles at equal frequency.

#### 3.10.7 Shannon's Information Index ( $H'$ or $I$ )

Shannon's information index is a diversity coefficient widely used to quantify the degree of polymorphism provided by banding patterns (Lewontin, 1972). It can be calculated as  $I = -\sum p_i \log_2 p_i$ , where  $I$  is diversity and  $p_i$  is the frequency of a particular RAPD band. It is assumed that (a) the population is large, (b) random mating takes place, (c) the population is isolated from others, and (d) the effects of migration, mutation and selection are not relevant. Allele-frequency statistics are commonly supplemented with another estimation that allows for the distortion from HWE. These methods depend on the extraction of allelic frequencies from the data, which can be accomplished by using a number of different procedures (Lynch and

Milligan, 1994; Stewart and Excoffier, 1996). The value of the Shannon index obtained from the data usually ranges between 1.5 and 3.5, and sometimes surpasses 4 (Margalef, 1972). It is only when there is a huge number of species in the samples a high value of the index is produced. The fact that the Shannon index is so narrowly constrained in most cases can make interpretation difficult. Therefore, some investigators sidestep the problem by using  $e^I$  instead of  $I$  (Kaiser et al, 2000).

### 3.10.8 Allelic Diversity ( $A$ )

One of the easiest ways to measure genetic diversity is to quantify the number of alleles present. Allelic diversity ( $A$ ) is the average number of alleles per locus and is used to describe genetic diversity (Frankham et al., 2002). When there is more than one locus,  $A$  is calculated as the number of alleles averaged across all loci:  $A = n_i / n_l$ , where  $n_i$  is the total number of alleles over all loci and  $n_l$  is the number of loci.

### 3.10.9 Heterozygosity ( $H$ )

Heterozygosity can be regarded as the average portion of loci with two different alleles at a single locus within an individual. It is commonly extended to the whole or a sub-portion of an entire population and differentiated into observed and expected heterozygosity. Expected heterozygosity ( $H_E$ ) or Nei's gene diversity ( $D$ ) is the expected probability that an individual will be heterozygous at a given locus, in multi-locus systems over the assayed loci. In other words, it is the estimated fraction of all individuals that would be heterozygous for any randomly chosen locus (Nei, 1973). It is often calculated based on the square of the frequency of the null (recessive) allele as follows:  $H_E = 1 - \sum_i^n p_i^2$ , where  $p_i$  is the frequency of the  $i^{\text{th}}$  allele. Observed heterozygosity ( $H_O$ ) is the portion of genes that are heterozygous in a population. It is calculated for each locus as the total number of heterozygotes divided by sample size. Typically, values for  $H_E$  and  $H_O$  range from 0 (no heterozygosity) to nearly 1 (a large number of equally frequent alleles). Expected heterozygosity is usually calculated when describing genetic diversity, as it is less sensitive to sample size than observed heterozygosity. If  $H_O$  and  $H_E$  are similar (they do not differ significantly) mating in the populations is approximately random. If  $H_O$  is less than  $H_E$ , the population is inbreeding; if  $H_O$  exceeds  $H_E$ , the population has a mating system avoiding inbreeding.

### 3.10.10 *F*-statistics

*F*-statistics, or fixation indices were originally designed to measure the amount of allelic fixation by genetic drift. It can describe the inbreeding (coefficient) of an individual relative to the total population ( $F_{IT}$ ) or the inbreeding (coefficient) of an individual relative to the subpopulation ( $F_{IS}$ ), and can also express the “fixation” (index) resulting from comparing subpopulations to the total population ( $F_{ST}$ ) (Wright, 1977). Moreover, inbreeding and genetic drift allow determining fixation indices when expected and observed heterozygosity are known. Thus the three indexes can be calculated as:  $F_{IT} = 1 - (H_I/H_T)$ ;  $F_{IS} = 1 - (H_I/H_S)$  and  $F_{ST} = 1 - (H_S/H_T)$ ; where  $H_I$  is the average observed heterozygosity within each population,  $H_S$  is the average expected heterozygosity of subpopulations, assuming random mating within each population, and  $H_T$  is the expected heterozygosity of the total population, assuming random mating within subpopulations and no divergence of allele frequencies among subpopulations. Fixation indices are interrelated as follows:  $1 - F_{IT} = (1 - F_{ST})(1 - F_{IS})$  (Wright, 1977). It varies from 0 (no isolation) to 1 (complete isolation), but in the case of a bias towards polymorphic loci, lower values (e.g. 0.5) may indicate complete isolation between populations.

### 3.10.11 Gene Flow ( $Nm$ )

The parameter of gene flow ( $Nm$ ) is the product of the effective size of individual populations ( $N$ ) and the rate of migration among them ( $m$ ). However, commonly used approach is to make an estimation of  $Nm$  is based on the degree of genetic differentiation among populations by calculating *F*-statistics.  $F_{ST}$  is related to population size and migration rate, based on a standard island model, which is not applicable to every natural population, but provides an easy way to calculate gene flow as the following:  $Nm = (1/F_{ST} - 1)$  (McDonald and McDermott, 1993). Gene flow among fragmented populations is related to dispersal ability, so it is expected to provide information about the dispersal ability of the organism and the degree of isolation among populations.

### 3.10.12 Shannon’s Evenness measure

The degree of evenness in species abundance is calculated by the Shannon’s evenness measure. The maximum diversity ( $H_{max}$ ) could occur when all species has equal abundance, in

other words if  $H' = H_{\max} = \ln S$ . The ratio of observed diversity to maximum diversity can, therefore, be used to measure evenness (J') (Pielou, 1975) therefore,  $J' = H'/H_{\max} = H'/\ln S$ .

### 3.10.13 Heip's Index of Evenness

Heip (1974) reported that the evenness measure should be independent of species richness and that they should have a low value in contexts where evenness is very low. The proposed measure was  $E_{\text{Heip}} = (e^{H'} - 1) / (S - 1)$ . The minimum value of Heip's measure is 0 and that it registers 0.006 when an extremely uneven community is used (Smith and Wilson, 1996).

### 3.10.14 SHE Analysis

Buzas and Hayek (1996) and Hayek and Buzas (1997) pointed out that Shannon's index is simply the sum of the natural log of evenness value ( $\ln E$ ) [where  $E = e^{H'}/S$ ] and natural log of species richness ( $\ln S$ ). This decomposed the Shannon's index into its two components  $H' = \ln E + \ln S$ . This allows the investigators to interpret the changes in diversity pattern in a spatial or temporal scale. The SHE analysis describes a relationship between the S (species richness), H (diversity measured by Shannon's index) and E (evenness). When there is a large number of samples are counted or estimated there are five scenarios possibly occur (Hayek and Buzas, 1997).

- i.  $S1=S2, H1=H2, E1=E2$ ; identical richness, evenness and relative abundance.
- ii.  $S1=S2, H1 \neq H2, E1 \neq E2$ ; richness remains constant but evenness changes.
- iii.  $S1 \neq S2, H1=H2, E1 \neq E2$ ; H remains constant because changes in S and E offset each other.
- iv.  $S1 \neq S2, H1 \neq H2, E1=E2$ ; E remains constant but S and therefore H changes
- v.  $S1 \neq S2, H \neq H2, E1 \neq E2$ , H changes because differences in S and E do not offset each other.

I have divided the seventeen riverine populations in case of *Badis badis* into ten groups and fourteen riverine populations of *Amblyceps mangois* into nine groups according to the continuity of the water flow through the river system from upstream to downstream. These groupings form an altitudinal gradient throughout the river system and easily observed the

relationship between the diversity index with that of richness and evenness. This allows interpreting the changes in diversity pattern of the two studied fishes within the three river system of the Terai and Dooars region of West Bengal, India.

#### 3.10.15 Phylogenetic Diversity or Phylogenetic Tree

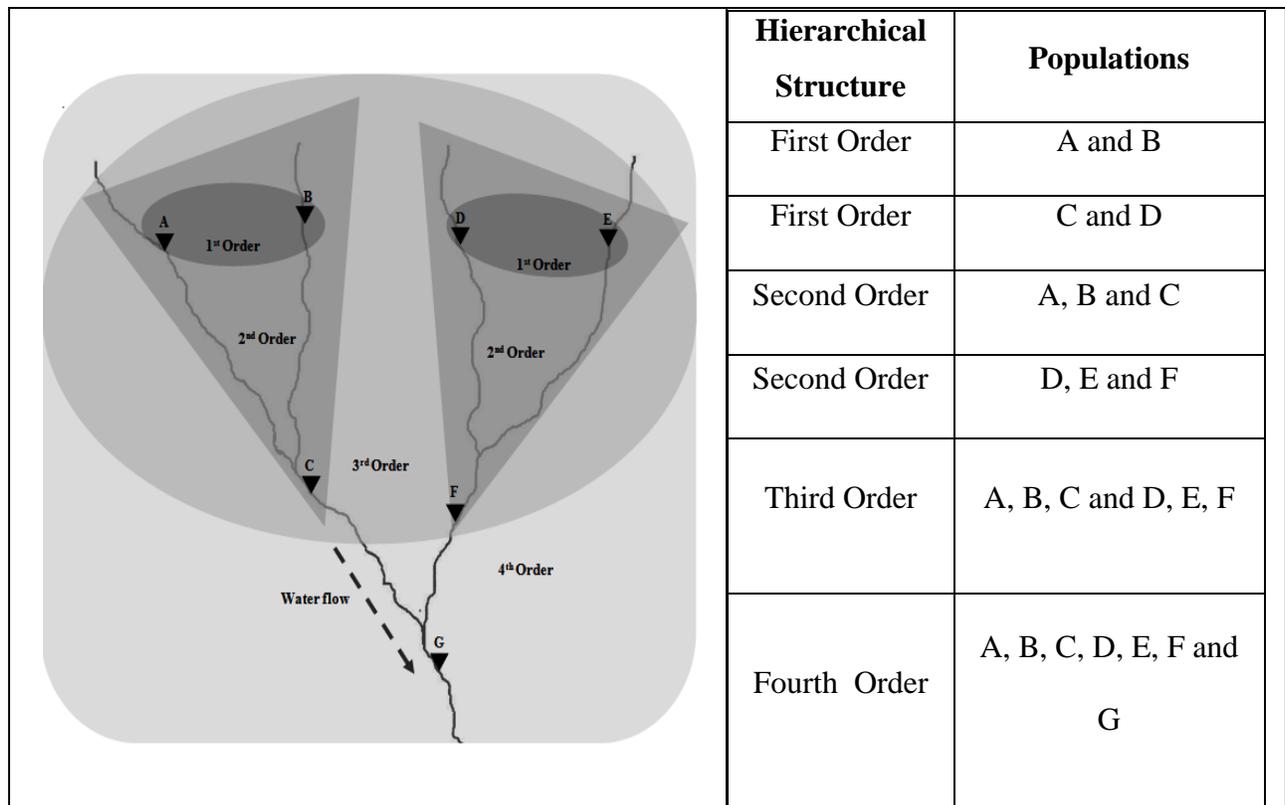
Besides genetic diversity estimation it is also obvious in molecular ecological studies to represent the genetic structure of the population(s) examined. Such information is extremely valuable because it can address many important issues such as estimation of migration pattern, identification of conservation units, or resolution of biogeographical patterns (Mantel et al., 2005). When geographical locations are available for populations and sampling allows the detection of spatial structure, then tree-based methods should be considered. These methods allow graphical representation of the relatedness of individuals and exploration of information on the spatial structure of populations. The simultaneous evaluation of the results of tree-based methods coupled with the distribution of individual markers within and among identified groups provides some information on the distribution of genetic diversity (Bonin et al., 2007). Knowledge of phylogenetic structure of genetic variation in populations is essential for understanding what the genetic diversity within a given species is likely to represent as part of the total diversity (Pleijel and Rouse, 2000).

Faith (1992) showed how the sum of the branch lengths on a minimum-spanning path between any set of terminals on a cladogram could be used to calculate a metric he termed Phylogenetic Diversity (PD) for that set. PD is important not only because species in isolated, monotypic lineages usually represent much more unique genetic diversity than ones in younger, more speciose groups, but also because the very nature of these species means that they may actually be disproportionately threatened by extinction (Purvis et al., 2000). Isaac et al. (2007) devised a measure they termed Evolutionary Distinctiveness (ED), which represents the relative contribution of an individual species to a clade's PD. They further showed how extinction risk could be combined with ED to identify species that are both Evolutionarily Distinct and Globally Endangered (EDGE) species. UPGMA (Unweighted Pair Group Method with Arithmetic mean) dendrogram based on Nei's (1978) unbiased genetic distance matrix was constructed by POPGENE v 1.2.1 software using default parameters (Yeh and Boyle, 1997).

### 3.10.16 Hierarchical Gene Diversity Analyses

Genetic hierarchy occurs because the divergence component of diversity ( $D_{pt}$ ) can be subdivided into geographic hierarchy in the distribution of species. Therefore, genetic divergence exists between different populations because of different level of gene flow among them. Populations that are geographically proximate show continuous gene flow and are genetically more similar than the more distant populations. Thus a species can be visualized as having a spatial genetic architecture.

Perhaps the clearer picture of genetic and geographic hierarchy is observed in riverine species because river forms a geographic structure with natural hierarchy (Meffe and Carroll, 1997). The river forms the first order, second order and so on hierarchical structure. The first order stream joins to form second order streams, and the second order streams join to form the third order streams and so on. In Figure 11, the first order hierarchy constitutes the A and B populations, and C and D populations. The second order hierarchy constitutes the A, B and C populations, and D, E and F populations. The third order hierarchy constitutes the A, B, C and D, E, F populations. Whereas the fourth order hierarchy constitute the A, B, C, D, E, F and G populations (Figure 11). Therefore, the overall diversity of any river stream is determined by calculating genetic diversity and genetic divergence of the individual hierarchical order of the river stream and finally adding the values of genetic diversity and genetic divergence of the individual first, second, third and fourth hierarchical orders.  $F_{ST}$ ,  $Nm$  and AMOVA (Analyses of Molecular Variance) among and within population was calculated to analyze the genetic hierarchical structure of the studied fishes across river system using GenAlex v 6.5 software (Peakall and Smouse, 2006; Peakall and Smouse, 2012). The genetic differentiation between populations was determined using  $PhiPT$  values.



**Figure 11:** Hypothetical genetic hierarchical map of seven different populations. The populations are arranged in a hierarchical order, viz., first order, second order, third order and fourth order populations. The different shaded areas represent different orders of the hierarchy. The triangles indicate the collection spots and curved line indicate the river streams (adapted from Meffe and Carroll, 1997).

### 3.11 mtCOI amplification and analyses

*Analyses of Mitochondrial Cytochrome Oxidase subunit I gene was undertaken as additional objective to support and substantiate RAPD-ISSR-based data.*

#### 3.11.1 Primer Selection

The primers were chosen from previously published report (Ward et al. 2005) and selected after nr-blast in NCBI database to amplify the cytochrome oxidase subunit 1 (COI) gene (approximately 700 bp long). The primers were synthesized by Imperial Life Science (Gurgaon, India). The primer sequences are follows:

**Forward: 5'- TCAACCAACCACAAAGACATTGGCAC-3'** (GC Content 42.30%)

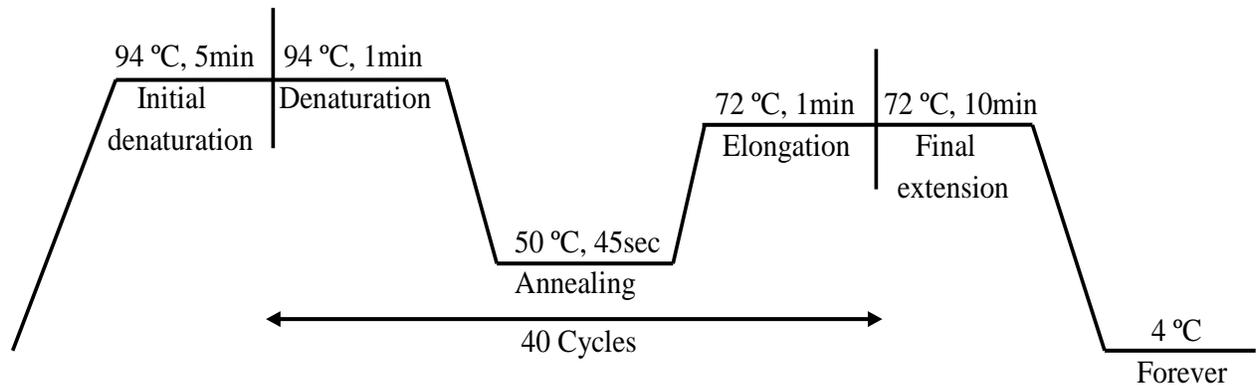
**Reverse: 5'- TAGACTTCTGGGTGGCCAAAGAATCA-3'** (GC Content 46.15%).

#### 3.11.2 PCR reaction mixture

Mitochondrial COI gene amplification was performed in a 96 wells Eppendorf® (Germany) Peltier thermal cycler. The final reaction volumes were 30µl, each containing a final concentration of ~100-150 ng of isolated gDNA, 1pM of each oligonucleotides primers (Forward and Reverse), 1X standard Taq Polymerase buffer (10mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>) (NEB, USA), 200 µM of each dNTPs (dATP, dTTP, dCTP, dGTP) (NEB, USA), and one unit of Taq DNA Polymerase (NEB, USA).

#### 3.11.3 PCR Programme

PCR cycling programs were as follows: initial denaturation at 94°C for 5 min followed by 40 cycles of 94°C, 1 min for denaturation; 50°C, 45 sec for annealing; 72°C, 1 min for elongation and finally an extension at 72°C for 10 min.



#### 3.11.4 Checking and documentation of amplified product

The amplified products were electrophoresed in an ethidium bromide (0.5 $\mu$ g/ml) pre-stained 1.5% (w/v) agarose gel (Lonza, Switzerland) at a constant voltage 100 V and current 100 mA in TAE buffer (40 mM Tris-HCl, pH 8.0; 20 mM Acetic acid; 1 mM EDTA, pH 8.0) using BenchTop Labsystems BT-MS-300 (Taiwan) electrophoretic apparatus. Molecular weight of each band was estimated using a standard 100 base pair ladder (NEB, USA). The gels were visualized on the UV-transilluminator (Spectroline BI-O-Vision<sup>®</sup>NY, USA) and photographed using a Nikon D3100 camera (Fig 2).

#### 3.11.5 Purification of PCR product, sequencing and submission of *Badis badis* and *Amblyceps mangois* cytochrome oxidase I (COI) partial coding sequences (CDS) to public database

The amplified products were purified using Invitrogen Purelink<sup>™</sup> PCR Purification Kit (Thermofisher Scientific, India), following manufacturer's protocol. The amplified products were confirmed as COI partial coding gene by restriction digestion using Hind III restriction enzyme (Hind III cutting point between 252<sup>th</sup>-253<sup>th</sup> bp). Each PCR product was sequenced at least twice by dye-dideoxy automated chain termination method (MWG-Biotech Pvt. Ltd., Bangalore India). All the COI partial CDS were used separately to search the GB/EMBL/DDBJ combined nr database in NCBI BLAST search through the implementation of Blastn and Blastx, optimized for highly similar sequences (Megablast), using stringent expect threshold (0.01) and excluding low complexity regions within the combined database. The curated sequences (approximately 650-690 bp) were submitted to GenBank and accession numbers were obtained (Table 7 and 8).

### 3.11.6 Retrieval of publicly available *Badis badis* and *Amblyceps mangois* COI partial CDS

Publicly available twentyfive *Badis badis* COI partial CDS corresponding to isolates submitted from northern, southern, north-easten India and Dhaka, Bangladesh were retrieved from GenBank (NCBI; National Centre for Biotechnology Information). The isolate names and their accession numbers are presented in Table 7. Moreover, publicly available four *Amblyceps mangois* COI partial CDS corresponding to isolates submitted from north-eastern India and Dhaka, Bangladesh were retrieved from GenBank (NCBI; National Centre for Biotechnology Information). The isolate names and their accession numbers are presented in Table 8.

**Table 7:** Retrieved publicly available 25 number of *Badis badis* COI gene sequence (No. 1-25) from NCBI database and ten number of presently studied sequences (No. 26-35) .

No.	Name	Accession No	Location	Broad Location Division	Institute
1	GPCR 5BB	KU043316.1	Ganges & Yamuna rivers, uttarkashi, garhwalthimalaya,uttarak hand.	North India	Zoology, Pt. LMS Govt. PG College, Rishikesh, Delhi Road, Rishikesh, Uttarakhand, India.
2	GPCR 6BB	KU043317.1			
3	GPCR 7BB	KU043318.1			
4	GPCR 8BB	KU043319.1			
5	GPCR 141BB	KR809715.1			
6	GPCR 282BB	KR809716.1			
7	GPCR 284BB	KR809717.1			
8	GPCR 287BB	KR809718.1			
9	NBFGR: BB8079A	FJ459452.1	Kulsi river, Rampur, Assam	N-E India	International Centre For DNA Barcoding, National Bureau of Fish Genetic Resources, Canal Ring Road, P. O. Dilkusha, Lucknow, Uttar Pradesh 226002, India
10	NBFGR: BB8079B	FJ459453.1			
11	NBFGR: BB8079C	FJ459454.1			
12	IASST F169	KY823511.1	Ranganadi river, Lakhimpur, Assam.	N-E India	Resource Management & Environment Section, Institute of Advanced Study in Science & Technology, Vigyan Path, Paschim Boragaon, Garchuk, Guwahati, Assam 781035, India
13	DOF 72	JN815304.1	Sunai river, karimganj, Assam		Assam University (Central University), Durgakona, Silchar, Assam 788011, India
14	DOF 73	JN815305.1			
15	DOF 89	JN815311.1			
16	BDB 1001	KJ936835.1	Jinjiram River, Lembucherra, Tripura West, Tripura		College of Fisheries, Central Agricultural University, Lembucherra, Tripura West, Tripura 799210, India
17	BDB 1002	KJ936836.1			
18	BDB 1003	KJ936837.1			
19	BDB 1004	KJ936838.1			
20	BDB 1005	KJ936839.1			
21	SGBK OFF29F	JQ713858.1	Tuipi river, Pherzawl, Manipur	Assam University (Central University), Durgakona, Silchar, Assam 788011, India	
22	WILD 15 PIS 144	KP666032.1	Cauvery river, Mysore, Karnataka	South India	Indian Institute of Science Education and Research, Pune, G1 Block, Dr. Homi Bhabha Road, Pashan, Pune, Maharashtra 411008, India
23	BabaH1	KC774640.1	Kochi, Kerala		Kochi Centre, National Bureau of Fish Genetic Resources, CMFRI Campus, PO. Ernakulam North, Kochi, Kerala 682 018, India
24	BabaH2	KC774641.1			
25	DUZM 196	KT364764.1	Tanguar Haor, Sunamganj, Dhaka, Bangladesh	Dhaka, Bangladesh	Department of Zoology, University of Dhaka, Karzon Hall, Dhaka, Dhaka 1000, Bangladesh.
26	DRBB 1	MF774665.1	Dooars region, Sub-Himalayan West Bengal.	Sub-Himalayan Terai and Dooars	THIS STUDY  Cell and Molecular Biology Laboratory, Dept. of Zoology, University of North Bengal, Raja Rammohanpur, Darjeeling, West Bengal, India.
27	DRBB 3	MF774666.1			
28	DRBB 6	MF774667.1			
29	DRBB 7	MF774668.1			
30	DRBB 8	MF774669.1			
31	DRBB 9	MF774670.1			
32	DRBB 11	MF774671.1			
33	TRBB 1	MF774672.1	Terai region, Sub-Himalayan West Bengal.		
34	TRBB 5	MF774673.1			
35	TRBB 6	MF774674.1			

**Table 8:** Retrieved publicly available four number of *Amblyceps mangois* COI gene sequence (No. 1-4) from NCBI database and ten number of presently studied sequences (No. 5-14).

No.	Name	Accession No	Location	Broad Location Division	Institute
1	UMMZ 244760	DQ508067.1	NA	NA	NA
2	BCF11	JX105483.1	Noadihang river, Assam, India	N-E India	NA
3	SGBK-OF31F	JQ713859.1	Thingou, Manipur		
4	DUZM140	KT762370.1	Sunamgan, Tanguar Haor, Bangladesh	Bangladesh	Department of Zoology, University of Dhaka, Karzon Hall, Dhaka 1000, Bangladesh
5	TRAM1	MK012251	Sub-Himalayan Terai region	Sub-Himalayan Terai and Dooars, Inmdia	THIS STUDY Cell and Molecular Biology Laboratory, Dept. of Zoology, University of North Bengal. Raja Rammohanpur, Darjeeling, West Bengal, India.
6	TRAM2	MK012252			
7	DRAM4	MK012253	Sub-Himalayan Dooars region		
8	DRAM6	MK012254			
9	DRAM7	MK012255			
10	DRAM8	MK012256			
11	DRAM9	MK012257			
12	DRAM10	MK012258			
13	DRAM12	MK012259			
14	DRAM14	MK012260			

### 3.11.7 Genetic diversity and population structure of two species

Levels of genetic variation of *Badis badis* and *Amblyceps mangois* within the Mahananda, Teesta, Jaldhaka and other populations were estimated in terms of the number of variable sites (S), total number of mutations (Eta), number of haplotypes (H), haplotype diversity (H), nucleotide diversity ( $\pi_n$ ) and average number of nucleotide differences. These indices were calculated by freely available DnaSP ver. 5.1 software (Librado and Rozas 2009). The genetic differentiation ( $F_{ST}$ ) and gene flow ( $N_m$ ) were also calculated following the method of Wright (1978), Govindaraju (1989) and Low et al. (2014). Tajima's  $D$  (Tajima 1989), Fu and Li's  $D$  and Fu and Li's  $F$  (Fu 1997) were calculated to verify selective neutrality in relation to mtDNA sequences, which would help to accumulate information regarding the demographic history of the population and used to deduce whether a population has undergone a sudden population expansion or construction.

### 3.11.8 Preparation of sequence dataset and phylogenetic analyses of mtDNA COI sequences

Raw sequence reads, each PCR product sequenced at least twice, were curated in BioEdit ver. 7.0.9 (Hall, 1999) for character uncertainty and then assembled in Geneious R7 ver 7.0.6 (trial) (Rozen and Skaletsky, 2000) using pro features and retrieved the final sequences as Fasta

files. Multiple sequence alignment was implemented by CLUSTALX in MEGA ver 7.10 (Kumar *et. al.*, 2016) using gap opening penalty 50, gap extension penalty 20, transition weight 0.5 and delay divergent cut-off 30%. The best model of DNA substitution selected was implemented in MEGA ver 7.10 (Kumar *et. al.*, 2016) following software default parameter settings.

### 3.11.9 Phylogenetic analyses based on a MtCOI partial CDSs of *Badis badis* and *Amblyceps mangois*

The evolutionary relationship between thirty five *Badis badis* (ten number of studied sequences and twenty five number of published sequences from NCBI ) and fourteen *Amblyceps mangois* (ten number of studied sequences and four number of published sequences from NCBI) mitochondrial cytochrome oxidase I sequences was estimated using phylogenetic inference and Maximum Likelihood (ML) method.

The best nucleotide substitution model for phylogenetic analyses of *Badis badis* was selected by model testing programme in MEGA7.10. The best nucleotide substitution model obtained was Kimura 2-parameter the following scores: Bayesian Information Criterion (BIC)=5863.595, Akaike Information Criterion corrected (AICc)= 5132.779, Maximum likelihood value (InL) = -2586.133, Assumed or estimated values of transition and transversion (R)=5.12, nucleotide frequencies (f): f(A)=0.250, f(T) =0.250, f(C) =0.250, f(G) =0.250; rates of base substitution (r): r(AT)=0.020, r(AC)=0.020, r(AG)=0.209, r(TA)=0.020, r(TC)=0.209, r(TG)=0.020, r(CA)=0.020, r(CT)=0.209, r(CG)=0.020, r(GA)=0.209, r(GC)=0.020, r(GT)=0.020. Kimura 2-parameter nucleotide substitution model was selected and total 500bootstrap replication was carried out during phylogenetic tree construction. The following default parameters are selected: uniform rates, 1<sup>st</sup>+2<sup>nd</sup>+3<sup>rd</sup>+noncoding codon positions, ML Heuristic method=Subtree-Pruning-Regrafting - Extensive (SPR level 5), Initial tree for ML= Make initial tree automatically (Default - NJ/BioNJ), strong branch swap filter and one number of thread.

The best nucleotide substitution model for phylogenetic analyses of *Amblyceps mangois* was selected by model testing programme in MEGA7.10. The best nucleotide substitution model obtained was Jukes-Cantor having the following scores: Bayesian Information Criterion (BIC)=2398.872, Akaike Information Criterion corrected (AICc)= 2222.754, Maximum likelihood value (InL) = -1086.300, Assumed or estimated values of transition and transversion

(R)=0.50, nucleotide frequencies (f): f(A)=0.250, f(T) =0.250, f(C) =0.250, f(G) =0.250; rates of base substitution (r): r(AT)=0.083, r(AC)=0.083, r(AG)=0.083, r(TA)=0.083, r(TC)=0.083, r(TG)=0.083, r(CA)=0.083, r(CT)=0.083, r(CG)=0.083, r(GA)=0.083, r(GC)=0.083, r(GT)=0.083. Jukes-Cantor nucleotide substitution model was selected and total 500bootstrap replication was carried out during phylogenetic tree construction. The following default parameters are selected: uniform rates, 1<sup>st</sup>+2<sup>nd</sup>+3<sup>rd</sup>+noncoding codon positions, ML Heuristic method=Subtree-Pruning-Regrafting - Extensive (SPR level 5), Initial tree for ML= Make initial tree automatically (Default - NJ/BioNJ), strong branch swap filter and one number of thread.

All the trees were built as midpoint-rooted consensus trees using FIGTREE v1.4.1 (<http://tree.bio.ed.ac.uk/software/figtree/>).

### 3.12 Table 9. Software used for statistical and genetic diversity analyses

Sl. No.	Software Program me	Operating Platform	Source	Description	Implemented In	Reference
	Arlequin v. 1.1	Windows	Freeware from <a href="http://anthropologie.unige.ch/arlequin">http://anthropologie.unige.ch/arlequin</a>	It can handle RFLPs, microsatellites, allozymes, RAPDs, AFLPs, allele frequencies, DNA sequences data. Calculates gene and nucleotide diversity, mismatch distribution, haplotype frequencies, linkage disequilibrium, tests of Hardy-Weinberg equilibrium, neutrality tests, pairwise genetic distances, analyses of molecular variance (AMOVA).	Calculation of genetic diversity indices, AMOVA.	Schneider <i>et al.</i> (1997)

Sl. No.	Software Program me	Operating Platform	Source	Description	Implemented In	Reference
	Fstat v. 1.2	DOS/ Windows	Freeware by writing to J. Goudetatjerome .goudet@izea.unil.ch	It can handle Allozymes, microsatellites, mtDNA RFLPs data. Calculates gene diversity statistics of Weir and Cockerham (Weir, 1996). Computes jackknife and bootstrap confidence intervals of the statistics or can test gene diversity statistics using a permutation algorithm.	Genetic diversity indices	Goudet (1995)
	POPGEN E v. 1.21,	Windows	Freeware from <a href="http://www.ualberta.ca/~fyeh/index.htm">http://www.ualberta.ca/~fyeh/index.htm</a>	It can handle Codominant or dominant markers using haploid or diploid data. Calculates standard genetic diversity measures, tests of Hardy-Weinberg Equilibrium, Wright's Fstatistics, genetic distances, UPGMA dendrogram, neutrality tests, linkage disequilibrium.	Calculation of Polymorphic loci, Nei's genetic diversity, Shannon's information index, UPGMA cluster analyses.	Yeh and Boyle (1997)
	TFPGA	Windows	Freeware from <a href="http://herb.bio.nau.edu/~miller">http://herb.bio.nau.edu/~miller</a>	It can handle Codominant (allozyme) and dominant (RAPD, AFLP) genotypes data. Calculates descriptive statistics, genetic distances, and F-statistics. Performs tests for Hardy-Weinberg equilibrium, exact tests for genetic differentiation, Mantel tests, and UPGMA cluster analyses.	Calculation of F statistics, UPGMA cluster analyses.	Miller (1997)

Sl. No.	Software Programme	Operating Platform	Source	Description	Implemented In	Reference
	GenAlEx 6.5	Windows, Mac OS X	Freeware from <a href="http://biology.anu.edu.au/GenAlEx">http://biology.anu.edu.au/GenAlEx</a>	GenAlEx offers analysis of codominant, haploid and binary genetic loci and DNA sequences. Both frequency-based ( <i>F</i> -statistics, heterozygosity, HWE, population assignment, relatedness) and distance-based (AMOVA, PCoA, Mantel tests, multivariate spatial autocorrelation) analyses are provided. In GenAlEx 6.5 introduce exciting new features including calculation of new estimators of population structure: <i>G'ST</i> , <i>G''ST</i> , Jost's <i>Dest</i> , and <i>F'ST</i> via AMOVA, Shannon Information analysis, linkage disequilibrium analysis for biallelic data, and heterogeneity tests for spatial autocorrelation analysis.	Calculation of <i>F</i> statistics and AMOVA, <i>PhiPT</i> values.	Peakall R and Smouse PE (2006); Peakall R and Smouse PE (2012)
	Phylip ver. 3.69	Windows, Mac OS X, UNIX	Freeware from <a href="http://evolution.genetics.washington.edu/phylip.html">http://evolution.genetics.washington.edu/phylip.html</a> .	Data types that can be handled include molecular sequences, gene frequencies, restriction sites, distance matrices, and 0/1 (binary) discrete characters. Methods that are available in the package include parsimony, distance matrix, and likelihood methods, including bootstrapping and consensus trees.		Felsenstein (2005)

<b>Sl. No.</b>	<b>Software Program me</b>	<b>Operating Platform</b>	<b>Source</b>	<b>Description</b>	<b>Implemented In</b>	<b>Reference</b>
	FigTree ver.1.3.1	Window, UNIX, Linux, Mac OS X	Freeware from <a href="http://tree.bio.ed.ac.uk/">http://tree.bio.ed.ac.uk/</a>	FigTree is designed as a graphical viewer of phylogenetic trees and as a program for producing publication-ready figures.	Construction of graphical viewer and publication ready phylogenetic trees	Rambaut and Drummond (2010)
	Geneious R7 ver 7.0.6 (Trial Version)	Windows, UNIX, Linux, Mac OS X	Trial Version: Freeware from <a href="http://www.geneious.com">http://www.geneious.com</a>	Sequence editing	Sequence editing, alignment	Rozen and Skaletsky (2000)
	BioEdit ver. 7.0.9	Windows, UNIX, Linux, Mac OS X	Freeware from <a href="http://www.mbio.ncsu.edu/BioEdit/bioedit.html">www.mbio.ncsu.edu/BioEdit/bioedit.html</a>	Sequence editing	Sequence editing, alignment	Hall (1999)
	MEGA ver 7.10	Windows, UNIX, Linux, Mac OS X	Freeware from <a href="http://www.megasoftware.net">http://www.megasoftware.net</a>	Sequence alignment and analyses	Sequence alignment, analyses of different indices and phylogenetic analyses.	Kumar et. al., (2016)
11	. DnaSP ver. 5	Windows, UNIX, Linux, Mac OS X	Freeware from <a href="http://www.ub.edu/dnasp">www.ub.edu/dnasp</a>	DNA polymorphism analyses from single locus and multiple loci. Measurement of sequence variation within and between populations. Perform Neutrality tests.	Sequence variation/ polymorphism analyses, haplotype, nucleotide diversity analyses. Neutrality test.	Librado and Rozas (2009)