Chapter 6. SUMMARY

Freshwater fishes in India are poorly known and plagued by unresolved cryptic species complex that masks many latent and endemic species. Limitations in traditional taxonomy have resulted in this crypticism. Therefore, for legible characterization of Indian freshwater fishes, there is an urgent need of species scrutiny using advanced molecular methods. Among the molecular methods, DNA barcoding has proven to be an effective tool in species identification and resolution of various taxonomic impediments. However, its efficacy in assigning species to their higher taxonomic ranks viz. genus, family, order is still dubious. Many methods of higher taxon assignment using DNA barcodes have been proposed but, none could be standardized. Further, the method used for interpreting the barcode also plays an important role in determining the efficiency of the barcoding gene. The barcodes are generally interpreted using distance based and character based methods. The latter is more suitable for field based application of DNA barcoding. However character based approach have some inherent problems like sequencing of full length barcode leading to use of sparse data matrix and lack of a uniform diagnostic position for each studied group. In this study, two important issues are addressed, first comprehensive assessment of DNA barcoding of Indian freshwater fishes and second improvisation in current methods of interpreting barcodes.

In this study, the entire dataset of both recorded and barcoded Indian freshwater fishes, so far available, were surveyed and valid barcode sequences retrieved. Among them, some DNA barcodes of freshwater species from the eastern part of India were developed. Altogether 1383 mitochondrial *COI* barcode sequences of 175 species of Indian freshwater fishes were analyzed. An attempt was made to resolve the taxonomic impediments inherent in the study of freshwater fishes of India using DNA barcoding method. Overall, 87% of the available Indian freshwater fish barcodes were diagnosed as true species in parity with the existing checklist which can act as reference barcode for the particular taxa. While, for the remaining 13% (21species) the correct species name was difficult to assign as they depicted some erroneous identification and cryptic species complex.

A detailed compositional analysis of the barcode region was carried out and sequence information in *COI* was calculated using R_{seq} value. The motive was to elucidate, whether nucleotide and amino acid sequence in the *COI* gene carried sufficient information to assign species to their higher taxonomic rank. Our results revealed that with increase in the taxonomic rank, sequence conservation decreased for both nucleotides and amino acids. Order level exhibited lowest conservation with 50% of the nucleotides and amino acids being conserved. However, there was sufficient information content at each taxon level to facilitate identification at that level.

In an attempt to utilize the information content latent in *COI* barcode, the existing methods of "reading" barcodes was modified by concentrating on a short continuous information rich segment. A 154 base pair fragment, from the transversion rich domain of *COI* barcode sequences, was found to successfully delimit species in the three most diverse orders of Indian freshwater fishes. Species-specific barcode motifs for 109 species were designed using the character based method of DNA barcoding. Through an in-house pattern-matching program, the motifs were shown to successfully identify the correct species and did not show any false positive assignments. Further, the motifs also correctly identified geographically isolated populations of conspecific sequences of the order Cypriniformes. Thus it is anticipated, that the use of such motifs will enhance the diagnostic power of DNA barcode and will greatly benefit the field-based system of rapid species identification.

Overall, the study covered a comprehensive assessment of DNA barcodes of Indian freshwater fishes. It re-evaluated and elucidated the actual species status of the majority of the studied species and particularly helped to flag the species whose statuses have been doubtful. The study also revealed the potential of *COI* gene in higher taxon assignment with ample inherent signal latent in the gene. Moreover, the proposed novel method of "barcode-motif" based identification promises to enhance the in-field applicability of DNA barcode.