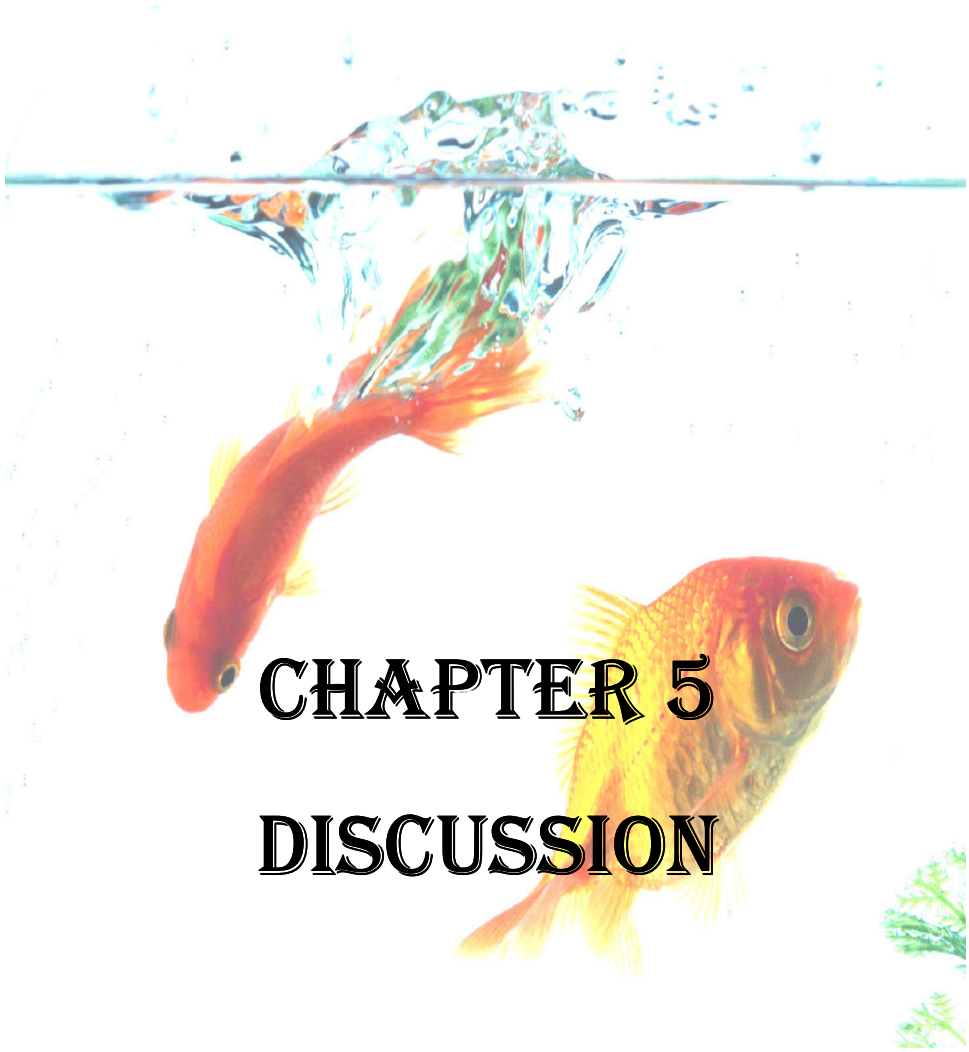

CHAPTER 5

DISCUSSION

CHAPTER 5
DISCUSSION

5.1 ORNAMENTAL FISH IN NORTHEAST INDIA

This study is undertaken to identify the ornamental fishes traded from NE India through DNA Barcoding approach, since, it is the harbor of many endemic and diverse fish species (Allen et al. 2010, Ponniah and Sarkar 2000). According to Ponniah and Sarkar, there were about 267 fish species being sold as either ornamental fish or as food from NE India. This figure increased in 2012, where Goswami et al. (2012) reported 422 fish species belonging to 133 genera, being traded for food as well as for aquarium pets. They showed out of total 1943 types of fishes (under 422 species), 1076 have high value for ornamenmtal relevance. It was also reported that highest number of ornamental fishes were found to be present in Manipur followed by Assam and Arunachal Pradesh.

In this study, the fish species being traded as ornamental fish for aquarium fish keeping was collected and vouchered based on available checklists (Bhattacharya and Choudhury 2004, Goswami et al. 2012, Jayalal and Ramachandran 2012, Mahapatra et al. 2004), from the different collection point of the major drainage systems in the Northeast and some ornamental fish traders from different regions.

5.2 DNA BARCODING OF ORNAMENTAL FISH

The effective monitoring of ornamental fish species traded is constrained by lack of accurate, quantitative and unbiased information (Murray et al. 2012, Raghavan et al. 2013) and shortage of taxonomic intervention. Since, ornamental fish trade is one of the large global industries (Tlusty et al. 2013a) hence; there is a vital need of inventorying the wealth of such bio-resources within the region. In the present study, the combined approach of *CO1* DNA barcoding and morphotaxonomy were employed to identify the collected 130 specimens into respective species. The meristic analysis of the ornamental fish provided

conclusive evidence to separate them as a distinct species. Our analyses indicate that number of caudal fin rays, lateral line scales, scales below lateral line; pre-anal length, etc. were the most prominent characters that can be used to distinguish the species. The *CO1* barcode region is already proven to be gold standard for discriminating animal species. Its short length can be sequenced quickly, easily and enough to distinguish species, due to its high sequence variability between the species as compared to within the species (Hebert et al. 2003b, Hebert et al. 2004c). The barcode gap which was calculated as the minimum interspecific distance readily delineates 130 specimens to 53 species. Among the certain cases, the similarity searches result, showed a range of 96-99% similarity with the database sequences viz. *Mastacembelus armatus, Channa striata, C. orientalis, Canthophrys gongota.* Besides, these sequences clustered as 2 or 3 cohesive units under a single node with respect to the species with which they showed similarity (Figure 4.3.2) and the K2P distance within, and between the clusters were slightly higher than the divergence that appeared for the other straightforward cases (Table 4.3.4). The sequences of *Channa orientalis* clustered separately from the database sequences forming two distinct clusters originating from the single node, the K2P distance between the two sub-clusters, although showed a higher divergence of 4.2%, but it is below the minimum interspecific K2P distance as barcode gap. Similarly, *Channa striata* clustered as a three cohesive unit under a unique node with a bootstrap support of 99%. The maximum K2P distance between the sub-cluster showed a higher divergence of 4.4%, but it is also below the barcode gap. This may arise due to geographic isolation of the population as there are few database sequences from Philippines (HQ654691), Tamil Nadu, southern part of India (EU342203, EU342204), etc. Such high conspecific divergence was also previously reported in *Channa striata* due to geographical isolation and substantial habitat re-organization (Jamsari et al. 2011), similar may be the case with *Mastacembelus armatus, Canthophrys gongota*. Thus, in congruence to the earlier studies, it was tentatively consider those cases as species having deep intraspecific divergence.

The minimum interspecific K2P distance as threshold for species delineation is put forwarded by many workers (Bhattacharjee et al. 2012, Meier et al. 2008) and is used in the current study to identify species that is high conspecific divergent so as to eliminate the possible misidentification induced. However, the misidentification can only be identified and eliminated effectively by the combined approach of morphology and distance based algorithms such as setting up threshold level for species delineation as well as clustering method. However a detailed study on the species diversity covering the entire taxa is further required for determining the presence of possible crypticism, overlooked species or possibility of misidentification. The barcode sequence compositional analysis revealed that there was a predisposition towards low 'G' content. The 2nd, 3rd codon position and the complete barcode region showed an AT bias. Among the different species, the amino acid composition showed highest frequency of Leucine followed by Alanine and Glycine. High Leucine content typically manage the correct assembly of collagen fibrils, regulate mineral deposition in bone, and modulate the activity of potent cellular growth factors through many signalling cascades (Park et al. 2008) which may be advantageous for muscle growth and its activity among fishes.

Among, all the above cases, some interesting instances were observed as already described in previous sections. In a few cases, although the specimens have different trade names, they represented same species. Many ornamental fish species, although having threatened status in IUCN, were marketed based on either multiple trade names or generic label of 'live ornamental fish' or 'live aquarium fish' or sometimes by the group label of 'Snakeheads', 'Barbs' rather than zoological nomenclature and thereby escapes regulations (Raghavan et al. 2013). The trade could still be regulated if each of the species were marketed with a single trade name with the corresponding zoological nomenclature. On the other hand, as DNA barcode is dependent on gene sequence of *CO1*, it will not change even if the species were misidentified or traded by multiple trade names. Hence, barcoding technology has an important role in regulating such case. Furthermore,

it was also evident throughout this study that, many threatened species were traded and for a few species, the IUCN and CAMP, revealed incongruous species status (Figure 4.3.4). As already discussed in result section, the species status of *Rhinomugil corsula*, *Pseudecheneis sulcata*, *Mystus vittatus* in IUCN is 'Least Concern' but the same is 'Vulnerable' as per CAMP, *Anabas testudineus,* the species status in IUCN is 'Data Deficient' but the same is 'Vulnerable' as per CAMP, similarly, the species status of *Puntius manipurensis, Glossogobius guiris* is 'Vulnerable' in IUCN, but the status of the same is 'Not Available' in CAMP. Therefore, species status of many fish appeared to be not accurately evaluated and needs revision. In this study, 53 species were identified from the traded samples. Among them, 14 species had multiple trade name, four species had single trade names. Importantly, 17 species belong to the threatened category and 31 species belong to lower risk near threatened/Least concern and the rest are deficient in data (Figure 5.1). Nevertheless, among the 14 species, four species having multiple trade names fall within the threatened category, and the species have been facing tremendous risks of jeopardizing due to the ongoing uncontrolled exploitation of resources through trade with them. This study focuses the imprudent exploitation scenario of high germplasm in nature. As anticipation, the survey highlighted the usefulness of DNA barcoding technique in monitoring the trade of threatened species and demand conservation strategy for sustaining wildlife.

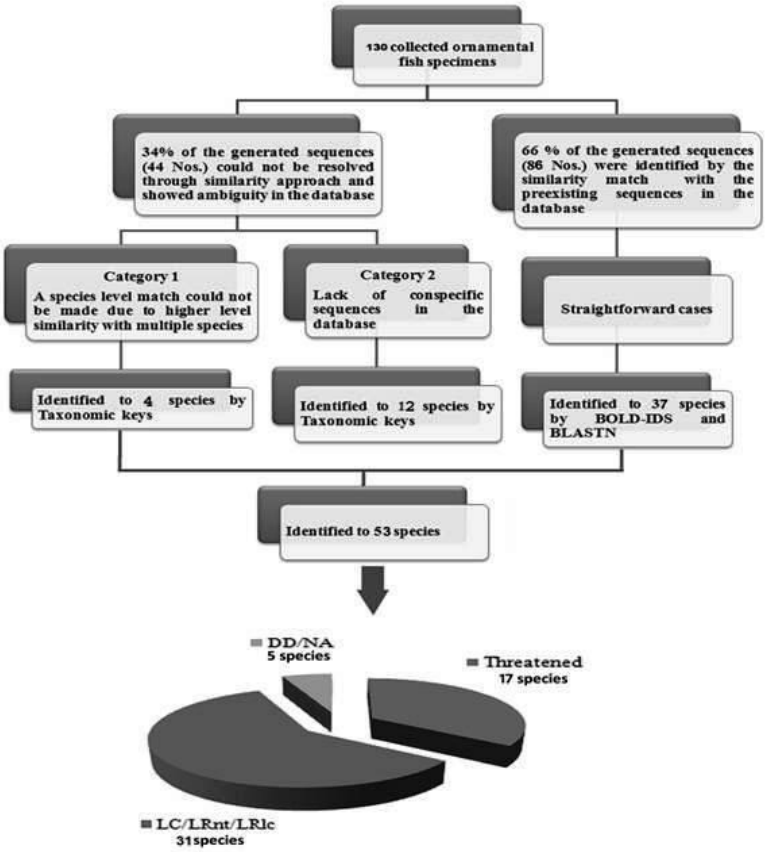

Figure 5.1: Species Status of Ornamental fish being traded from Northeast India. A total of 130 sequences were generated for the study out of which 66% of the cases were found to be straightforward to identify 37 respective species by similarity match approach. The remaining 34% ambiguous cases were solved by morphotaxonomy and found to represent 16 species. Among the total 53 identified, 17 species were in a threatened category.

5.3 DEVELOPMENT OF MINI-BARCODE FRAGMENT

Among the nucleotide substitutions, transition was found to be more dominant over transversion for entire barcode length which indicates less amino acid variation among the species as compared to nucleotide variation. Thus *CO1* nucleotide variation facilitates the identification of the taxa without altering the biological function CO1 protein. However, transversions in 3rd codon was found to be increased as compared to other sites. The comparison of interspecific K2P

distance with average interspecific transition and transversion in terms of correlation coefficient predicts the role involved separately of each type of nucleotide substitution (transition and transversion) in species differentiation. The comparison showed the genetic divergence between the species is positively correlated with both transitions and transversions, however, transversions showed slightly high positive correlation (Pearson correlation, $R = 0.95$, $p<0.001$). As observed in this study, transitions and transversions follow a distinct pattern of distribution in all the three orders of fishes and found the transversion biased sites are scattered randomly throughout the full-length barcode. Finally 171bp mini barcode (positions BP_{261} –BP_{432}) was proposed within the whole length barcode for ornamental targeting a variable concentric site rich with transversional substitution (Figure 4.4.9), as it has major effect in bringing divergence between the species. The identification of high informative segment within the full-length barcode is important to select a subset of mini barcode. To minimize the barcode segment, the number of selected features has to be minimized, under some constraints, while ensuring that the genetic information content of the segment is high enough to delimit species. Furthermore, the length of the segment should be sufficient to generate requisite number of barcodes to describe all the extant species of the studied group. Here in this study, transversion rich sites were used for generating a short species-specific barcode motif as transversion mutations have been known to play a dominant role over transitions in the evolution of species and thus are rich in information content (Mitrofanov et al. 2002). This further validates that the proposed mini barcode region is competent with the full-length barcode in delimiting species. This 171bp segment is found to have a better coverage of the transversion dominant segment for the different fish groups similar to the 119bp segment for catfish, as proposed earlier, Bhattacherjee et al. (Bhattacharjee and Ghosh 2013)

The NJ trees constructed from the full length and the mini barcode region (171bp) have shown a comparable clustering pattern of the 53 species (Figure 4.4.10). Some sequences, that show intraspecific divergence from the expected pattern of clustering in the full-length barcode NJ tree, also showed similar trend

in the mini barcode tree. These included instances of either, mislabeled sequences in the database, presence of haplotype diversity or cases of crypticism, that have been discussed in previous studies (Benziger et al. 2011, Bhattacharjee et al. 2012). Similarly, the K2P distances in both the cases (full length vs mini-barcode) showed similar trend of intra-specific as well as inter-specific divergences. Furthermore, successful PCR amplification with the designed common primer for the three orders of fishes (from archived and field based collected sample) confirmed specific amplification of the targeted mini-barcode fragment as already described (Figure 4.4.13). Sequencing of the 221bp amplicons with the common primers further validated the presence of mini-barcode and the recovered sequence data from this region showed congruency with the *in-silico* result which well characterized the fish species. Thus, it can be said that the mini-barcode approach is more effective than conventional DNA barcoding with an added advantage of being short in size and have better coverage of the transversion rich domain thereby making it more preferable for rapid field-based species identification as well as from archived reference sample, to confirm the species identification.

5.4 *CO1* BASED PHYLOGENY

CO1 DNA barcode sequences proved to be very effective in species differentiation and many studies had proposed that the *CO1* sequences also carries phylogenetic signal (Benziger et al. 2011, Hajibabaei et al. 2007b, Kerr et al. 2009). In this study as it was observed that all the species had clustered distinctly and originated from a common ancestral node and most of the species within a family had clustered within a single node in ML phylogeny. Therefore, it revealed its efficacy for higher-level (family level) phylogenetic analysis. The families of the Percifirmes clustered distinctly with respect to other families originating from a common ancestor but *Glossogobius guiris* (family Gobiidae) under the order Perciformes clustered out and placed itself near the members of the family Sisoridae of the order Siluriformes indicating it to be paraphyletic or its evolution to be convergent. Such discrepancies are also described previously (Dawson et al.

2002, Ellingson et al. 2014). Hence, the family Gobiidae needs to be assessed more carefully.

The families of Siluriformes showed distinct clusters (monophyletic origin), in some cases it indicated the presence of paraphyly. The family Siluridae and Erethistidae of the order Siluriformes revealed as the monophyletic families in this study. While the other members of the same order viz. *Mystus vittatus* and *M. gulio,* within the family Bagridae is consistent with the morphological classification. Similar observation was with the *Ompok bimaculatus* and *O. pabda* of the family Siluridae, while *Clarias batrachus* clustered separately as a monophyletic group. The family Schilbeidae reflected as a true family since the species *Eutropiichthys murius* and *Ailia coila* had clustered within a node.

Sisoridae revealed that the species *Sisor rhabdophorus, Pseudechineis sulcata* had clustered from a common ancestral node. On the other hand the genus *Gagata* which is claimed to be under Sisoridae revealed here as distinct monophyletic unit separately clustered from the other sisorid members, thus claiming this family to be paraphyletic in origin. However, detailed study of this family is further required.

The families of the Cyprinoformes (Cyprinidae, Psilorhynchidae and Cobitidae) in the present study clustered distinctly with respect to other families originating from a common ancestor. The members of the family Cyprinidae viz. *Labeo bata, L. dero, Puntius sophore; P. chola, Gara arupi, Barilius bendelisis, Amblypharingodon mola, etc.* clustered distinctly with respect to other species of the Cyprinidae, originating from a common node. Thus, cyprinidae is confirmed to be a true family. Similar trend was observed for family Psilorhynchidae and Cobitidae. However, *Botia dario* and *B. rostrata* did not clustered with the other species of the family Cobitidae, rather it clustered with the members of the Cyprinidae as a monophyletic unit. This indicated that the family Cobitidae as a paraphyletic in origin.

Evolution time for all theses taxa were determined by molecular clock under General Time Reversible model. In this study, the phylogenetic tree was

calibrated with two alternative constraints, one based on the oldest known fossil of channids (from middle Eocene period) and the other based on the available molecular divergence time estimate for the emergence of the genus Channa based on previous records which is 84-112 MYA (Benziger et al. 2011). This divergence time value was attained based on the continental breakup of African and South American landmasses (100–120 MYA) and the estimated divergence time between Sarcopterygians and Actinopterygians (420–500 MYA), which has been successfully used previously to date old divergence times in actinopterygian fishes (Kumazawa and Nishida 2000). The fossil records (oldest known channid fossil) from Northwest Pakistan had faunal affinities towards both Asia and Africa, which might be due to the touch with the drifting Indian subcontinent, with Africa, during its northward movement allowing the dispersion of African fauna into Asia (Murray and Thewissen 2008, West 1980).

In this study, the divergence of Cypriniformes from the Perciformes and Siluriformes from the ancestral node was found to be started during late Triassic period of the Mesozoic era upto early Cenozoic era (130.04 – 12.31 MYA). On the other hand Perciformes emerged from the other orders during early Jurassic Period. The different members of Perciformes viz. Channa, Parachanna, Nandus, etc. under the current study emerged during the late Cretaceous to Paleogene period. While, the different members of the order Siluriformes like Pseudecheneis, Gagata, Mystus, etc was found to be emerged in between late Cretaceous to Neogene period date back to 105 – 13.44 MYA. The continents stayed united as the supercontinent Pangea. However, in the middle Jurassic (157–178 MYA), Pangea splitted into Laurasia and Gondwanaland, which were further fragmented into smaller landmasses like Eurasia, North America, and Greenland from the former, and Africa, South America, Australia, Antarctica, Madagascar, and India from the latter (Kumazawa and Nishida 2000). These tectonic plate movements not only shaped our present but also created cosmopolitan distribution of the Cypriniformes. In this study, it was found that mean divergence between *Botia dario* and *B. rostrata, Labeo bata* and *L. dero*, *L. gonius* and *L. clabasu,* and the divergence of *Devario devario* and *D. aequipinnatus* from a common ancestral

node during Miocene epoch (17.81, 15.94, 12.31 and 14.18 MYA respectively). The Continental drift was continued to present position; many mountain range formations took place in America, Europe and Asia (Kumazawa and Nishida 2000). This might have dispersed the said genera throughout Southern part of Asia including India and other drainage of Southeast Asia. Whereas, the emergence of the genus Barilius, Cirrhinus, Puntius, Crossocheilus occured during the Paleogene period (33-66.3 MYA). In this period many lakes and fresh water bodies formed along with the distribution of many fresh water fishes. There are some fossil records which can be considered in evaluating the migration history of the Asian freshwater fish. The Indian subcontinent became connected to Eurasia by the late Early Eocene. Thus, the migration of many diverse groups of fish could have come to India through the connection of terrestrial freshwater habitats or vice-versa (Metcalfe 2013).

The mean divergence time to common ancestor (TMRCA) between the members of Channa and another genus Parambassis ranges between early Eocene to Miocene epoch of Cenozoic era (26.84 - 58.76 MYA), While, the members of Anabus, Nandus, Rhinomugil emerged between late Cretaceous period. As projected in the Satpura hypothesis the westward relocation of Malayan fishes deflected southwards in the late Miocene due to the formation of a ridge in the North of the Himalayan range. But, this may be likely for other torrential freshwater fishes, but the dispersion of the congeners of the Channa, all the way through this route would be complicated to explain. This is because in the early Miocene there was absence of any geographic connections towards south; rather it was hypothesize the divergence of Channa genera during the Gondwana land breakup, with the genus Channa scattering into Eurasia (Benziger et al. 2011).