

DNA-BASED SPECIES IDENTIFICATION OF FOUR IMPORTANT CYPRINIDAE FISH IN THE NORTH OF VIETNAM

Luu Thi Ha Giang¹*, Pham Hong Nhat¹, Vu Thi Trang¹, Vu Thi Huyen¹, Vu Van Sang¹

ABSTRACT

The four Cyprinidae fish including *Cyprinus carpio*, *Semilabeo notabilis*, *Spinibarbus denticulatus* and *Bangana* sp. are valuable fish for aquaculture and fish conservation in the North of Vietnam. DNA-based molecular method is a potential technique for fish species identification and genetic analysis. In this study, 21 mitochondrial DNA cytochrome oxidase I (COI) sequences of the four Cyprinidae fish were collected from previous projects of Research Institute for Aquaculture No.1 and minned on GenBank for further analysis of molecular identification. The nucleotide discrimination showed higher average AT content (55.17%) compared to average GC content (44.84%) in the Cyprinidae fish. The obtained sequences were compared to the GenBank database (NCBI) using the Basic Local Alignment Search Tool (BLAST), a total of 37 sequences with high percentage identity (92.88 to 99.84%) were found. Phylogenetic tree analysis for 58 COI sequences revealed four main branches corresponding to the four genera in the Cyprinidae. The mean genetic Kimura 2 - parameter distances illustrated the close relationship diversity between species (0.004 to 0.058) and genera (0.105 to 0.151). Single nucleotide polymorphism (SNP) analysis of all species in the study revealed distinct features regarding unique sites (31 sites) for four species. DNA sequence based QR codes were generated based on the short sequence of COI containing SNP data for accurate identification of fish species. In addition, the results showed the high Potential of using the DNA-based method to identify Cyprinidae fish in Vietnam.

Keywords: *Bangana* sp., *COI*, *Cyprinus carpio*, *Semilabeo notabilis*, *Spinibarbus denticulatus*, *DNA-based species identification*.

Received: 10 June 2023; revised: 8 August 2023; accepted: 19 August 2023.

1. INTRODUCTION

Cyprinus carpio, *Semilabeo notabilis*, *Spinibarbus denticulatus*, *Bangana* sp. are freshwater fish belonging to the carp family (Cyprinidae), which are important cyprinid fish in the North of Vietnam. Common carp (*Cyprinus carpio*) is one of the main fish species cultured and of great economic importance in Northern Vietnam [1]. Meanwhile, *Semilabeo notabilis*, *Spinibarbus denticulatus* and *Bangana* sp. are

unique and rare fish species and have high economic value in the Northern midlands and mountainous region. The distribution range of these fish is in large rivers of the Red river system (Thao River, Chay River, Lo River, Gam River) and the Thai Binh river system [1]. These are all species of high gastronomy and are considered the king of freshwater fish in the North of Vietnam. Nevertheless, along with the overfishing and the degradation of the environment, these fish are increasingly scarce as well as in danger of extinction, and have been named in the Vietnam Red data book (1992) [2].

¹ Research Institute for Aquaculture No. 1

* Email: hagiang@ria1.org

There was a report of several confusions among these species from different local areas in Vietnam due to the similar morphology of Ciprinid fish [3]. *Semilabeo notabilis*, *Spinibarbus denticulatus*, *Bangana* sp. have almost the same shape of body; they all live in the same habitat. These fish have the habit of eating moss on rocks with their mouths, it formed the structure of the fish's mouth is quite special for the purpose of finding food. It is challenging for non-taxonomists to accurately recognize and identify these species based on the conventional methods. Therefore molecular identification can complement morphological identification data.

Mitochondrial DNA cytochrome oxidase I (COI) sequence information plays a vital role in identifying species [4]. This gene sequence is conserved within a species, and the rate of mutation is fast enough to distinguish between closely related species [5]. Therefore, the COI sequencing method, i.e., DNA barcoding, has been used as a standard tool for species identification [6]. DNA barcoding can help to eliminate existing misidentification and the availability of cryptic species that mimic and equally compromise the accuracy of fishes in research, fishery management and conservation [7]. A number of countries have had DNA barcoding done on freshwater and marine

fishspecies. Until now, more than 502,000 specimens, belonging to 26,187 fish species, have been successfully identified in BOLD [8].

In Vietnam, seven hundred and sixty-five COI sequences of coastal ray-finned fishes were generated, belonging to 458 species, 273 genera, 113 families and 43 orders, was reported for DNA barcoding study by Pham The Thu *et al.* (2019) [9]. The mtDNA COI sequence analysis also has been used to understand genetics structure and species identification of several fish species including common carp [10]; the Mastacembelidae [11].

The present study was aimed to estimate genetic differences among the four cyprinid species collected from the North of Vietnam using the different computational methods e.g., data mining, nucleotide discrimination, distance, phylogenetics, and Single Nucleotide Polymorphism (SNP) of fish species in the carp family (Cyprinidae). The DNA sequence data generated from this study was used to develop a "Quick Response Code" (QRC). This study provides important information for future conservation and fisheries resource management of four important cyprinid fish species in Vietnam.

2. MATERIALS AND METHODS

2.1. Fish database of COI gene

Table 1. List of fish species (Order: Cypriniformes; Family: Cyprinidae) and sample sequences used in the analysis

Vietnamese name	Scientific name	Genus	Location	Genbank (Accession Number)	Source
Anh Vũ	<i>Semilabeo notabilis</i>	Semilabeo	Thai Nguyen, Ha Giang, Phu Tho, Cao Bang	OP269716 → OP269720	RIA1 (2023)
Cá Bống	<i>Spinibarbus denticulatus</i>	Spinibarbus	Ha Giang, Tuyen Quang, Hoa Binh	MW446147 → MW446157	RIA1 (2021)
Cá chép	<i>Cyprinus carpio</i>	Cyprinidae	Cao Bang	MW450995 → MW450999	RIA1 (2021)
Rầm xanh	<i>Bangana</i> sp.	Bangana	Ma river, Thanh Hoa	KY498533	[12]

COI gene sequences were collected from several previous reports which focused on the *Cyprinidae* in the mountainous of Northern Vietnam, involving *Semilabeo notabilis*; *Spinibarbus denticulatus*; *Cyprinus carpio*, *Bangana* sp. The sampling and sequencing activities were conducted by our team in Research Institute for Aquaculture No. 1 from 2021 to 2023. In total, 20 sequence data (COI gene) of fish species (*Semilabeo notabilis*; *Spinibarbus denticulatus*; *Cyprinus carpio*) were collected and recently published in GenBank. The *Bangana* sp. isolated from Vietnam were mined on NCBI (<https://www.ncbi.nlm.nih.gov/>). Total 21 COI sequences subjected to this study were shown in the table 1.

2.2. Nucleotide discrimination and BLAST annotation

Compositional features for COI sequences were calculated for the base number of the four nucleotides and GC content for each sequence as well as an overall average by using MEGAX software [13].

The BioEdit software [14] was also used to check and determine the degree of similarity of sequences and to create the consensus sequence of each species.

The Basic Local Alignment Search Tool (BLAST) is highly efficient for determining sequence similarities with reference sequences from GenBank. The input consensus sequences were compared with the maximum similarity data sets of fish species based on high BLAST identity percentage with the lowest E-value for the generated alignment. The 37 validated reference sequences for all fish species were downloaded from GenBank for utilization in the construction of a phylogenetic evolutionary tree (neighbor-joining tree).

2.3. Multiple sequence alignment, estimates of evolutionary divergence and phylogenetic analysis

A sequence file including the 21 Vietnamese sequences and 37 reference sequences was aligned through MEGAX using the CLUSTAL

alignment tool. Additionally, all sequences were edited manually, i.e., similar, highly mismatched sites and gaps were removed; each base of the spliced sequence was checked before submission to other analysis. Analysis of genetic distance between populations was conducted using MEGA; the evolutionary history was inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site.

The phylogenetic trees were constructed by using algorithm programs in MEGAX software. The bootstrap support of the Neighbor-Joining tree was assessed using 1000 replicates.

2.4. Single Nucleotide Polymorphism (SNP) Screening and DNA barcode generation

Based on the above alignment data and SNPs were detected manually for estimation of unique sites same as described by Yang *et al.* (2019) [15].

QR code is easily accessible two-dimensional barcode, readable by smartphones. It allows to encode over 4000 characters in a two-dimensional barcode. The COI fragments containing SNPs were used for the development of DNA barcodes for each species using an online QR code generator (<https://www.online-qrcode-generator.com/>).

3. RESULTS AND DISCUSSION

3.1. Nucleotide composition

A total of 21 COI sequences were generated from 4 Vietnamese cyprinid fish mined from previous reports and GenBank (Table 1). The sequence length ranged from 604 bp to 884 bp (mean 725 bp) and there was not any codon, insert or deletion in the sequences. The average nucleotide compositions (Table 2) were 26.32% adenine (A), 28.87% cytosine (C), 17.93% guanine (G) and 26.88% thymine (T). This result corresponds to previous studies in genus *Capoeta* (Family Cyprinidae) that showed the average nucleotide frequencies were C (29.24%), T (26.82%), A (26.53%), G (17.41%) [16].

Table 2. Comparison of the nucleotide composition among the 4 Vietnamese Cyprinidae

Fish species	Total (bp)	A (%)	T (%)	G (%)	C (%)	GC (%)	AT (%)
<i>Semilabeo obscurus</i>	652	26.17	28.65	17.82	27.36	45.16	54.84
<i>Spinibarbus denticulatus</i>	884	26.73	29.36	18.12	25.78	44.00	56
<i>Cyprinus carpio</i>	760	27.05	29.47	17.55	25.92	43.48	56.52
<i>Bangana sp.</i>	604	25.33	27.98	18.21	28.48	46.7	53.3
<i>Average</i>	725	26.32	28.87	17.93	26.88	44.84	55.17

The Cyprinidae fish species nucleotide discrimination revealed varied AT (the percentage of adenine + thiamine) and GC (the percentage of guanine+ cytosine) contents. Among the four species, the observed nucleotide base composition of all analyzed sequences was average at 55.17% AT (range from 53.3 to 56.52%) and 44.84% GC (from 43.48 to 46.7%) (Table 2, figure 1). Overall, the results demonstrated that for the Cyprinidae

fish species, the average AT content (55.17%) was higher than average the GC content (44.84%). This result is consistent with previous studies in fish species that reported higher AT content than GC content [17]. Reviewed data from the 655 bp *COI* region was on average GC content in the 44 freshwater species about 45.2% [18]; from 89 fish species (GC content = 45.48%) [19], and from 79 species of fish with GC content=45.78% [20] .

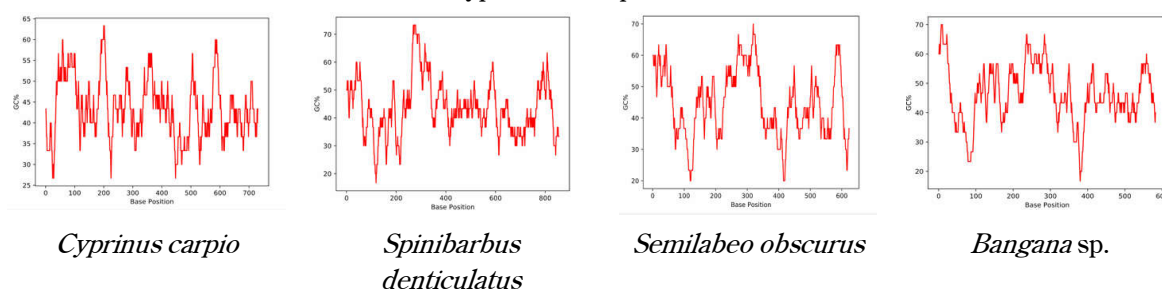


Figure 1. Comparison of the average GC among the 4 Cyprinidae

Genomic DNA base composition and especially its GC content plays an important role in the functioning and regulation of genes [21]. The GC content is known to vary widely between genomes of different species and also serve as an important criterion in taxa delimitation [22]. This genomic trait has been widely studied, and its evolution has been proposed to be associated with numerous mutational and selective forces driven by genetic, metabolic, and ecological factors [23]. Our analysis revealed that the highest nucleotide base composition was 46.7% GC in *Bangana sp.* and 45.6% in *Semilabeo obscurus*. In contrast, *Spinibarbus denticulatus* has lower GC content with *COI* gene at 44% and lowest GC content was observed in *Cyprinus carpio* at 43.38%. It could be explained by the different habitats of the fish

species because higher GC contents were found to be mostly associated with species inhabiting more stable environments. Hence, *Bangana sp.* and *Semilabeo obscurus* are sharing the same ecological conditions (high water quality, low water temperatures), less suffer from the stress of the environment and similar morphology. In the other hand, *Spinibarbus denticulatus* and *Cyprinus carpio* have a broader spectrum of habitats.

3.2. Similarity-based search of the GenBank databases

The results of comparing the nucleotide consensus sequences of studied fish samples with GenBank data through NCBI were shown in table 3. The BLAST results showed all of the taxa available in a database that have sequence similarity with the query sequence. A total of

fifteen COI sequences of fish were selected with percentage identity ranging from 92.88 to 99.84% across the fish sequences (Table 3). The expected value, e-value and maximum coverage obtained from the sequences ranged from 0.0 and 93 - 100% (online data), respectively. A total of 4 genera in the Cyprinidae were also detected among the fish samples. Total 37 sequences with maximum identity percent (ID) score and query cover were selected for estimating of evolutionary divergence and building phylogenetics tree. These were *Cyprinus carpio* (number, n=18), *Semilabeo* sp.

(n=11), *Spinibarbus denticulatus* (n=5), and *Bangana* sp. (n=3). For the total number of species, 6 (*Cyprinus carpio*, *Semilabeo obscurus*, *Semilabeo notabilis*, *Spinibarbus denticulatus*, *Bangana decora*, *Bangana lemasoni*) of them were found among the fish sequences. According to the sampling countries, 22 sequences isolated from the fish belonging to China; 6 sequences from USA; 4 sequences from Vietnam, 2 sequences from Taiwan, one from Japan and one sequence was recorded in Japan.

**Table 3. Similarity results and reference sequence (Accession no.)
for fish species identified by BLAST/GenBank**

Scientific name	BLAST hits in NCBI database	Percent identities (%)	GenBank Accession No.	Year of publication	Source of sequence
<i>Cyprinus carpio</i>	<i>Cyprinus carpio</i>	98.95	MK291479.1	2018	China
	<i>Cyprinus carpio</i>	98.95	MH202953.1	2019	China
	<i>Cyprinus carpio</i>	98.95	MG570435.1	2021	USA
	<i>Cyprinus carpio</i>	98.95	KP993139.1	2015	China
	<i>Cyprinus carpio</i>	98.95	KP993137.1	2015	China
	<i>Cyprinus carpio</i>	98.95	OM736811.1	2022	USA
	<i>Cyprinus carpio</i>	98.95	OL693871.1	2022	USA
	<i>Cyprinus carpio</i>	98.95	OL457418.1	2022	USA
	<i>Cyprinus carpio</i>	98.95	MZ713633.1	2022	Taiwan
	<i>Cyprinus carpio</i>	98.95	MW125611.1	2022	China
	<i>Cyprinus carpio</i>	98.95	OM234682.1	2022	China
	<i>Cyprinus carpio</i>	98.95	KJ511883.1	2014	Hungary
	<i>Cyprinus carpio</i>	98.95	KF856965.1	2014	China
	<i>Cyprinus carpio</i>	98.95	MW680808.1	2021	China
	<i>Cyprinus carpio</i>	98.95	NC_018039.1	2023	USA
	<i>Cyprinus carpio</i>	98.95	NC_018037.1	2023	USA
	<i>Cyprinus carpio</i>	98.95	JN673560.1	2016	Taiwan

	<i>Cyprinus carpio</i>	98.68	NC_018366.1	2013	China
<i>Semilabeo obscurus</i>	<i>Semilabeo notabilis</i>	99.84	KY615364.1	2017	Vietnam
	<i>Semilabeo notabilis</i>	99.69	NC_045916.1	2023	China
	<i>Semilabeo obscurus</i>	99.69	MN229749.1	2021	China
	<i>Semilabeo obscurus</i>	99.69	MN229748.1	2021	China
	<i>Semilabeo obscurus</i>	99.68	GU086581.1	2016	China
	<i>Semilabeo notabilis</i>	99.67	KY615365.1	2017	Vietnam
	<i>Semilabeo obscurus</i>	99.54	NC_037408.1	2023	China
	<i>Semilabeo obscurus</i>	99.54	MN229747.1	2021	China
	<i>Semilabeo notabilis</i>	96.39	KY615363.1	2017	Vietnam
	<i>Semilabeo notabilis</i>	96.26	KT633633.1	2017	Vietnam
	<i>Semilabeo notabilis</i>	95.98	JX074195.1	2012	China
<i>Spinibarbus denticulatus</i>	<i>Spinibarbus denticulatus</i>	97.99	JX042168	2013	China
	<i>Spinibarbus denticulatus</i>	97.97	GU086582	2016	China
	<i>Spinibarbus denticulatus</i>	97.96	AP013335	2016	Japan
	<i>Spinibarbus denticulatus</i>	97.79	KJ994631	2017	China
<i>Bangana</i> sp.	<i>Bangana decora</i>	95.7	NC_026221	2023	China
	<i>Bangana decora</i>	95.96	MG732840	2019	China
	<i>Bangana lemassoni</i>	92.88	GU86575	2016	China

The three species of *Semilabeo obscurus*, *Spinibarbus denticulatus* and *Bangana* sp. are native fish, distributing mainly in the highlands of Vietnam, Lao and China and rarely in other Asian countries. Therefore, it is predictable that fewer number of identified Cyprinid species in GenBank for these three species compared to *Cyprinus carpio*. In this study, there were two species found by BLAST for the *Semilabeo obscurus* (RIA1's specimen), which are *Semilabeo notabilis* and

Semilabeo obscurus. Interestingly, the *Semilabeo notabilis* matched with higher percent identities (99.84% and 99.69%, respectively) and also lower percent identities (from 95.98 to 96.39%) compared to the *Semilabeo obscurus* from China (from 99.54 to 99.69% identities). It might be missidentification of this species of previous study, since literally only specimen with a sequence similarity of 97% or above are considered to be a single species. In addition, BLAST results (Table 3) indicated that

the *Bangana* sp. from Vietnam might be not one of *Bangana decora* or *Bangana lemasoni* species (from 92.88 to 95.7% identities), more studies and larger databases are needed to identify exactly the

species name of *Bangana* specimens from Vietnam.

3.3. Genetic diversity among samples and evolutionary relationship Phylogenetic analysis

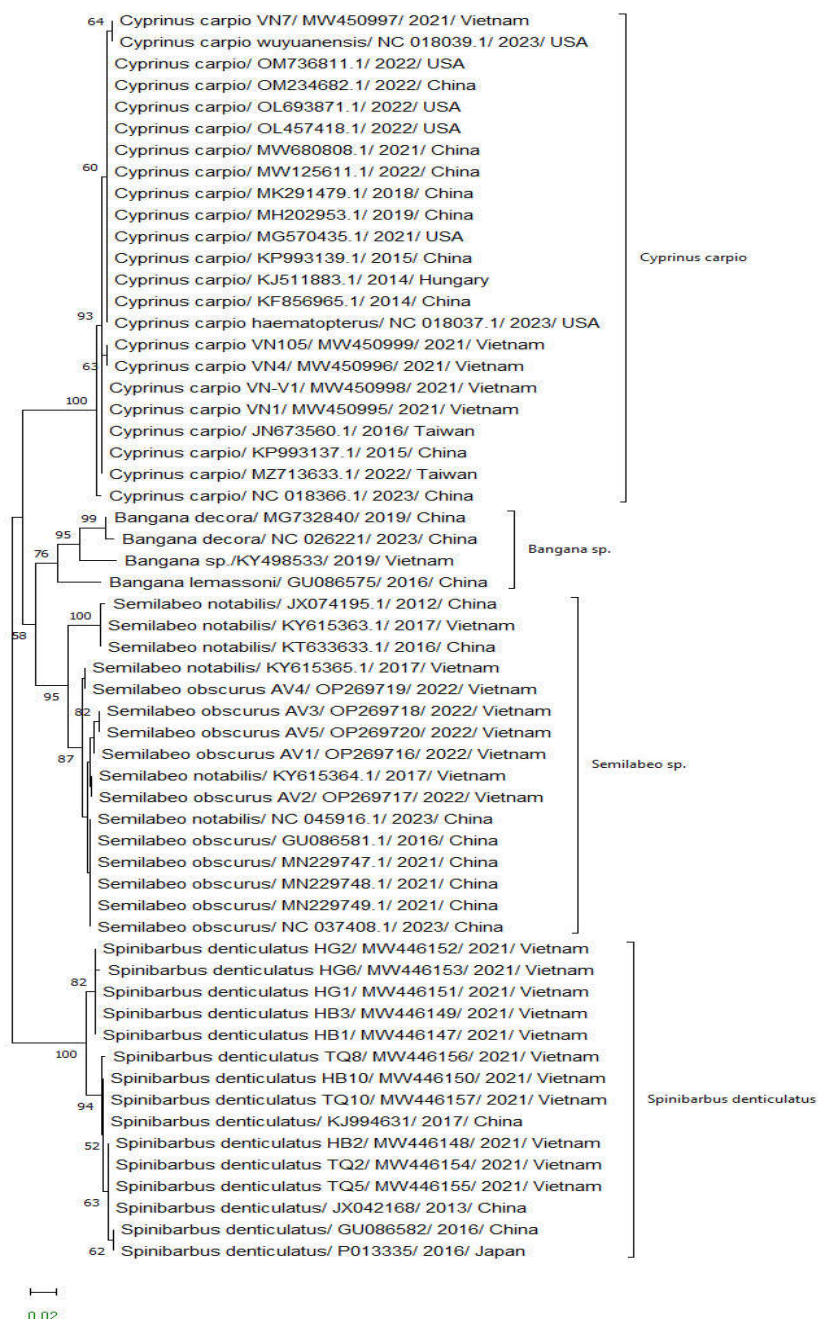


Figure 2. Neighbor-Joining tree of the Cyprinidae based on 58 COI sequences.

Note: Numbers at tree branches indicate the percentage of replicate trees in which the same clusters are found in the bootstrap tests of 1000 replicates.

The phylogenetic tree of the studied fish in figure 2 was divided into 4 main groups corresponding to the 4 subfamilies in the Cyprinidae. Therein, The first group is *Spinibarbus denticulatus* included 2 clusters indicated with the highest bootstrap value (100%). Interestingly, among 11 fish samples isolated from Vietnam, there are 5 haplotypes from Ha Giang and Hoa Binh provinces which were grouped into separate clusters, while other 6 haplotypes have closer relationship with *Spinibarbus denticulatus* from China and Japan. Group 2 is *Cyprinus carpio*, which show close relationship between fish from variety regions China, Taiwan, The USA and Vietnam (100% value of bootstrap). Group 3 included two branches of *Semilabeo* (95% value of bootstrap); but it seems to be no distinction between *Semilabeo obscurus* and *Semilabeo notabilis* in both countries. Group 4 indicated with lowest bootstrap value (76%) included 4 divergence *Bangana* sp. The isolated from Vietnam could be more closed to *Bangana decora*, since it cluster with two *Bangana decora* isolated from China with well supported by bootstrap value with 95%, the BLAST search (Table 3) showed similar result that Vietnamese *Bangana* sp. is 95.96% identities with Chinese *Bangana decora*.

The phylogenetic relationship of Cyprinidae has also been published by many authors. In which, Thai *et al.* (2007), examined the phylogenetic relationships in the subfamilies within Cyprinidae using mitochondrial 16S rRNA, D-loop and cytochrome b gene sequences from 25 species of cyprinids collected from Vietnam combined with sequences of cyprinids available in GenBank, the result showed that there are two principal lineages within the Cyprinidae: Cyprinines and Leuciscines, but many of the subfamily boundaries remained unclear [24]. Other study using 16S rRNA by Sharma *et al.* (2014), indicated that the family Cyprinidae of Agra region was resolved as a paraphyletic group which shows the divergence may occur, but they share the same common ancestor [25]. In recent

study, Alam *et al.* (2021) provided more information about the phylogenetic relationship between cyprinid fishes species in Bangladesh, in trees, Leuciscinae subfamily showed monophyletic lineage - where Garrinae, Schizothoracinae, Rasborinae, Cultrinae, Cyprininae subfamilies were polyphyletic [26]. These studies all revealed a complex evolutionary relationship of the Cyprinidae and suggested further studies on phylogenetic analysis of Cyprinidae fishes below the level of family.

In Vietnam, several studies conducted in RIA1 have been reported the phylogenetics relationship within species of *Semilabeo obscurus* [27], *Spinibarbus denticulatus*, *Cyprinus carpio* [10]. In this study, the combination in the Neighbor - Joining tree indicated the relationship between these species, *Semilabeo* sp. might be closely related to *Bangana* sp., since two genera were placed in one branch supported by bootstrap value with 58% (Figure 2).

3.4. Genetic Divergence (K2P) Among Taxa

The estimates of evolutionary divergence between sequences of the COI were determined using Kimura's two-parameter model of nucleotide substitution (Table 4). Intergeneric genetic distances ranged from 0.105 to 0.151. The highest estimates of evolutionary divergence among groups were found between *Spinibarbus denticulatus* and *Bangana* sp. (0.151 ± 0.027), and the lowest between *Spinibarbus denticulatus* and *Cyprinus carpio* (0.105 ± 0.019). The divergence between *Cyprinus carpio* and *Semilabeo* sp. and the divergence between *Cyprinus carpio* and *Bagana* sp. were similar value at 0.135 ± 0.023 compared to 0.136 ± 0.023 , respectively.

The mean K2P distance of species within genus was from 0.004 to 0.058. The highest K2P genetic distances was found in *Bangana* sp. (0.058 ± 0.011), then *Semilabeo* sp. (0.018 ± 0.005), *Spinibarbus denticulatus* (0.013 ± 0.005), and lowest genetic distances was detected in *Cyprinus carpio* (0.004 ± 0.002).

Table 4. Summary of Kimura 2-parameter genetics divergence within (the diagonal) and between species (below the diagonal) of cyprinid fishes based on 58 COI sequences

	<i>Cyprinus carpio</i>	<i>Semilabeo</i> sp.	<i>Spinibarbus denticulatus</i>	<i>Bangana</i> sp.
<i>Cyprinus carpio</i>	0.004 ± 0.002			
<i>Semilabeo</i> sp.	0.135 ± 0.023	0.018 ± 0.005		
<i>Spinibarbus denticulatus</i>	0.105 ± 0.019	0.120 ± 0.023	0.013 ± 0.005	
<i>Bangana</i> sp.	0.136 ± 0.023	0.145 ± 0.026	0.151 ± 0.027	0.058 ± 0.011

In case of the *Semilabeo* sp., the BLAST result illustrated two species with high percentage of identities (Table 3). Additionally, the output of the phylogenetic tree (Figure 2) and the genetics divergence (Table 4) showed conservative sequences of COI gene within the *Semilabeo* genus. These findings showed close evolution and lower genetic distances of *Semilabeo* species between different regions (0.018 ± 0.005). In contrast, Ha T.T.T *et al.* (2018) indicate higher genetic distance (0.04) between two *Semilabeo* groups from different provinces in Vietnam [28].

In the case of other species *Cyprinus carpio*, *Spinibarbus denticulatus*, *Bangana* sp., the Kimura 2 - parameter, genetics divergence result in this study is consistent with previous studies, the intraspecific genetic distances based on K2P are usually low (below 1) and are rarely greater than 2 across a broad range of taxa [29]. All the genetics distances in this study showed less than 1 sequence diversity within species and showed less than 2 sequence diversity between species, which indicated no increase in genetic variability relative to species (Table 4).

3.5. Single Nucleotide Polymorphism screening and QR Codes generation

From the sequence alignments, there were genetic variations at a nucleotide level as determined at different positions of the representative sequences. Base on the result of

phylogenetics tree, the four species including 51 sequences was selected for identification of variable sites (Table 5). Consequently, the original data of COI - studied sequences were selected into a specific short segments (235 bp) of COI barcode (from site 465 to site 699). There were 31 character attribute sites were observed over the DNA barcode. The sample data generated in the present study were showing lowest number of variable sites (4 positions) for the *Bangana* genus. Whereas, there was nine diverse sites detected for each species involving *Semilabeo obscurus*, *Spinibarbus denticulatus*, *Cyprinus carpio*. These differences may be due to the replacement mutation occurred in the groups that can cause the phenotypic differences in fish species, which could be used as potential markers for identification of the Cyprinidae.

At the nucleotide level, the sequence alignment revealed much polymorphism at different positions and this similar degree of variations had been previously observed in several fish species. Previous reports yielded 56 variable sites in 43 sequences of *Steindachneridi on scriptum* [30]; and 76 variable sites in 74 fish of *Engraulis encrasicolus* [31]. The existing variations show a high degree of heterogeneity within the studied fishes. In this study, the polymorphic sites were observed for different species, however it could not represent for different countries.

Table 5. Single Nucleotide Polymorphism observed in the different species of the Cyprinidae

	Nucleotide positions	465	474	475	483	486	498	504	519	525	531	547	549	555	561	564	573	576	582	585	588	595	597	600	624	636	645	651	672	684	693	699
1	Bangana decora/ MG732840/ 2019/ China	C	T	T	T	C	T	C	C	C	C	C	A	C	C	A	T	C	A	T	T	C	A	T	G	T	A	C	A	A	C	C
2	Bangana decora/ NC 026221/ 2023/ China	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	Bangana sp./KY498533/ 2019/ Vietnam	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	Cyprinus carpio haematopterus/ NC 018037.1/ 2023/ USA	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
5	Cyprinus carpio VN-V1/ MW450998/ 2021/ Vietnam	T	.	C	A	.	.	.	.	.	.	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
6	Cyprinus carpio VN1/ MW450995/ 2021/ Vietnam	T	.	C	A	.	.	.	.	.	.	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
7	Cyprinus carpio VN105/ MW450999/ 2021/ Vietnam	T	.	C	A	.	.	.	.	.	.	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
8	Cyprinus carpio VN4/ MW450996/ 2021/ Vietnam	T	.	C	A	.	.	.	.	.	.	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
9	Cyprinus carpio VN7/ MW450997/ 2021/ Vietnam	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
10	Cyprinus carpio wuyuanensis/ NC 018039.1/ 2023/ USA	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
11	Cyprinus carpio/ JN673560.1/ 2016/ Taiwan	T	.	C	A	.	.	.	.	.	.	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
12	Cyprinus carpio/ KF856965.1/ 2014/ China	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
13	Cyprinus carpio/ KJ511883.1/ 2014/ Hungary	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
14	Cyprinus carpio/ KP993137.1/ 2015/ China	T	.	C	A	.	.	.	.	.	.	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
15	Cyprinus carpio/ KP993139.1/ 2015/ China	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
16	Cyprinus carpio/ MG570435.1/ 2021/ USA	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
17	Cyprinus carpio/ MH202953.1/ 2019/ China	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
18	Cyprinus carpio/ MK291479.1/ 2018/ China	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
19	Cyprinus carpio/ MW125611.1/ 2022/ China	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
20	Cyprinus carpio/ MW680808.1/ 2021/ China	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
21	Cyprinus carpio/ MZ713633.1/ 2022/ Taiwan	T	.	C	A	.	.	.	.	.	.	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
22	Cyprinus carpio/ NC 018366.1/ 2023/ China	T	.	C	A	.	.	.	.	.	.	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
23	Cyprinus carpio/ OL457418.1/ 2022/ USA	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
24	Cyprinus carpio/ OL693871.1/ 2022/ USA	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
25	Cyprinus carpio/ OM234682.1/ 2022/ China	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
26	Cyprinus carpio/ OM736811.1/ 2022/ USA	T	.	C	A	.	.	.	.	.	T	.	G	.	.	G	C	.	G	C	.	T	.	.	.	C	T	.	.	.	T	.
27	Semilabeo obscurus AV1/ OP269716/ 2022/ Vietnam	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
28	Semilabeo obscurus AV2/ OP269717/ 2022/ Vietnam	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
29	Semilabeo obscurus AV3/ OP269718/ 2022/ Vietnam	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
30	Semilabeo obscurus AV4/ OP269719/ 2022/ Vietnam	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	.	.	T	T	.
31	Semilabeo obscurus AV5/ OP269720/ 2022/ Vietnam	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
32	Semilabeo obscurus/ GU086581.1/ 2016/ China	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
33	Semilabeo obscurus/ MN229747.1/ 2021/ China	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
34	Semilabeo obscurus/ MN229748.1/ 2021/ China	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
35	Semilabeo obscurus/ MN229749.1/ 2021/ China	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
36	Semilabeo obscurus/ NC 037408.1/ 2023/ China	A	.	C	C	T	.	.	.	.	.	T	.	T	A	.	C	.	.	.	.	.	G	C	.	T	.	G	.	T	T	.
37	Spimbarbus denticulatus HB1/ MW446147/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	.	T	.	.
38	Spimbarbus denticulatus HB10/ MW446150/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.
39	Spimbarbus denticulatus HB2/ MW446148/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.
40	Spimbarbus denticulatus HB3/ MW446149/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	.	T	.	.
41	Spimbarbus denticulatus HG1/ MW446151/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	.	T	.	.
42	Spimbarbus denticulatus HG2/ MW446152/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	.	T	.	.
43	Spimbarbus denticulatus HG6/ MW446153/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	G	.	T	.	.
44	Spimbarbus denticulatus TQ10/ MW446157/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.
45	Spimbarbus denticulatus TQ2/ MW446154/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.
46	Spimbarbus denticulatus TQ5/ MW446155/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.
47	Spimbarbus denticulatus TQ8/ MW446156/ 2021/ Vietnam	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	G	.	A	.	C	T	.	.	T	T	.	.
48	Spimbarbus denticulatus/ GU086582/ 2016/ China	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.
49	Spimbarbus denticulatus/ JX042168/ 2013/ China	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.
50	Spimbarbus denticulatus/ KJ994631/ 2017/ China	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.
51	Spimbarbus denticulatus/ P013335/ 2016/ Japan	.	C	C	C	.	C	T	T	T	.	.	.	T	.	C	T	.	C	.	.	.	A	.	C	T	.	.	T	T	.	.

Note: “.” Represented the similarity of sequences, and Character attributes were highlighted in red. Nucleotide positions noted from the trimmed COI sequence after alignment, and based on the complete COI gene reference of *Cyprinus carpio* 1551 bp from the GenBank database (accession number OM736811.1)

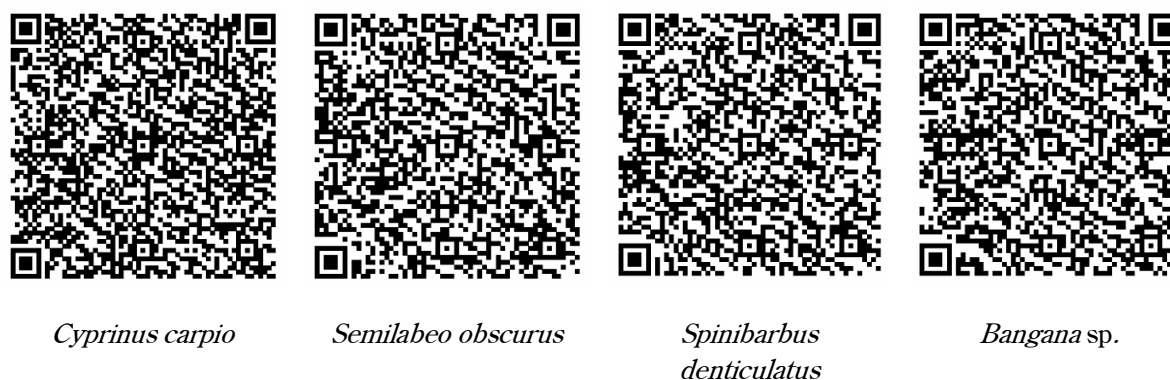


Figure 3. QR codes generated using the 235 bp COI gene fragments

Moreover, all the Cyprinidae COI barcoding fragment (235 bp) were used to generate scannable QR codes. DNA sequence based QR codes for the Vietnamese *Bangana* sp. *Semilabeo obscurus*, *Spinibarbus denticulatus*, *Cyprinus carpio* are given in figure 3. We have developed DNA sequence based QR codes that can be scanned using mobile phone applications in the same way that barcodes are scanned in supermarkets. This study generated QR codes for the identification of 4 fish species belonging to the family of Cyprinidae based on molecular approaches. The DNA QR code can be used as the tag or label in the tracking system, this technique can be applied to develop the native fish products on blockchains for different Cyprinid, however it is not suitable for different countries. According to the previous reports, Yang *et al.* (2019) developed one-dimensional DNA barcode as an example for the precise identification of Teleost fish species [15]. Ghouri *et al.* (2020) targeting 17 commercially available freshwater and marine fish species of Pakistan, based on using QR barcodes of DNA unique SNP [32]. The use of species authentication supported by DNA barcoding could provide an effective approach for monitoring, management, and conservation of the fisheries sector.

4. CONCLUSION

DNA-based molecular method showed the efficiency in identification of the four Cyprinidae fishes at the species level in Vietnam. In this study, 21 Vietnamese specimens and 37

sequences mined from Genbank were analyzed of nucleotide composition, phylogenetic relationship, and SNP using DNA-based method. *Semilabeo notabilis* and *Bangana* sp. showed genetically closer relationship compared to *Spinibarbus denticulatus* and *Cyprinus carpio*. The 31 polymorphic sites were observed for different species, however it could not represent for different countries. DNA sequence based QR codes were generated based on the short sequence of COI containing SNP data for accurate identification of fish species. The protocol of DNA-based analysis in this present study may serve as a reference for accurate identification of fishes that may facilitate ichthyological research, biodiversity management and authentication of fish products in Vietnam.

REFERENCES

1. Hao N.V and Van N.S. (2001). *Freshwater fish in Vietnam. Carp family (Volume 1)*, Agriculture Publishing House, Ha Noi.
2. Vietnam Ministry of Science Technology and Environment (1992). *Red data book of Vietnam, vol 1 Animals*, Ha Noi. Science and Technics Publishing House.
3. Vietnam Agriculture Newspaper (2020). Dam xanh culture. <https://nongnghiep.vn/nuoi-ca-dam-xanh-d266083.html> accessed: 09/07/2023.
4. Bingpeng X., Heshan L., Zhilan Z., Kumar N.K. and Ramamurthy V.V. (2018). DNA barcoding for identification of fish species in the Taiwan Strait. *PLOS ONE*, 13 (6), e0198109.

5. Rebijith K.B., Asokan R., Krishna V., *et al.* (2012). Development of Species-Specific Markers and Molecular Differences in Mitochondrial and Nuclear DNA Sequences of *Aphis Gossypii* and *Myzus Persicae* (hemiptera: Aphididae). *The Florida Entomologist*, 95 (3), 674 - 682.
6. Hebert P.D.N. and T. Ryan Gregory (2005). The Promise of DNA Barcoding for Taxonomy. *Systematic Biology*, 54 (5), 852 - 859.
7. Bortolus A. (2008). Error Cascades in the Biological Sciences: The Unwanted Consequences of Using Bad Taxonomy in Ecology. *ambi*, 37 (2), 114-118.
8. Actinopterygii | Taxonomy Browser | BOLDSYSTEMS.<http://www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=77>, accessed: 09/08/2023.
9. Thu P.T., Huang W.-C., Chou T.-K., Van Quan N., Van Chien P., Li F., Shao K.T. and Liao T.Y. (2019). DNA barcoding of coastal ray-finned fishes in Vietnam. *PLOS ONE*, 14 (9), e0222631.
10. Trang V.T., Huyen V.T., Nhat P.H., Giang L.T.H., Chi C.T.L., Lua D.T., and Khoi L.V. (2021). Genetic diversity of *spinibarbus denticulatus* populations using molecular marker. *TNU Journal of Science and Technology*, 226 (14), 114 -121.
11. Duong T.-Y., Tran L.V.D., Nguyen N.-T.T., Jamaluddin J.A.F. and Azizah M.N.S. (2020). Unravelling taxonomic ambiguity of the Mastacembelidae in the Mekong Delta (Vietnam) through DNA barcoding and morphological approaches. *Tropical Zoology*, 33 (2).
12. Zheng L.-P., Chen X.-Y., and Yang J.-X. (2019). Molecular phylogeny and systematic revision of *Bangana sensu lato* (Teleostei, Cyprinidae). *Journal of Zoological Systematics and Evolutionary Research*, 57 (4), 884-891.
13. Kumar S., Stecher G., Li M., Knyaz C. and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*, 35 (6), 1547-1549.
14. Hall T., Hall T., and Alzohairy D.A.M. BioEdit: An important software for molecular biology.
15. Yang C.-H., Wu K.-C., Chuang L.-Y., and Chang H.W. (2019). Decision Theory-Based COI-SNP Tagging Approach for 126 Scombriformes Species Tagging. *Frontiers in Genetics*, 10.
16. Bektaş Y., Aksu I., Kaya C., and Turan D. (2019). DNA barcoding of the Genus *Capoeta* (Actinopterygii: Cyprinidae) from Anatolia. *Turkish Journal of Fisheries and Aquatic Sciences*.
17. Ward R.D., Zemlak T.S., Innes B.H., Last P.R. and Hebert P.D. (2005). DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci*, 360 (1462), 1847-1857.
18. Ude G.N., Igwe D.O., Brown C., Jackson M., Bangura A., Ozokonkwo-Alor O., Iheahalu O.C., Chosen O., Okoro M., Ene C. and Chieze V. (2020). DNA barcoding for identification of fish species from freshwater in Enugu and Anambra States of Nigeria. *Conservation Genet Resour*, 12 (4), 643 - 658.
19. Henriques J.M., Silva G.J.C., Ashikaga F.Y., Hanner R., Foresti F. and Oliveira C. (2015). Use of DNA barcode in the identification of fish species from Ribeira de Iguape Basin and coastal rivers from São Paulo State (Brazil). *DNA Barcodes*, 3 (1).
20. Díaz J., Villanova G.V., Brancolini F., Del Pazo F., Posner V.M., Grimberg A. and Arranz S.E. (2016). First DNA Barcode Reference Library for the Identification of South American Freshwater Fish from the Lower Paraná River. *PLOS ONE*, 11 (7), e0157419.
21. Li X.-Q. and Du D. (2014). Variation, Evolution, and Correlation Analysis of C+G Content and Genome or Chromosome Size in Different Kingdoms and Phyla. *PLOS ONE*, 9 (2), e88339.
22. Šmarda P., Bureš P., Horová L., Leitch I.J., Mucina L., Pacini E., Tichý L., Grulich V. and Rotreklová O. (2014). Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proceedings of the National Academy of Sciences*, 111 (39), E4096-E4102.
23. Chaurasia A., Uliano E., Berná L., Agnisola C., and D'Onofrio G. (2011). Does Habitat Affect the Genomic GC Content? A Lesson from

Teleostean Fish: A Mini Review. *Fish Ecology*, 61–80.

24. Thai B.T., Si V.N., Phan P.D., and Austin C. M. (2007). Phylogenetic evaluation of subfamily classification of the Cyprinidae focusing on Vietnamese species. *Aquat Living Resour*, 20 (2), 143-153.

25. Sharma U., Singhal V., Gupta D., and Mohanty P.S. (2014). Phylogenetic analysis among Cyprinidae family using 16SrRNA. *International Journal of Fisheries and Aquatic Studies*.

26. Alam M.A., Parvez I., Ara Y., Khan M.N., Nehrin S., and Mahajebin T. (2021). Phylogenetic relations of the cyprinid fishes (Cyprinidae) in Bangladesh inferred from morphological traits and cytochrome b gene sequences. *AACL Bioflux*. 14 (3).

27. Nhat P.H., Trang V.T., Vinh T.V., Cong L.V., Huong T.T.M., and Tuan N.D. (2023). The genetic diversity of dusky labiate barbel fish *Semilabeo obscurus* in Thai Nguyen province based on Mitochondrial COI gene sequences. *TNU Journal of Science and Technology*, 228 (05), 213 - 221.

28. Ha T.T.T., Thuy N.T.T., and Van N.S. (2018). Identification of Fish in *Semilabeo* Genus Using Morphological Taxonomy and Molecular Biology Methods. *GenAqua*, 2 (1).

29. Ward R.D. (2012). FISH-BOL, a case study for DNA barcodes. *Methods Mol Biol*, 858, 423 - 439.

30. Paixão R.V., Ribolli J., and Zaniboni-Filho E. (2018). Genetic Variation of the Endangered Neotropical Catfish *Steindachneridion scriptum* (Siluriformes: Pimelodidae). *Frontiers in Genetics*, 9.

31. Pappalardo A. M., Federico C., Sabella G., Saccone S., & Ferrito V. (2015). A COI Nonsynonymous Mutation as Diagnostic Tool for Intraspecific Discrimination in the European Anchovy *Engraulis encrasicolus* (Linnaeus). *PLOS ONE*, 10 (11), e0143297.

32. Ghouri M. Z., Ismail M., Javed M. A., Khan S. H., Munawar N., Umar A. B., Ahmad A. (2020). Identification of Edible Fish Species of Pakistan Through DNA Barcoding. *Frontiers in Marine Science*, 7.