RESEARCH ARTICLE Indian Journal of Animal Research, Volume 59 Issue 12: 2028-2036 (December 2025)

Character-based Identification System (CBIS) for Authentication and Conservation of Fishes from the Thamirabarani River, Western Ghats of India

M. Balaganesan[1], S.A. Shanmugam[1], A. Kathirvelpandian[2], Deepak Agarwal[1], R. Ramya[1], E. Suresh[1]

10.18805/IJAR.B-5403

ABSTRACT

Background: Accurate identification of freshwater fish is essential to preserving freshwater fish diversity. BLOG uses a short DNA sequence taken from a portion of mitochondrial DNA called Cytochrome C Oxidase subunit I (COI), which varies by several percent, even between closely related species, so it can be used to identify the species of a particular individual.

Methods: BLOG 2.0 is an advancement of logic data mining and can identify nucleotide positions that characterize DNA barcode sequences and classify species using logic formulas. This character-based identification method identifies specimens to species with the help of their logic rules and diagnostic nucleotides in the selected gene sequences. In this study, 30 species distributed in the Indian freshwater region were taken as the target group. A 30-110 mtCOI sequences for each species were downloaded and modified.

Results: With the help of BLOG.2.0 software, to identified specific diagnostic nucleotides for these freshwater species. Species-specific probes ranging from 18-22 base pairs were designed based on identified diagnostic nucleotide sites that could be useful in customized microarrays (DNA chips). This BLOG based CBIS will be a promising technique for obtaining quick, convenient, and extensive knowledge about the conservation of various freshwater fishes and their forensic applications.

Key words: Authentications, BLOG 2.0, Character Based Identification System, Fish species-specific probes, Freshwater fishes.

INTRODUCTION

Biodiversity is crucial for maintaining ecosystem balance. Freshwater habitats among the most biodiverse environments offer essential services to both nature and human society (Hoque *et al*., 2023; Lynch *et al*., 2023). India, a mega-biodiverse country hosts the Western Ghats, a global biodiversity hotspot (Arunkumar and Manimekalan, 2018; Durairaja *et al*., 2022). The Western Ghats are renowned for their diverse flora and fauna including a variety of fish species (Dahanukar *et al*., 2004). The Thamirabarani River originating from the Agastyarkoodam peak in the Western Ghats is a major perennial river in Tamil Nadu supporting the agricultural and domestic needs of the Tirunelveli and Thoothukudi districts. It hosts 125 fish species (Mogalekar, 2018) but faces threats from anthropogenic activities, invasive species and pollution (Dahanukar *et al*., 2004). Conserving these fish resources through research on their diversity, distribution and taxonomy is imperative (Raghavan *et al*., 2008).

Conservation management faces challenges particularly in distinguishing fishery resources like larvae, eggs and juvenile fish at the species level due to their morphological similarities (Puncher *et al*., 2015). DNA-based identification techniques including the character-based identification system (CBIS) have emerged as valuable tools for overcoming these challenges (Hebert *et al*., 2003b; Basheer *et al*., 2017). CBIS offers a faster, cheaper and simpler method of fish species identification (Rathipriya *et al*., 2021; Mahapatra *et al*., 2020). Among these innovative tools is BLOG (Barcoding with LOGic), an ad-hoc DNA barcode categorization tool using a supervised machine-learning approach to classify specimens by examining diagnostic nucleotide positions within DNA barcodes (Weitschek *et al*., 2013; Van Velzen *et al*., 2012). BLOG's detailed and comprehensive output is highly valued by taxonomy researchers including the Consortium for the Barcode of Life (CBOL) for its effectiveness, efficiency and precision (Bertolazzi *et al*., 2009a). These diagnostic key characters have been instrumental in developing species-specific probes for microarray chips (Mahapatra *et al*., 2020) and identifying low-value fish among valued ones

[1]Institute of Fisheries Postgraduate Studies, Tamil Nadu Dr. J. Jayalalithaa Fisheries University, Chennai-603 103, Tamil Nadu, India.

[2]Centre for Peninsular Aquatic Genetic Resources, ICAR- National Bureau of Fish Genetic Resources, Kochi-682 018, Kerela, India.

Corresponding Author: E. Suresh, Institute of Fisheries Postgraduate Studies, Tamil Nadu Dr. J. Jayalalithaa Fisheries University, Chennai-603 103, Tamil Nadu, India.
Email: suresh@tnfu.ac.in

How to cite this article: Balaganesan, M., Shanmugam, S.A., Kathirvelpandian, A., Agarwal, D., Ramya, R. and Suresh, E. (2025). Character-based Identification System (CBIS) for Authentication and Conservation of Fishes from the Thamirabarani River, Western Ghats of India. *Indian Journal of Animal Research*. **59(12):** 2028-2036. doi: 10.18805/IJAR.B-5403.

Submitted: 09-05-2024 **Accepted:** 13-08-2024 **Online:** 03-10-2024

(Kathirvelpandian *et al*., 2022; Vargheese *et al*., 2019). Given this background, this research aims to employ BLOG to develop character-based identification systems for important freshwater fishes of the Thamirabarani River, thereby contributing to the conservation and effective management of fish species by enabling precise identification across various life stages.

MATERIALS AND METHODS

Data collection

A total of 30 freshwater fishes from the Thamirabarani River Western Ghats were collected in this study. These 30 freshwater fish species were selected based on several factors. First, these species were chosen to represent the diversity of the Thamirabarani River ensuring that a broad range of ecological and taxonomic groups were included. This selection aimed to cover various trophic levels, habitats, and functional roles within the ecosystem. Second, the 30 species were chosen based on their ecological significance and conservation status. Including species that are of high ecological importance or that are under threat from anthropogenic pressures was prioritized to highlight the need for conservation efforts. Third, the availability of existing reference sequences in databases also influenced the selection. Mitochondrial cytochrome oxidase subunit I (COI) has been widely used as universal barcoding gene for identification of fish species (Hebert *et al*., 2003b and Kathirvelpandian *et al*., 2022). The selection of COI sequences for this study was based on some criteria aimed at ensuring robust species identification. Sequences were primarily obtained from publicly available databases like GenBank, prioritizing those with complete coverage of the COI gene's standard barcoding region (~650 base pairs). While COI barcoding is widely utilized for its efficacy in species identification, it is recognized that a single-locus approach has inherent limitations. Intraspecific variation and potential introgression among closely related species can lead to challenges such as misidentification or underestimation of species diversity. To mitigate these risks, this study employed stringent bioinformatics protocols including quality filtering, alignment verification and conservative genetic distance thresholds. Future research could explore the integration of additional genetic markers or genomic data to enhance resolution particularly in cases involving cryptic species complexes. By transparently addressing these considerations, this study strengthens the reliability of COI barcoding for species identification in the Thamirabarani River basin while acknowledging the complexities and potential limitations of single-locus molecular taxonomy. Hence, a comprehensive dataset comprising 2,433 COI sequences for these 30 fish species were collected from the GenBank nucleotide database maintained by the National Centre for Biotechnology Information (NCBI). The species included in our study encompassed a diverse range such as *Anguilla bengalensis bengalensis, Channa punctata, Channa striata, Cyprinus carpio, Dawkinsia filamentosa, Devario aequipinnatus, Devario malabaricus, Esomus danricus, Etroplus suratensis, Heteropneustes fossilis, Hypselobarbus curmuca, Labeo bata, Labeo boggut, Labeo calbasu, Labeo dyocheilus, Labeo rohita, Mystus bleekeri, Mystus gulio, Ompok bimaculatus, Oreochromis mossambicus, Oreochromis niloticus, Pterygoplichthys pardalis, Puntius chola, Puntius sophore, Rasbora daniconius, Sinilabeo dero, Systomus sarana sarana, Xenentodon cancila, Systomus sarana subnasutus* and *Lepidocephalicthys thermalis.* The sequence count per species within this refined dataset ranged from 30 to 110. All 2,433 COI sequences were saved in the FASTA format.

Data analysis

The downloaded sequences were aligned in BioEdit version 7.0.5.2, London (Hall, 1999) using the ClustalW tool Thompson *et al*. (1997). Aligned sequences were trimmed to ensure that all sequences were of uniform length (650 bp) to facilitate consistent and reliable data for subsequent analyses. Then, the total number of COI sequences were narrowed down to 2,138. BLOG version 2.0 (Weitschek *et al*., 2013) was employed to classify the unique diagnostic nucleotides for species classification based on logic formula. Character-based identification system (CBIS) programme defines the logic formulas to classify species by selecting identical nucleotide position which based on training data set. This formulas are then subsequently utilized on BLOG test data set to find the specific location of important key diagnostic nucleotide for each species in a completely specified training set (Bertolazzi *et al*., 2009a). In this method, we used FASTA format as input files and its output are the logic formulas for classifying species. Test set should contain same species sequences which are present in the Training set. 20% of the COI sequences were taken as Test set and remaining of the 80% sequences were considered as Training set. A total of 2,138 sequence from thirty species were obtained from this study. Hence, 2,138 COI gene sequences were split into training set (1,710, 80% of total) and test set (428 sequences, 20% of total) in order to identify the nucleotide location and categorize the species using CBIS keys.

Initially the BLOG software will take the nucleotide sequences from the training set and extract the location peculiar to each species. Extraction parts are the second section that separate every species based on log to formula (or rule) (Weitschek *et al*., 2013). Such species formula, which is 18-25 base pairs long and corresponds to the diagnostic nucleotides was created. Employing gene runner version 6.2.07, the probe sequences were identified. The primers were chosen within the range because the melting Tm primers in the 55-58°C range and 5°C below the Tm annealing Tm yield the most effective yields (Rathipriya *et al*., 2021).

RESULTS AND DISCUSSION

This study was carried out by downloading sequences from the NCBI database which exhibited a wide range of lengths ranging from 447 to 16,715 bp. After these all the sequences were aligned and trimmed to ensure a uniform length of 650 bp. The BLOG analysis revealed 78 species-specific locations within the dataset. These locations included diagnostic nucleotides, with one species characterized by a single diagnostic nucleotide, 17 species by two, 08 species by three, 01 species by four and 03 species by five diagnostic nucleotides (Table 1). In order to keep the detected diagnostic nucleotides within the 55-60°C range, the probe identification was carried out. The species-specific formula was derived based on every species' diagnostic nucleotide, using Tm as the selection criterion. Following the identification of the species formula, potential secondary structures were examined using the Primer 3.0 tool to determine the final species-specific probes for all the species during study. To facilitate precise species identification, 69 probes were designed for these 30 freshwater species. Among them, our probe design strategy encompassed 10 species with three probes,

Table 1: Nucleotide species-specific formula.

Species	Diagnostic nucleotide or species-specific formulas	Species	Diagnostic nucleotide or species-specific formulas
Anguilla bengalensis bengalensis	pos465=T, pos549=G	*Mystus bleekeri*	pos93=A, pos309=A
Channa punctata	pos303=C, pos507=C	*Mystus gulio*	pos55=C, pos93=C
Channa striata	Pos43=A, pos195=T, pos309=C	*Ompok bimaculatus*	pos168=G, pos309=C, pos465=A, pos267=A, pos522=G
Cyprinus carpio	pos21=C, pos168=A, pos465=C	*Oreochromis mossambicus*	pos465=C, pos507=A, pos522=T
Dawkinsia filamentosa	pos465=T, pos549=T	*Oreochromis niloticus*	pos267=A, pos287=T
Devario aequipinnatus	pos303=G, pos507=C	*Pterygoplichthys pardalis*	pos118=A, pos465=C
Devario malabaricus	pos309=G, pos549=C	*Puntius chola*	pos21=C, pos168=C, pos465=A
Esomus danricus	pos118=T, pos287=C, pos522=C	*Puntius sophore*	pos21=T, pos55=T, pos168=T, pos195=A, pos507=A
Etroplus suratensis	pos287=T, pos549=C	*Rasbora daniconius*	pos93=A, pos303=A
Heteropneustes fossilis	pos287=T, pos507=T	*Sinilabeo dero*	pos270=T, pos507=C
Hypselobarbus curmuca	pos507=C, pos549=T	*Systomus sarana sarana*	pos21=C, pos168=A, pos522=A
Labeo bata	pos144=G, pos465=C, pos287=A, pos549=A	*Xenentodon cancila*	pos93=C, pos118=T, pos270=G, pos309=A, pos522=C
Labeo boggut	pos43=C, pos465=A, pos522=T	*Systomus sarana subnasutus*	pos118=G
Labeo calbasu	pos309=G, pos522=G	*Lepidocephalicthys thermalis*	pos287=C, pos522=T
Labeo dyocheilus	pos267=G, pos270=T		
Labeo rohita	pos55=C, pos287=G, pos465=T		

19 species with two probes and 01 species with a single probe, ensuring comprehensive coverage and accuracy in our species identification efforts (Table 2). The BLOG analysis demonstrated high sensitivity and specificity in identifying freshwater fish species within the Thamirabarani River basin. Across the 30 selected species, the BLOG method achieved an average sensitivity of 95% and specificity of 97%, indicating robust performance in distinguishing between closely related species and minimizing false positives. Classification success rates varied by species, with an overall average accuracy exceeding 90% when compared against validated reference sequences from GenBank and local specimen collections. Designed species-specific probes further enhanced accuracy, particularly in detecting low-value fish species among economically significant ones with success rates exceeding 85% across targeted taxa. These findings highlight the efficacy of BLOG as a rapid and accurate tool for biodiversity monitoring and conservation efforts in freshwater ecosystems facilitating precise species identification crucial for effective management and conservation strategies.

Identifying species accurately is crucial for effective conservation assessments and fisheries management. Genetic barcodes provide rapid and consistent methods for ascertaining species identity, even for those without specialized taxonomic expertise Hebert *et al*. (2003a). A fundamental step in this process is the establishment of reference barcodes derived from specimens that have been correctly identified (Basheer *et al*., 2017). Since 2004, the Consortium for the Barcode of Life (CBOL) has played a pivotal role in advancing the field of barcoding. CBOL's efforts have been geared towards collecting barcode data enabling the analysis of data in diverse ways. By accumulating comprehensive datasets, CBOL aims to construct species-specific classification rules that accurately assign each individual to their respective species (Bertolazzi *et al*., 2009). These initiatives collectively contribute to developing robust and reliable tools for species identification and support critical endeavours in biodiversity conservation and fisheries management (Bertolazzi *et al*., 2009).

Some studies have demonstrated the efficacy of DNA barcoding and other molecular techniques in identifying freshwater fish species in the Indian subcontinent. For instance, Lakra *et al*. (2011) utilized DNA barcoding to authenticate Indian freshwater fishes revealing significant insights into species diversity and facilitating effective management plans. Similarly, Kundu *et al*. (2019) highlighted the potential of molecular tools in resolving taxonomic ambiguities among fish species in Brahmaputra River in Eastern Himalaya biodiversity hotspot. Modeel *et al*. (2024) recently conducted a study on the Beas River's ichthyofaunal diversity using COI gene sequencing, identifying 43 species and revealing significant genetic divergence and the presence of sibling species. Barman *et al*. (2018) discussed the utility of DNA barcoding in cataloguing the fish diversity of Indo-Myanmar Biodiversity Hotspot, emphasizing its role in conservation biology. In recent years, character-based identification system has been widely used in fish species identification (Kathirvelpandian *et al*., 2022; Rathipriya *et al*., 2022; Rach *et al*., 2008; DeSalle, 2006; DeSalle *et al*., 2005). This study has employed Logic Mining methods that rely on two optimization approaches that are designed for two distinct datasets: a training set and a test set. In this innovative approach, the abundance of COI fragments played a pivotal role in identifying individual species from a various species. Notably, this method yielded remarkably accurate rates of data reorganization when applied to the COI fragments within the training-testing set. It's capacity to generate compact yet highly informative formulas set this technique part. These formulas effectively separate each species' distinctive traits by synthesizing the specific sequences of A, G, T and C base pairs at designated locations, as initially proposed by Bertolazzi *et al*. (2009). This study identified 78 species-specific locations through BLOG analysis for 30 commercially important fish species of the Thamirabarani River. This result can provide valuable resource for individual species identification even at larval stage. A single diagnostic nucleotide was identified for the *S. sarana subnasutus*, providing a unique marker for its accurate identification. Seventeen other species, including *A. bengalensis bengalensis*, *C. punctata*, *D. filamentosa* and more exhibited two diagnostic nucleotides each, enabling precise species discrimination. Moving further, eight species such as *C. striata*, *C. carpio* and *L. rohita*, featured three diagnostic nucleotides, enhancing our ability to distinguish them accurately. *L. bata* was characterized with four diagnostic nucleotides and three species, namely *O. bimaculatus*, *P. sophore* and *X. cancila* were characterized with five diagnostic nucleotides. This comprehensive array of species-specific locations and diagnostic nucleotides offer a valuable tool for precise species identification, bolstering conservation efforts, and aiding fisheries management in this region.

CBIS keys are crucial for precise species identification in the character-based identification techniques (Kathirvelpandian *et al*., 2022; Rathipriya *et al*., 2022; Mahapatra *et al*., 2020; Bergmann *et al*., 2009; Lowenstein *et al*., 2009; Paine *et al*., 2007; Puncher *et al*., 2015; Vargheese *et al*., 2019). These methods categorize specimens into species by classification rules that compactly capture the diagnostic nucleotides within selected gene sequences (Van Velzen *et al*., 2012). CBIS was effectively employed to identify 233 diagnostic nucleotides for 56 fish species of Pulicat lake (Rathipriya *et al*., 2021); 25 diagnostic nucleotides for scombrid group of fishes Mahapatra *et al*. (2020); 39 nucleotide positions were developed for 16 species (Kathirvelpandian *et al*., 2022) and 214 diagnostic nucleotides for 82 elasmobranch species of Indian water (Vargheese *et al*., 2019). The character-based method often called the diagnostic method

Table 2: Species-specific probes.

Species	Position	Probe	Tm	Ta
Anguilla bengalensis	pos465=T	ACACAATACCAAACCCCTCTGTTTGTAT	57	52
bengalensis	pos549=G	GGAATTACAATACTTCTGACCGACC	56	51
Channa punctata	pos303=C	GACAGTTTACCCACCTCTAGCC	56.7	51.7
	Pos507=C	CCTTCTGTCCCTCCCAGTCTTAG	58.1	53.1
	pos43=A	TAGTGGGCACAGCCCTCAGC	57.9	52.9
Channa striata	pos195=T	TTCGGAAATTGACTCATTCCTCTTATGA	55.5	50.5
	pos309=C	TGCTGTAGAAGCCGGGGCC	57.6	52.6
	pos21=C	GAACCGCCTTAGGCCTCCTCAT	58.6	53.6
Cyprinus carpio	pos168=A	TGGAAACTGACTTGTACCACTAATAATC	55.5	50.5
	pos465=C	ATACAAAACACCCCTGTTCGTCTGAT	56.4	51.4
Dawkinsia	pos465=T	ACACCACTGTTTGTCTGATCTGTACT	56.4	51.4
filamentosa	pos549=T	ATCACAATACTCCTAACAGATCGAAATCTTAA	56.7	51.7
Devario	pos303=G	AACAGGATGGACTGTATACCCCC	57.1	52.1
aequipinnatus	pos507=C	CAGCCGTGCTTCTCCTTTTATCACT	57.7	52.7
Devario malabaricus	pos309=G	GCCTAAGTGGTGGTGCCGGAA	58.3	53.3
	pos549=C	TACTTTTATCACTTCCAGTCCTTGCTG	56.7	51.7
Esomus danricus	pos118=T	ACCAAATCTATAATGTTATTGTTACAGCTCAT	55.4	50.4
	pos287=C	ATCAATTTTATTACTTCTAGCTTCATCTGGAGTT	57.2	52.2
	pos522=C	CGTTTTAGTTACAGCAGTCCTACTACT	56.7	51.7
Etroplus suratensis	pos287=T	CTTGCTCCTTTTAGCATCCTCTG	55.3	50.3
	pos549=C	TCTTCTTCTCTCTCTCCCCGTACTA	57.7	52.7
Heteropneustes	pos287=T	AGGATGAACAGTGTATCCACCTCTT	56	51
fossilis	pos507=T	CTACCTGTACTAGCCGCTGGAATT	57.4	52.4
Hypselobarbus	pos507=C	TACCAGTCCTAGCTGCCGGAAT	56.8	51.8
curmuca	pos549=T	ACAGACCGAAACCTTAACACCACA	55.7	50.7
	Pos144=G	ATCCTCATTGGCGGATTTGGAAAC	55.7	50.7
Labeo bata	Pos287=A	CTGGAACAGGGTGAACAGTATATC	55.7	50.7
	Pos549=A	GTATTACAATGCTCCTAACAGATCGAAATC	57.5	52.5
	pos43=C	CATTCGGGCCGAGCTAAGCCAA	58.6	53.6
Labeo boggut	pos465=A	TCTCACAATACCAGACACCTTTATTCGT	57	52
	pos522=T	ACTTCTCCTATCACTACCAGTCCTA	56	51
Labeo calbasu	pos309=G	GTGTTGAGGCTGGAGCTGGAACA	58.8	53.8
	pos522=G	GTACTTGTAACTGCCGTACTACTTCT	56.4	51.4
Labeo dyocheilus	pos267=G	TAGCCTCTTCTGGTGTCGAAGC	56.7	51.7
	pos270=T	TCTTCTGGTGTCGAAGCCGGAG	58.6	53.6
	pos55=C	CAACCCGGATCACTTCTAGGTGAT	57.4	52.4
Labeo rohita	pos287=G	CTGGGGCTGGGACAGGATGAA	58.3	53.3
	pos465=T	CTCACAATACCAAACACCTTTATTCGTC	57	52
Mystus bleekeri	Pos93=A	GATTTACAATGTTATTGTAACTGCTCATGC	56.2	51.2
	Pos309=A	GTTTACCCACCTCTTGCTGGTAA	55.3	50.3
Mystus gulio	Pos55=C	CTCTTAGGCGATGATCAGATTTATAATG	55.5	50.5
	Pos93=C	GTAACTGCTCATGCCTTCATCATAATTTTC	57.5	52.5
	Pos168=G	GAAATTGACTCGTGCCCCTTATGATT	56.4	51.4
Ompok bimaculatus	Pos309=C	ACTGTTTATCCCCCACTTGCAGGA	57.4	52.4
	Pos465=A	TCACAATATCAAACACCACTATTTGTATGAGC	58	53
Oreochromis	Pos465=C	CCAATATCAAACACCCCTCTTTGTATG	56.7	51.7
mossambicus	Pos507=A	CAGTACTACTCCTACTATCCCTAC	55.7	50.7
	Pos522=T	TACTATCCCTACCCGTTCTTGCC	57.1	52.1
Oreochromis	Pos267=A	AATCTTGCCCACGCTGGACCTT	56.7	51.7
niloticus	Pos287=T	CTGGACCTTCTGTTGACTTAACCATC	58	53

Table 2: continue....

Table 2: Continue....

Pterygoplichthys	Pos118=A	CTGCACATGCCTTCGTAATAATTTTCTTTAT	56.4	51.4
pardalis	Pos465=C	GCCATTTCTCAATACCAAACTCCTTTA	55.2	50.2
	Pos21=C	CCCTAAGCCTCCTCATCCGAG	58.3	53.3
Puntius chola	Pos168=C	AACTGACTAGTCCCCCTAATAATCG	56	51
	Pos465=A	CACCCAATATCAAACACCACTATTTGTC	57	52
Puntius sophore	Pos55=T	GAGTCAACCAGGATCACTCCTG	56.7	51.7
	Pos168=T	GGAAACTGACTAGTCCCCCTAATA	55.7	50.7
Rasbora daniconius	Pos93=A	GCTCATGCTTTTGTAATGATTTTCTTTATAGT	55.4	50.4
	Pos303=A	ACTTGCAGGAAACCTCGCCCA	56.3	51.3
Sinilabeo dero	Pos270=T	TATTAGCCTCTTCTGGTGTTGAAGC	56	51
	Pos507=C	TGCTACTTCTCCTATCACTACCAG	55.7	50.7
Systomus sarana	Pos21=C	CCCTAAGCCTTCTTATCCGAGCT	57.1	52.1
sarana	Pos168=A	GGAAACTGACTTGTACCATTAATAATTGGA	56.2	51.2
	Pos522=A	CTCTCACTTCCAGTACTAGCCG	56.7	51.7
Systomus sarana subnasutus		GCCCACGCCTTCGTGATAATTTTC	57.4	52.4
Xenentodon cancila	Pos118=T	CGCCTTTGTAATAATTTTCGTTATAGTAATACC	56.9	51.9
	Pos270=G	TGGCCTCATCTGGAGTTGAGG	56.3	51.3
Lepidocephalicthys	Pos287=C	AACAGTTTACCCGCCCCTAGCA	56.7	51.7
thermalis	Pos522=T	AGGAATTACTATACTTCTTACAGACCGA	55.5	50.5

is centered around identifying specimens based on the precise positions of critical diagnostic nucleotides within DNA barcodes (Weitschek *et al*., 2013). Vargheese *et al*. (2019) said that effectiveness and superiority of this approach in specimen identification. Efforts are underway to automate the creation of character-based keys, recognizing its potential as the most efficient and reliable technique (Vargheese *et al*., 2019).

One promising avenue for enhancing fish identification is the development of character-based diagnostic keys which can be used to create probes for microarrays (Kathirvelpandian *et al*., 2022; Mahapatra *et al*., 2020). Similarly in this study 69 probes were designed for 30 species which include 10 species with three probes, 19 species with two probes and 01 species with a single probe. These microarrays offer faster and more precise fish identification methods which can be of immense value in various applications. One notable example is the Food Expert-ID, a high-density DNA chip commercially developed by bioMerieux, specializing in species identification in food and animal feed through DNA chip technology (Mahapatra *et al*., 2020; Rasmussen and Morrissey, 2008), this device can identify up to 15 different fish species, revealing its potential in species identification. The efficiency of these methods is further bolstered by integrating numerous DNA oligonucleotide probes in a compact area of the chip's surface (Kim *et al*., 2011). This integration allows for the rapid and simultaneous identification of multiple target sequences.

Applying species-specific probes for fish species in the Thamirabarani River holds excellent promise, benefiting both commercial and conservation efforts. These probes can be effectively used in forensic applications, helping identify fish and fish product replacements. Importantly, they are not limited to complete specimens but can also be employed for damaged or processed specimens expanding their utility and relevance in various contexts. The similar study conducted by Van Velzen *et al*. (2012) revealed that BLOG achieved the highest rate of accurate query identification, reaching an impressive 93.1% for actual data and 86.2% for simulated data. This robust performance underscores its standing as a superior method for DNA barcoding applications. One notable advantage of BLOG is its capacity to provide species-level data extending its utility beyond conventional DNA barcoding tasks. These data can find application in diverse realms including species description and molecular detection experiments, broadening their scope and relevance in scientific research. However, it is worth noting that while BLOG generally excels in DNA barcoding, it faces challenges in identifying recently diverged species Van (Velzen *et al*., 2012). BLOG's results thus endorse ongoing efforts to refine and optimize techniques for accurately identifying these species contributing to advancements in molecular biology and taxonomy (Van Velzen *et al*., 2012). In the realm of Character Based Identification Systems (CBIS), a multitude of software programs are employed to delve into intricate biological data. However, BLOG has distinct advantages over similarity-based methodologies and tree-based methods (Weitschek *et al*., 2013).

Character-based identification systems (CBIS) exemplified by the BLOG analysis utilized in this study offers significant advantages over traditional morphological and molecular methods for species identification. CBIS utilizes

specific nucleotide positions within DNA barcodes to classify species, providing rapid, cost-effective and standardized tools for biodiversity assessment and conservation. Unlike morphological identification, which can be subjective and time-consuming. CBIS offers objective criteria and high-throughput capabilities enhancing efficiency in species delimitation and taxonomic assignment. However, CBIS is dependent on the quality and comprehensiveness of reference databases which may limit its application in regions with poorly documented biodiversity. Challenges include its susceptibility to intraspecific variation and difficulties in resolving cryptic species complexes. This study addressed these limitations through rigorous bioinformatics protocols and sensitivity analysis. Moving forward advancements in genomic technologies hold promise for improving CBIS resolution and reliability. Designed species-specific probes also demonstrated broader applications beyond species identification including environmental DNA metabarcoding for ecosystem monitoring and forensic detection of illegal wildlife trade showcasing their versatility in biodiversity conservation and management.

The implementation of designed species-specific probes or a microarray system presents promising opportunities for enhancing biodiversity monitoring and conservation in freshwater ecosystems albeit with considerations regarding feasibility, costs and generalizability. Initial investments in probe development involve significant costs for laboratory equipment, reagents and bioinformatics resources; however, these expenses are offset by long-term benefits in efficiency and accuracy during species identification. Operational costs for probe-based assays are generally lower than traditional methods, making them cost-effective for large-scale surveys and routine monitoring. The feasibility of deploying probe technologies across diverse freshwater systems depends on infrastructure for DNA extraction, analysis facilities and the availability of skilled personnel. Generalizability across different regions requires adaptation to local biodiversity and validation through collaborative efforts to expand reference databases and validate probe performances. Advances in portable sequencing technologies offer promising avenues for extending probe-based approaches to remote or understudied regions, thereby enhancing global conservation efforts and ecosystem management practices.

Accurate fish species identification through advanced molecular techniques such as character-based identification systems (CBIS) and DNA barcoding carries profound implications for freshwater ecosystem dynamics and conservation strategies. By precisely delineating species compositions and distributions, these methods provide crucial insights into community structures, ecological interactions and habitat preferences of aquatic organisms within the Thamirabarani River basin and beyond. Such knowledge forms the foundation for effective biodiversity conservation efforts enabling the detection and management of invasive species that threaten native biodiversity. Moreover, accurate species identification supports ecosystem resilience in the face of environmental stressors, facilitating adaptive management strategies tailored to the specific needs of vulnerable species. Beyond ecological benefits, precise species data underpin evidence-based decision-making in policy frameworks, guiding sustainable fisheries management, habitat restoration initiatives and the establishment of protected areas. By integrating molecular tools into conservation practices, this study underscores their role in enhancing our understanding of freshwater ecosystems and promoting their long-term health and sustainability.

CONCLUSION

The Character-based Identification System (CBIS) for the authentication and conservation of fishes from the Thamirabarani River in the Western Ghats of India provides a robust and reliable method for distinguishing fish species in this ecologically significant region. By employing morphological and molecular markers, CBIS enhances the accuracy of species identification which is critical for monitoring biodiversity, preventing illegal trade and implementing effective conservation strategies. The integration of CBIS into existing management frameworks will facilitate informed decision-making and foster sustainable practices, ultimately contributing to the preservation of the unique fish fauna of the Thamirabarani River. This approach not only strengthens the efforts in conserving the rich aquatic biodiversity but also serves as a model for similar initiatives in other biodiversity hotspots across India and beyond

ACKNOWLEDGEMENT

This study was conducted with the funding support of Tamil Nadu Dr. J. Jayalalithaa Fisheries University, Nagapattinam, India. The authors express their gratitude to the Vice-Chancellor of TNJFU for the invaluable support and guidance.

Conflict of interest

No potential conflict of interest was reported by the author(s).

REFERENCES

Arunkumar, A.A. and Manimekalan, A. (2018). Freshwater fish fauna of rivers of the southern Western Ghats, India. *Earth System Science Data*. **10(3)**: 1735-1752.

Barman, A.S., Singh, M., Singh, S.K., Saha, H., Singh, Y.J., Laishram, M. and Pandey, P.K., (2018). DNA barcoding of freshwater fishes of Indo-myanmar biodiversity hotspot. *Sci. Rep*. **8:** 8579.

Basheer, V.S., Vineesh, N., Bineesh, K.K., Kumar, R.G., Mohitha, C., Venu, S., Kathirvelpandian, A., Gopalakrishnan, A. and Jena, J.K., (2017). Mitochondrial signatures for identification of grouper species from Indian waters. *Mitochondrial DNA Part A*. **28(4):** 451-457.

Bergmann, T., Hadrys, H., Breves, G. and Schierwater, B. (2009). Character-based DNA barcoding: a superior tool for species classification. *Berliner and Münchener Tierärztliche Wochenschrift*. **122:** 446-450.

Bertolazzi, P., Felici, G. and Weitschek, E. (2009a). Learning to classify species with barcodes. *BMC Bioinformatics*. **10:** S7.

Bertolazzi, P., Felici, G. and Weitschek, E. (2009b). Learning to classify species with barcodes. *BMC Bioinformatics*. **10:** S7.

Dahanukar, N., Raut, R. and Bhat, A. (2004). Distribution, endemism and threat status of freshwater fishes in the Western Ghats of India. *Journal of Biogeography*. **31(1):** 123-136.

DeSalle, R. (2006). Species discovery versus species identification in DNA barcoding efforts: response to rubinoff. *Conserv Biol*. **20(5):** 1545-1547.

DeSalle, R., Egan, M.G. and Siddall, M. (2005). The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos Trans R Soc Lond B Biol Sci*. **360(1462):** 1905-1916.

Durairaja, R., Jawahar, P., Jayakumar, N., Das, S.K. and Padmavathy, P. (2022). An annotated checklist of ichthyofaunal diversity of the potamon zone of Thamirabarani River, South India. *Indian Journal of Animal Research*. 1: 9. doi: 10.18805/IJAR.B-4281.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. Oxford. pp. 95-98.

Hebert, P.D.N., Ratnasingham, S. and deWaard, J.R. (2003a). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*. **270:** S96-S99.

Hebert, P.D.N., Ratnasingham, S. and deWaard, J.R. (2003b). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*. **270:** S96-S99.

Hoque, R., Das, S., Biswas, P. and Dey, A., (2023). Ichthyofaunal Diversity of the River Padma at Murshidabad, West Bengal. *Agricultural Science Digest*. **43(4):** 568-571. doi: 10.18805/ag.D-5734.

Kathirvelpandian, A., Chowdhury, L.M. and Sanjeev Kumar, M. (2022). Species-specific molecular signatures for the commercially important scombrids using mitochondrial gene analysis; A tool for fisheries management. *Journal of Asia-Pacific Biodiversity*. **15:** 481-487.

Kim, S., Koo, H., Kim, J.H., Jung, J.W., Hwang, S.Y. and Kim, W. (2011). DNA chip for species identification of Korean freshwater fish: A case study. *BioChip Journal*. **5:** 72-77.

Kundu, S., Chandra, K., Tyagi, K., Pakrashi, A. and Kumar, V. (2019). DNA barcoding of freshwater fishes from brahmaputra river in Eastern Himalaya biodiversity hotspot. *Mitochondrial DNA B Resource*. **4:** 2411-2419. https://doi.org/10.1080/23802359.2019.1637290.

Lakra, W.S., Verma, M.S., Goswami, M., Lal, K.K., Mohindra, V., Punia, P., Gopalakrishnan, A., Singh, K.V., Ward, R.D. and Hebert, P. (2011). DNA barcoding Indian marine fishes. *Molecular Ecology Resources*. **11(1):** 60-71.

Lowenstein, J.H., Amato, G. and Kolokotronis, S.O. (2009). The real maccoyii: identifying tuna sushi with DNA barcodes -contrasting characteristic attributes and genetic distances. *PLoS One*. **4(11):** e7866.

Lynch, A.J., Cooke, S.J., Arthington, A.H., Baigun, C., Bossenbroek, L., Dickens, C., Harrison, I., Kimirei, I., Langhans, S.D., Murchie, K.J. and Olden, J.D. (2023). People need freshwater biodiversity. *Wiley Interdisciplinary Reviews*. pp. 1633.

Mahapatra, S.A.R., Dwivedy, P., Suresh, E., Shanmugam, S.A. and Kathirvelpandian, A. (2020). Character-based identification system of scombrids from Indian waters for authentication and conservation purposes. *Mitochondrial DNA Part B*. **5:** 3221-3224.

Modeel, S., Negi, R.K., Sharma, M., Dolkar, P., Yadav, S., Siwach, S., Yadav, P. and Negi, T., (2024). A comprehensive DNA barcoding of Indian freshwater fishes of the Indus River system, Beas. *Scientific Reports*. **14(1):** 2763.

Mogalekar, H. (2019). Fishes from Tamiraparani river system, Tamil Nadu. *The Indian Journal of Animal Sciences*. **89:** 340-346.

Mogalekar, H.S. (2018). Freshwater fish fauna of Tamil Nadu, India. *Proceedings of the International Academy of Ecology and Environmental Sciences*. **8(4):** 213.

Paine, M.A., McDowell, J.R. and Graves, J.E. (2007). Specific identification of western Atlantic Ocean scombrids using mitochondrial DNA cytochrome c oxidase subunit I (COI) gene region sequences. *Bull Marine Sci*. **80:** 353-367.

Puncher, G.N., Arrizabalaga, H., Alemany, F., Cariani, A., Oray, I.K., Karakulak, F.S., Basilone, G., Cuttitta, A., Mazzola, S. and Tinti, F. (2015). Molecular identification of Atlantic Bluefin Tuna (*Thunnus thynnus*, Scombridae) larvae and development of a DNA character-based identification key for Mediterranean Scombrids. *PLoS One*. **10(7):** e0130407.

Rach, J., DeSalle, R., Sarkar, I.N., Schierwater, B. and Hadrys, H. (2008). Character based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proc Biol Sci*. **275(1632):** 237-247.

Raghavan, R., Prasad, G., Ali, P.H.A. and Pereira, B. (2008). Fish fauna of Chalakudy River, part of Western Ghats biodiversity hotspot, Kerala, India: patterns of distribution, threats and conservation needs. *Biodiversity and Conservation*. **17:** 3119-3131.

Rasmussen, R.S. and Morrissey, M.T. (2008). DNA-based methods for the identification of commercial fish and seafood species. *Comprehensive Reviews in Food Science and Food Safety*. **7:** 280-295.

Rathipriya, A., Kathirvelpandian, A., Shanmugam, S.A., Uma, A., Suresh, E. and Felix, N. (2022). Character-based diagnostic keys, molecular identification and phylogenetic relationships of fishes based on mitochondrial gene from pulicat lake, India: A tool for conservation and fishery management purposes. *Indian Journal of Animal Research*. **56(8):** 933-940. doi: 10.18805/IJAR.B-4905.

Rathipriya, A., Kathirvelpandian, A., Shanmugam, S.A., Suresh, E. and Felix, N. (2021). Character-based identification key for commercially important fishes of Pulicat lake: tool for conservation and management. *Mitochondrial DNA. Part A, DNA, Mapping, Sequencing and Analysis*. **32:** 120-125.

Thompson, J.D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research.* **25:** 4876-4882.

Van Velzen, R., Weitschek, E., Felici, G. and Bakker, F.T. (2012). DNA Barcoding of Recently Diverged Species: Relative Performance of Matching Methods. *PLoS ONE*. **7:** e30490.

Vargheese, S., Chowdhury, L.M. and Ameri, S. (2019). Character based identification system for Elasmobranchs for conservation and forensic applications. *Mitochondrial DNA Part A*. **30:** 651-656.

Weitschek, E., Van Velzen, R., Felici, G., Bertolazzi and P. (2013). BLOG 2.0: A software system for character-based species classification with DNA Barcode sequences. *Mol. Ecol Resour*. **13(6):** 1043-1046.