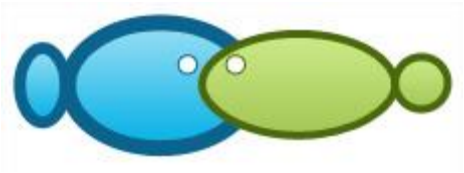

Phylogenetic and genetic diversity of Serranidae species from the South China Sea based on COI sequences

[1]Quyet T. Vu, [2]Van Q. Nguyen

[1] Joint Vietnam-Russia Tropical Science and Technology Research Center, Cau Giay, Ha Noi, Vietnam; [2] Institute of Marine Environment and Resources, Vietnam Academy of Science and Technology, Hanoi, Vietnam. Corresponding author: Q. T. Vu, vuquyetthanh@gmail.com

Abstract. This study explores the genetic diversity and phylogenetic relationships of economically important grouper species in Vietnam's coastal regions. Using mitochondrial COI gene sequencing, we identified high genetic diversity in the Epinephelinae sub-family. Analysis of 26 grouper samples from *Epinephelus*, *Cephalopholis*, and *Diploprion* revealed significant interspecies divergence, especially between genera, highlighting distinct evolutionary lineages. *Epinephelus* and *Cephalopholis* species showed closer genetic affinity within genera, while *Diploprion* displayed unique genetic characteristics. Phylogenetic analysis confirmed distinct clades for each genus, elucidating evolutionary divergence and relationships. Minimal genetic differences were observed between species collected in Son Cha and Cat Ba in Vietnam, contributing to our understanding of genetic relationships and evolutionary processes in Epinephelinae, supporting taxonomic classification and conservation efforts.
Key Words: Epinephelinae, geographic differentiation, grouper species, Vietnam.

Introduction. Grouper designates a group of fish predominantly found in coral reef ecosystems and shallow rocky waters. Certain species thrive in shallower waters with sandy mud substrates, contributing to the overall diversity of grouper habitats (Andriyono et al 2020). Grouper holds significant commercial importance in the global marine fish market, driven by its delectable taste and high demand. Its suitability for intensive aquaculture is underscored by its rapid growth, resilience to environmental stress, and efficient feed conversion (Craig & Hastings 2007). The appeal of grouper in the aquaculture sector is primarily attributed to their impressive growth potential (Heemstra & Randall 1993; Tupper & Sheriff 2008). As a result, grouper farming has seen a substantial increase in popularity, with production rates rising by 8-16% since the 1900s, spanning both floating net systems and ponds (Pomeroy 2002). This heightened demand for grouper, whether for consumption or ornamental purposes, has driven market prices upward and stimulated export-oriented initiatives (Halim 2001). Notably, among the grouper species that hold significant importance in both capture and aquaculture are *Cephalopholis boenak*, *Cephalopholis cyanostigma*, *Cephalopholis formosa*, *Cephalopholis sonnerati*, *Diploprion bifasciatum*, *Epinephelus areolatus*, *Epinephelus bleekeri*, *Epinephelus corallicola*, *Epinephelus fuscoguttatus*, *Epinephelus quoyanus*, *Epinephelus sexfasciatus*, and *Plectropomus maculatus* (Le et al 2011; Tran et al 2022). These carnivorous species primarily feed on small fish and crustaceans (Nagelkerken 1979) and are characterized by slow growth rates and long lifespans, which render them relatively immobile for extended periods and susceptible to overexploitation (Zhu & Yue 2008).

In the realm of genetic analysis, the mitochondrial DNA (mtDNA) genome is notable for its evolutionary conservation across various animal lineages, limited duplications, rapid mutational rate, and ease of isolation. Specifically, the COI gene, due to its moderate evolutionary pace compared to other mtDNA genes, serves as a valuable

marker for intraspecies genetic investigations (Desalle et al 2017; Zhang et al 2020). Moreover, COI gene sequences offer substantial and informative data for conducting phylogenetic analyses (Halasan et al 2021). Within the mtDNA genes, COI has been extensively employed for taxonomic and phylogenetic studies at both the species and family levels (Kumar et al 2018). It provides essential insights into population structure and facilitates intraspecific phylogenetic examinations (Sharina & Kartavtsev 2010; Shamblin et al 2015; Halasan et al 2021).

Despite numerous phylogenetic investigations into grouper species within the subfamily Epinephelinae, our comprehension of their phylogenetic relationships and species identification remains limited and beset with challenges. For example, Craig & Hastings (2007) explored closely related genera, *Epinephelus* and *Mycteroperca*, in the Eastern Atlantic Seas using mitochondrial cytochrome b (cyt b) and 16S ribosomal DNA. Their findings indicated minimal genetic distinctions between these two genera. Similarly, Ding et al (2006) conducted a phylogenetic analysis involving 28 species across six genera within the subfamily Epinephelinae, with a primary focus on the cyt b gene. Their results revealed that *Promicrops lanceolatus* and *Cromileptes altivelis* clustered within the *Epinephelus* clade, suggesting a reclassification of both species under the genus *Epinephelus*. Furthermore, Craig & Hastings (2007) conducted a comprehensive study utilizing multiple genes (16S, 12S, TMO-4C4, and histone H3) to construct a molecular phylogeny, leading to proposed revisions in the classification of certain Epinephelinae species. In parallel, Baharum & Nurdalila (2011) delved into the phylogenetic relationships of groupers from Malaysia, focusing on the cyt b gene, which revealed very minimal genetic differences between *Epinephelus fuscoguttatus* and *Epinephelus hexagonatus*, suggesting the intriguing possibility of conspecificity.

The lack of genetic diversity information about grouper species in Vietnam, especially along the eastern coast, exacerbates the challenge of selecting elite species for grouper breeding programs. Many grouper species, particularly those of the *Epinephelus* genus, are distributed along the Vietnamese coast (Tran et al 2022). However, comprehensive examinations of the phylogenetics of these frequently traded species in Vietnam remain insufficient. As a result, the current research endeavors to employ mitochondrial COI to elucidate the genetic affiliations among economically significant grouper species residing in the coastal areas of Vietnam. Two prolific grouper harvesting sites in Vietnam were selected for this study. One site, Cat Ba (CB), represents the northern region, while Son Cha (SC) is located on the boundary between the northern and central regions.

Material and Method

Sample collection. A total of 26 specimens belonging to the sub-family Epinephelinae were collected from two locations: Cat Ba and Son Cha (Table 1 and Figure 1).

Table 1

The number of individuals and their locality

No	*Name*	*SC*	*CB*	*n*
1	*Cephalopholis boenak*	2	2	4
2	*Cephalopholis cyanostigma*	2	0	2
3	*Cephalopholis formosa*	1	1	2
4	*Cephalopholis sonnerati*	2	0	2
5	*Diploprion bifasciatum*	2	2	4
6	*Epinephelus areolatus*	2	0	2
7	*Epinephelus bleekeri*	0	2	2
8	*Epinephelus corallicola*	2	0	2
9	*Epinephelus fuscoguttatus*	0	2	2
10	*Epinephelus quoyanus*	0	2	2
11	*Epinephelus sexfasciatus*	2	0	2
	Total	15	11	26

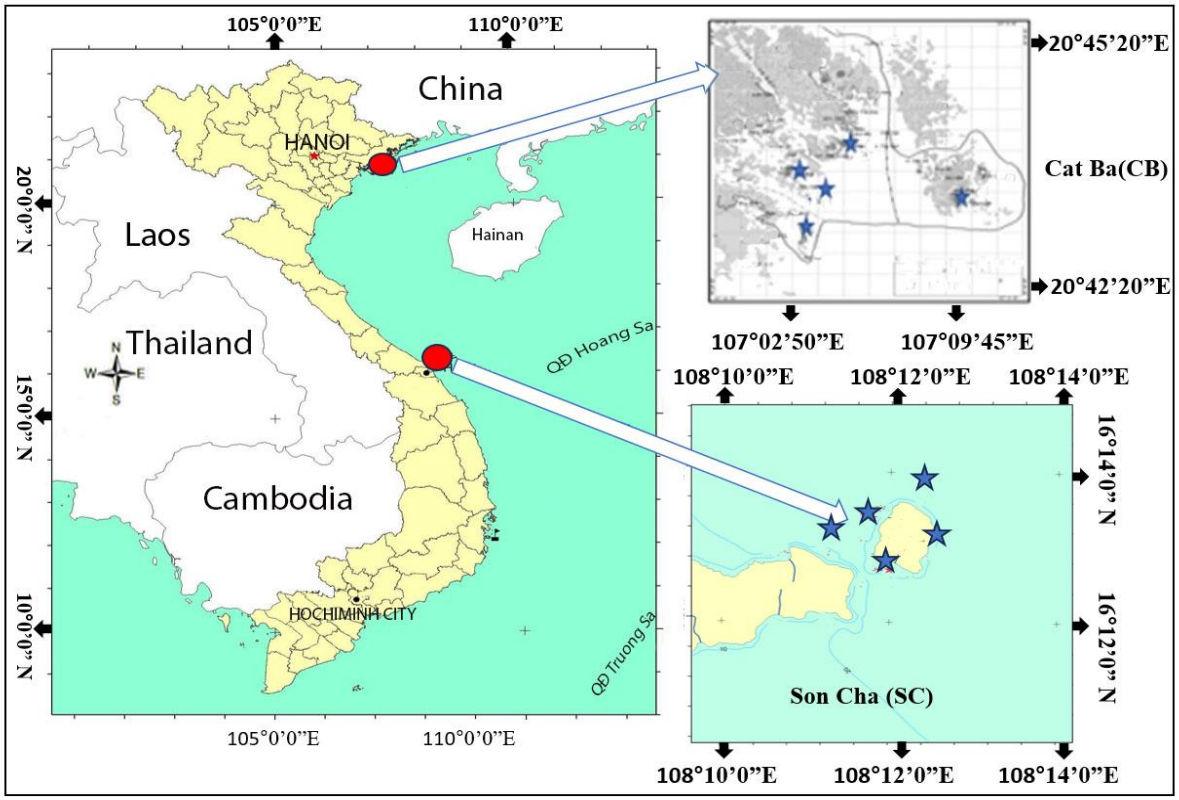

Figure 1. Sampling locations of Epinephelinae (*Epinephelus*, *Cephalopholis*, and *Diploprion*) in Cat Ba (CB) and Son Cha Lang Co (SC) in Vietnam. Note: the blue stars indicate specific fishing points.

The sampling was carried out over five distinct trips during May and June 2024. Specimens were captured using a trammel net with an inner mesh size of 20 mm and outer mesh size of 60 mm to ensure the appropriate selection of individuals of various sizes. After collection, all samples were either preserved in 95% ethanol or frozen at -10°C to maintain sample integrity for subsequent DNA extraction.

Morphologically identification was conducted according to the guideline from FAO (Heemstra & Randall 1993), and species confirmation has been carried out with molecular identification carried out in this study using the COI gene region.

DNA extraction, amplification, cloning, and sequencing. DNA was extracted from a tissue specimen with a DNeasy Blood and Tissue kit (Qiagen). The COI gene sequence was amplified using the universal primers L1490 and H2198 (Folmer et al 1994). PCR reaction was performed in a 24 µL volume: 2 µL of genomic DNA (total 40 ng), 12.5 µL master mix 2X (Thermo Scientific) and 9.5 µL deionized H_2O. The PCR amplifications were carried out on a Mastercycler X50s Eppendorf, under the following thermal cycler: initial denaturation at 94°C for 2 min, followed by 40 cycles consisting of 1 min at 94°C for denaturation, 30 s alignment at the annealing temperature for each primer pair at 56°C and 1 min at 72°C for extension and 10 min at 72°C for the final cycle to complete the extension of any remaining products before holding the samples at 4°C until they were analyzed. PCR amplicons were sent to Genlab (Hanoi, Vietnam) for bidirectional sequencing using the mentioned respective PCR primers.

Phylogenetic analysis. All sequences of the species were subjected to alignment in R programming language. The pairwise evolutionary distances within the family were determined using the p-distance method. Subsequently, a Neighbor-joining (NJ) tree was constructed, and a bootstrap analysis comprising 1000 replicates was performed using the Mega11 program (Tamura et al 2021).

Diversity of genes and haplotype network. The COI sequences were edited and aligned using MEGA v.11 (Tamura et al 2021) and the R programming language. The DnaSP v6.0 software (Rozas et al 2017) was employed to calculate variable sites, nucleotide composition, and assess the diversity of mitochondrial DNA. This assessment encompassed parameters such as Polymorphic Sites (S), Total Mutations (Eta), Haplotypes (h), Haplotype Diversity (Hd), Nucleotide Diversity (Pi), Theta (per site) derived from Eta, Theta (per site) derived from S (Theta-W), Average Nucleotide Differences (k), Stochastic Variance of k (No Recombination), Stochastic Variance of k (Free Recombination), and Theta (per sequence) derived from S, as well as Theta-W. The MEGA version 11 program (Tamura et al 2021) was used to compute the Maximum Likelihood Estimate of the Substitution Matrix and the count of polymorphic sites (ps). Genetic distances within and between species of the sub-family Epinephelinae from CB, SC were computed using the Tamura-Nei model for the COI dataset in MEGA 11.

Results and Discussion. A total of 26 successful identification samples encompassed three genera: *Epinephelus* (12), *Cephalopholis* (10), and *Diploprion* (4). In this study, the genus *Epinephelus* predominated in the overall fish catches, including species such as *E. areolatus*, *E. bleekeri*, *E. corallicola*, *E. fuscoguttatus*, *E. quoyanus*, and *E. sexfasciatus*. Regarding the sequence results, 583 base pairs of COI sequences were aligned, arranged in a horizontal manner, and visually represented using the R software (Figure 2).

Table 2 provides a comprehensive overview of genetic diversity metrics within the three distinct genera - *Epinephelus*, *Cephalopholis*, and *Diploprion* - as well as for the combined dataset of all genera. These metrics offer insights into the genetic variation, haplotype diversity, and evolutionary dynamics among the examined sequences, shedding light on the genetic intricacies within the grouper sub-family Epinephelinae (Table 2).

Table 2

Genetic diversity metrics for COI sequences in three genera - *Epinephelus*, *Cephalopholis*, and *Diploprion*

Genus	*Epinephelus*	*Cephalopholis*	*Diploprion*	*Genera*
Number of sequences	12	10	4	26
Total sites	583	583	583	583
Polymorphic sites (S)	171	124	1	213
Total mutations (Eta)	209	137	1	316
Haplotypes (h)	10	9	2	21
Haplotype diversity (Hd)	0.970	0.978	0.5	0.982
Nucleotide diversity (Pi)	0.125	0.091	0.001	0.154
Theta (per site) from Eta	0.119	0.083	0.001	0.142
Theta (per site) from S (Theta-W)	0.097	0.075	0.001	0.096
Average nucleotide differences (k)	73.136	52.956	0.5	90.058
Stochastic variance of k (no recombination)	956.989	501.135	0.162	1476.51
Stochastic variance of k (free recombination)	24.379	17.652	0.167	30.019
Theta (per sequence) from S, Theta-W	56.625	43.832	0.545	55.818

Figure 2. Alignment of 26 samples originating from three genera, namely *Epinephelus*, *Cephalopholis*, and *Diploprion*.

Note: The alignment highlights regions of both congruence and divergence within the COI sequences across these genera. Subsequent analyses may elucidate whether specific sequence disparities can be attributed to distinct genera. 1-SC159-1-*Epinephelus corallicola*, 2 - SC159-*Epinephelus corallicola*, 3 - SC158-1a-*Cephalopholis sonnerati*, 4 - SC158-1-*Cephalopholis sonnerati*, 5 - SC58-3-*Diploprion bifasciatum*, 6 - SC58-22-*Diploprion bifasciatum*, 7 - SCG-*Epinephelus sexfasciatus*, 8 - SCSCG2-*Epinephelus sexfasciatus*, 9 - SC164-g-*Cephalopholis cyanostigma*, 10 - SC164-g1-*Cephalopholis cyanostigma*, 11 - SC7172144G-*Epinephelus areolatus*, 12 - SC7172144G2-*Epinephelus areolatus*, 13 - SC61-3149-*Cephalopholis boenak*, 14 - SC61-1-*Cephalopholis boenak*, 15 - SC57-*Cephalopholis formosa*, 16- CB57- *Cephalopholis formosa*, 17 - CB22-*Diploprion bifasciatum*, 18 - CB22a-*Diploprion bifasciatum*, 19 - CB21-*Epinephelus fuscoguttatus*, 20 - CB21a-*Epinephelus fuscoguttatus*, 21 - CB24C-*Cephalopholis boenak*, 22 - CB04d-*Cephalopholis boenak*, 23 - CB03-*Epinephelus quoyanus*, 24 - CB03-21-*Epinephelus quoyanus*, 25 -CB02-*Epinephelus bleekeri*, 26 - CB02-21-*Epinephelus bleekeri*.

Epinephelus. *Epinephelus* demonstrates a significant level of genetic diversity, as indicated by its 171 polymorphic sites (S). The presence of 209 total mutations (Eta) further supports the species' genetic variation. With 10 haplotypes (h), this species showcases a diverse gene pool. The high haplotype diversity (Hd = 0.970) and nucleotide diversity (Pi = 0.125) underscore the rich genetic variation within this species. The theta values (Theta per site from Eta: 0.118 and Theta per site from S - Theta-W: 0.097) validate the observed genetic diversity. The average nucleotide differences (k) of 73.136 highlight substantial genetic variability. The stochastic variances of k, both with and without recombination, support the robust diversity within this species (Table 2).

Cephalopholis. *Cephalopholis* also displays notable genetic diversity, with 124 polymorphic sites (S) indicating significant genetic variation. The presence of 137 total mutations (Eta) further confirms genetic diversity. The species exhibits 9 haplotypes (h), reflecting a diverse gene pool. A high haplotype diversity (Hd = 0.978) and nucleotide diversity (Pi = 0.09083) underscore genetic variation. The theta values (Theta per site from Eta: 0.08307 and Theta per site from S - Theta-W: 0.07518) validate the genetic diversity observed. The average nucleotide differences (k) of 52.956 suggest substantial genetic variability, which is supported by stochastic variances of k, both with and without recombination (Table 2).

Diploprion. *Diploprion* exhibits limited genetic diversity with only 1 polymorphic site (S) and a single total mutation (Eta). The presence of just 2 haplotypes (h) suggests a less diverse gene pool. A low haplotype diversity (Hd = 0.500) and extremely low nucleotide diversity (Pi = 0.0008) reflect the limited genetic variation. Theta values (Theta per site from Eta: 0.00094 and Theta per site from S - Theta-W: 0.00094) are low, indicating low genetic diversity. The average nucleotide differences (k) of 0.500 confirm the limited genetic variability. Stochastic variances of k, both with and without recombination, support the low genetic diversity within this species (Table 2).

All genera (combined). When considering all genera together, the analysis reveals a high overall genetic diversity across the entire grouper sub-family Epinephelinae. The presence of 213 polymorphic sites (S) and 316 total mutations (Eta) indicates significant genetic variation within the sub-family. Twenty-one haplotypes (h) suggest a diverse gene pool. A high haplotype diversity (Hd = 0.982) and nucleotide diversity (Pi = 0.15447) reflect the rich genetic variation across the sub-family. Theta values (Theta per site from Eta: 0.14204 and Theta per site from S - Theta-W: 0.09574) validate the genetic diversity observed. The average nucleotide differences (k) of 90.058 highlight substantial genetic variability, which is further supported by stochastic variances of k, both with and without recombination. Theta values per sequence (Theta per sequence from S - Theta-W: 55.818) indicate genetic diversity on a per-sequence level (Table 2).

These comprehensive observations provide insights into the intricate genetic dynamics within each species and across the sub-family, offering crucial information for understanding their evolutionary history and conservation strategies.

In this haplotype network analysis, we explored the genetic diversity and evolutionary relationships among three closely related fish genera: *Epinephelus*, *Cephalopholis*, and *Diploprion*. Our investigation revealed varying degrees of haplotype diversity within each genus, with distinct haplotypes representing unique genetic lineages. The network structures exhibited interesting patterns, including the presence of shared haplotypes among species and isolated haplotypes, suggesting complex evolutionary histories. Genetic distance calculations provided insights into the genetic differentiation between haplotypes, while geographic distribution analysis highlighted potential patterns of isolation or connectivity. Overall, this analysis offers a glimpse into the intricate genetic relationships and historical processes that have shaped the genetic diversity of these fish genera. Significantly, during the haplotype network analysis, we observed a clear branching pattern in *Diploprion* (Figure 3). This indicates genetic divergence and diversity within this species, possibly as a result of long-term evolutionary processes or geographic events. However, it is noteworthy that *Diploprion*

and *Epinephelus* have not exhibited distinct branching patterns within the network. The absence of pronounced branching may be linked to their unique evolutionary histories or current population sizes.

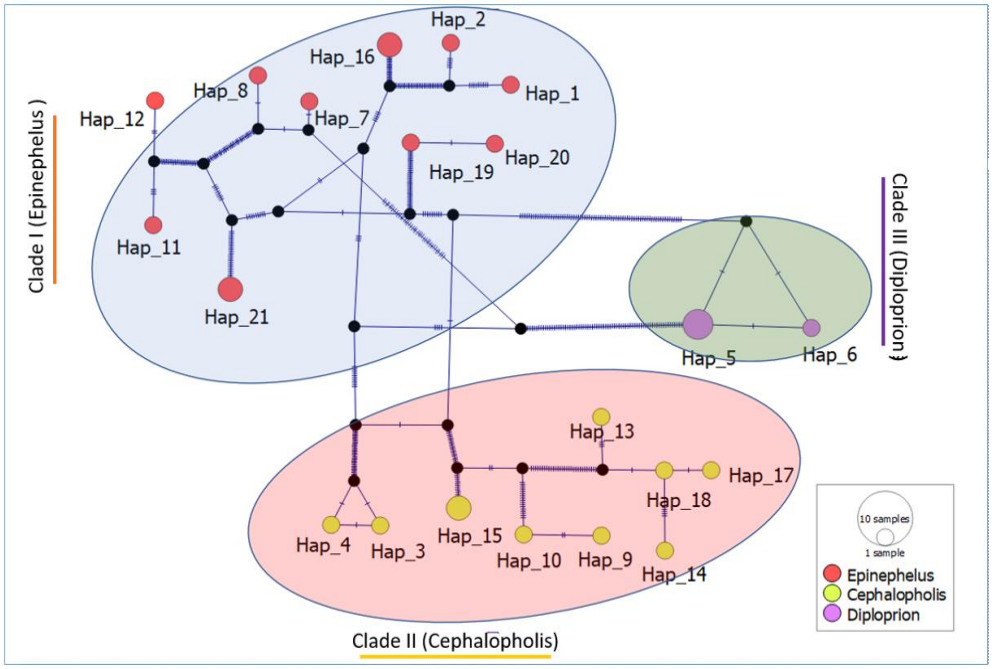

Figure 3. Haplotype network of 26 samples from three genera: *Epinephelus*, *Cephalopholis*, and *Diploprion*.

Estimate of the pattern of nucleotide substitution. The nucleotide frequencies are 24.52% (A), 30.35% (T), 27.69% (C), and 17.44% (G). This analysis involved 26 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 583 positions in the final dataset. In the study by Nor Rahim (2016), sequence analysis involved aligning a total of 1857 bp from combined cyt b (750 bp) and RAG-1 (1107 bp) gene sequences. This alignment revealed 1404 conserved sites, 453 variable sites, and 383 parsimony informative sites. The mean nucleotide composition among the species was 24.7% for T, 28.1% for C, 24.3% for G, and 22.91% for A.

In this scientific investigation, we examine nucleotide substitution patterns within the DNA of the Epinephelinae sub-family, encompassing three primary genera: *Epinephelus*, *Cephalopholis*, and *Diploprion*. Employing the Maximum Composite Likelihood Estimate method, we calculate probabilities of nucleotide substitutions among the A, T, C, and G bases. The results provide a comprehensive perspective on these substitution patterns, shedding light on the diversity and evolutionary dynamics that have shaped the genomes of Epinephelinae species. Our analysis, based on a dataset of 26 nucleotide sequences with a total of 583 positions, reveals significant variations in substitution probabilities between different nucleotide bases. Notably, we observe elevated substitution rates between G and A (13.25 and 9.42, respectively), suggesting a unique genetic interaction during the evolutionary trajectory of Epinephelinae species. Furthermore, our findings highlight pronounced substitution rates between T and C (21.02 and 23.04, respectively), indicating a noteworthy genetic relationship between these nucleotide bases. This study offers insights into the evolutionary processes and genetic interactions within the Epinephelinae sub-family (Table 3). In practical terms, it is noteworthy that there are instances of hybridization and even cases of misidentification of hybrid individuals, despite DNA analysis. Understanding these nucleotide substitution patterns is pivotal for advancing our knowledge of the evolutionary mechanisms within the diverse Epinephelinae sub-family, opening up new avenues for future research into their genetic diversity and evolution.

Table 3

Maximum composite likelihood estimate of the pattern of nucleotide substitution

	A	*T*	*C*	*G*
A	-	*5.05*	*4.61*	**9.42**
T	*4.08*	-	**21.02**	*2.9*
C	*4.08*	**23.04**	-	*2.9*
G	**13.25**	*5.05*	*4.61*	-

Note: Each entry shows the probability of substitution (r) from one base (row) to another base (column). For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in **bold** and those of transversionsal substitutions are shown in *italics*. This analysis involved 26 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 583 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 (Tamura et al 2021).

Based on nucleotide sequence analysis of 26 Serranidae samples across 3 genera (*Epinephelus*, *Cephalopholis*, and *Diploprion*), several key observations emerge. First, there is sequence variation within species at a low level (≤ 0.002, Table 5). Second, notable sequence divergence exists among different species (0.024-0.220, Table 4, Table 5), reflecting genetic differences and supporting phylogenetic tree branching (Figure 4). The highest genetic distance (0.22) between *E. corallicola* and *D. bifasciatum* indicates substantial divergence between these genera. Third, within *Cephalopholis*, sequence differences among species (0.089-0.210) are lower than those between *Cephalopholis* and *Epinephelus* (0.120-0.210, Table 4, Table 5). Fourth, we observed minimal genetic differences between the species collected in SC and CB (≤ 0.002). Fifth, within each genus (*Epinephelus*, *Cephalopholis*, and *Diploprion*), specific remarks can be made: Epinephelus species showed sequence differences (0.135-0.186); *E. corallicola*, *E. fuscoguttatus*, and *E. bleekeri* exhibited high similarities, while *E. areolatus* and *E. quoyanus* showed greater divergence. In *Cephalopholis*, interspecies sequence variation ranged from 0.089 to 0.211; *C. boenak* and *C. cyanostigma* displayed high genetic resemblance, with *C. formosa* showing larger differences. For *Diploprion*, only one species was studied; *D. bifasciatum* samples showed a sequence difference of 0.002. In summary, these findings illuminate genetic variations at species and genus levels within the studied Serranidae, providing insights into their genetic evolution. Sequence analysis forms a molecular basis for further exploring evolutionary relationships and taxonomic classification.

The bootstrap tree has relatively high reliability, mostly with bootstrap values above 70 (18/26, 69.23%); nodes with a value of 100 are 13/26 (50%); The lowest bootstrap value is 24. The phylogenetic tree is divided into three distinct branches representing three genera: *Epinephelus*, *Cephalopholis*, *Diploprion*.

The primary clade comprised sequences from species within the *Epinephelus* genus, including *E. areolatus*, *E. bleekeri*, and *E. quoyanus*, whereas the secondary clade exclusively featured sequences from two species, *E. fuscoguttatus* and *E. corallicola*. Notably, *E. sexfasciatus* formed a distinct branch within this phylogenetic analysis.

In the case of the *Cephalopholis* genus, the main clade included sequences from species such as *C. sonnerati*, *C. formosa*, and *C. cyanostigma*, with *C. boenak* forming a separate and distinct branch within the phylogenetic analysis.

Diploprion forms a distinct branch separated from the remaining branches consisting of the two genera *Epinephelus* and *Cephalopholis*.

Epinephelus demonstrates substantial genetic diversity, characterized by 171 polymorphic sites (S) and 209 total mutations (Eta), indicating a diverse gene pool with 10 haplotypes (h). High haplotype diversity (Hd = 0.970) and nucleotide diversity (Pi = 0.12545) underscore rich genetic variation. *Cephalopholis* exhibits notable genetic diversity, featuring 124 polymorphic sites (S) and 137 total mutations (Eta), with 9 haplotypes, and high Hd (0.978) and Pi (0.09083). In contrast, *Diploprion* displays limited diversity, with only 1 polymorphic site (S) and 2 haplotypes (h), Hd (0.500), and Pi (0.0008). However, due to the study's focus on a single species, definitive conclusions regarding genetic diversity are not possible. The combined analysis of Epinephelinae

demonstrates extensive genetic diversity with 213 S, 316 Eta, 21 haplotypes (h), Hd (0.982), Pi (0.15447), and substantial genetic variability (k = 90.058). These findings significantly contribute to our understanding of evolutionary dynamics and inform conservation strategies.

In investigating genetic differentiation based on geographical regions of sample collection, we observed minimal differences between the species collected in SC and CB. The SC group exhibited a broader range of maximum sequence divergence from 0.158 to 0.220, while the CB group displayed a range from 0.108 to 0.220, both showing minimal within-species differences of 0.002. According to principles of climatic hydrology, ecological biogeography, and terrestrial-aquatic interactions, the Vietnamese coastal region is categorized into three distinct coastal zones. The Northern zone, spanning from Mong Cai (21°29'08.1"N, 108°03'50.4"E) to Hai Van Cape (16°11'22.6"N, 108°07'46.6"E), experiences a tropical monsoon climate with cold winters. The term "CB" refers to this particular zone. The intermediate zone, from Hai Van Cape to Dai Lanh (12°50'08.0"N, 109°22'07.0"E), represents a convergence of the subequatorial monsoon climate of southern Vietnam and the marine climate of the Eastern Sea, following a tropical monsoon regime. SC locality is situated at the boundary between the northern and central regions. While we observed no significant differences in genetic distances between these two locales, this may still signal initial signs of ecological group differentiation. In order to clarify this observation regarding genetic differentiation based on geographical regions of sample collection, further investigation is required in southern regions. Moreover, the coastal area designated as CB, with its distinct monsoonal climate and isolation due to numerous small islands, likely facilitated the emergence of genetically distinct populations.

Table 4

The summary table of estimates of evolutionary divergence between sequences

STT	*Species*	*Max*	*Min*	*Local*
1	SC159-1-*Epinephelus corallicola*	0.220	0.024	SC
2	SC159-*Epinephelus corallicola*	0.024	0.024	SC
3	SC158-1a-*Cephalopholis sonnerati*	0.158	0.002	SC
4	SC158-1-*Cephalopholis sonnerati*	0.158	0.002	SC
5	SC58-3-*Diploprion bifasciatum*	0.220	0.002	SC
6	SC58-22-*Diploprion bifasciatum*	0.220	0.002	SC
7	SCG-*Epinephelus sexfasciatus*	0.201	0.005	SC
8	SCSCG2-*Epinephelus sexfasciatus*	0.202	0.005	SC
9	SC164-g-*Cephalopholis cyanostigma*	0.206	0.003	SC
10	SC164-g1-*Cephalopholis cyanostigma*	0.202	0.003	SC
11	SC7172144G-*Epinephelus areolatus*	0.209	0.009	SC
12	SC7172144G2-*Epinephelus areolatus*	0.208	0.009	SC
13	SC61-3149-*Cephalopholis boenak*	0.211	0.086	SC
14	SC61-1-*Cephalopholis boenak*	0.206	0.015	SC
15	SC57-*Cephalopholis formosa*	0.201	0.063	SC
16	CB57-*Cephalopholis formosa*	0.201	0.063	CB
17	CB22-*Diploprion bifasciatum*	0.220	0.002	CB
18	CB22a-*Diploprion bifasciatum*	0.220	0.002	CB
19	CB21-*Epinephelus fuscoguttatus*	0.201	0.118	CB
20	CB21a-*Epinephelus fuscoguttatus*	0.201	0.118	CB
21	CB24C-*Cephalopholis boenak*	0.206	0.009	CB
22	CB04d-*Cephalopholis boenak*	0.206	0.002	CB
23	CB03-*Epinephelus quoyanus*	0.213	0.139	CB
24	CB03-21-*Epinephelus quoyanus*	0.214	0.002	CB
25	CB02-*Epinephelus bleekeri*	0.199	0.108	CB
26	CB02-21-*Epinephelus bleekeri*	0.199	0.108	CB

Note: The number of base differences per site between sequences is shown. This analysis involved 26 nucleotide sequences. All ambiguous positions were removed for each sequence pair (using the pairwise deletion option). There were a total of 583 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 (Tamura et al 2021). Min: the smallest value among different species (excluding identical species).

Table 5

The mean distance between the groups for all eight species analysed under 1000 bootstrap replication. The analyses were conducted using Maximum Composite Likehood model (Kimura 1980) with rate variation of gamma distribution

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	0																									
2	0.024																									
3	0.158	0.158																								
4	0.158	0.158	0.002																							
5	0.220	0.214	0.199	0.199																						
6	0.220	0.214	0.199	0.199	0.002																					
7	0.142	0.141	0.160	0.160	0.199	0.201																				
8	0.144	0.139	0.160	0.160	0.201	0.202	0.005																			
9	0.180	0.180	0.146	0.146	0.206	0.206	0.166	0.165																		
10	0.177	0.177	0.142	0.142	0.202	0.202	0.163	0.161	0.003																	
11	0.158	0.151	0.173	0.172	0.209	0.209	0.141	0.139	0.175	0.172																
12	0.158	0.151	0.170	0.168	0.208	0.208	0.141	0.139	0.172	0.168	0.009															
13	0.163	0.166	0.153	0.153	0.211	0.209	0.163	0.165	0.089	0.086	0.182	0.180														
14	0.166	0.170	0.153	0.153	0.206	0.204	0.163	0.165	0.091	0.087	0.178	0.177	0.015													
15	0.168	0.170	0.141	0.141	0.201	0.201	0.163	0.161	0.067	0.063	0.153	0.149	0.091	0.093												
16	0.168	0.170	0.141	0.141	0.201	0.201	0.163	0.161	0.067	0.063	0.153	0.149	0.091	0.093	0.000											
17	0.220	0.214	0.199	0.199	0.000	0.002	0.199	0.201	0.206	0.202	0.209	0.208	0.211	0.206	0.201	0.201										
18	0.220	0.214	0.199	0.199	0.000	0.002	0.199	0.201	0.206	0.202	0.209	0.208	0.211	0.206	0.201	0.201	0.000									
19	0.123	0.118	0.148	0.148	0.201	0.201	0.136	0.137	0.161	0.158	0.144	0.144	0.163	0.158	0.161	0.161	0.201	0.201								
20	0.123	0.118	0.148	0.148	0.201	0.201	0.136	0.137	0.161	0.158	0.144	0.144	0.163	0.158	0.161	0.161	0.201	0.201	0.000							
21	0.161	0.165	0.153	0.153	0.206	0.204	0.161	0.163	0.087	0.084	0.180	0.178	0.009	0.010	0.089	0.089	0.206	0.206	0.158	0.158						
22	0.163	0.166	0.151	0.151	0.206	0.204	0.160	0.161	0.086	0.082	0.178	0.177	0.007	0.009	0.087	0.087	0.206	0.206	0.160	0.160	0.002					
23	0.161	0.154	0.153	0.153	0.213	0.213	0.153	0.151	0.175	0.172	0.144	0.141	0.177	0.175	0.175	0.175	0.213	0.213	0.139	0.139	0.175	0.173				
24	0.163	0.156	0.154	0.154	0.214	0.214	0.154	0.153	0.177	0.173	0.146	0.142	0.178	0.177	0.177	0.177	0.214	0.214	0.141	0.141	0.177	0.175	0.002			
25	0.144	0.137	0.154	0.154	0.199	0.199	0.137	0.136	0.161	0.158	0.110	0.108	0.166	0.166	0.163	0.163	0.199	0.199	0.115	0.115	0.165	0.163	0.127	0.129		
26	0.144	0.137	0.154	0.154	0.199	0.199	0.137	0.136	0.161	0.158	0.110	0.108	0.166	0.166	0.163	0.163	0.199	0.199	0.115	0.115	0.165	0.163	0.127	0.129	0.000	

Note: 1-SC159-1-*Epinephelus corallicola*, 2 - SC159-*Epinephelus corallicola*, 3 - SC158-1a-*Cephalopholis sonnerati*, 4 - SC158-1-*Cephalopholis sonnerati*, 5 - SC58-3-*Diploprion bifasciatum*, 6 - SC58-22-*Diploprion bifasciatum*, 7 - SCG-*Epinephelus sexfasciatus*, 8 - SCSCG2-*Epinephelus sexfasciatus*, 9 - SC164-g-*Cephalopholis cyanostigma*, 10 - SC164-g1-*Cephalopholis cyanostigma*, 11 - SC7172144G-*Epinephelus areolatus*, 12 - SC7172144G2-*Epinephelus areolatus*, 13 - SC61-3149-*Cephalopholis boenak*, 14 - SC61-1-*Cephalopholis boenak*, 15 - SC57-*Cephalopholis formosa*, 16- CB57- *Cephalopholis formosa*, 17 - CB22-*Diploprion bifasciatum*, 18 - CB22a-*Diploprion bifasciatum*, 19 - CB21-*Epinephelus fuscoguttatus*, 20 - CB21a-*Epinephelus fuscoguttatus*, 21 - CB24C-*Cephalopholis boenak*, 22 - CB04d-*Cephalopholis boenak*, 23 - CB03-*Epinephelus quoyanus*, 24 - CB03-21-*Epinephelus quoyanus*, 25 -CB02-*Epinephelus bleekeri*, 26 - CB02-21-*Epinephelus bleekeri*.

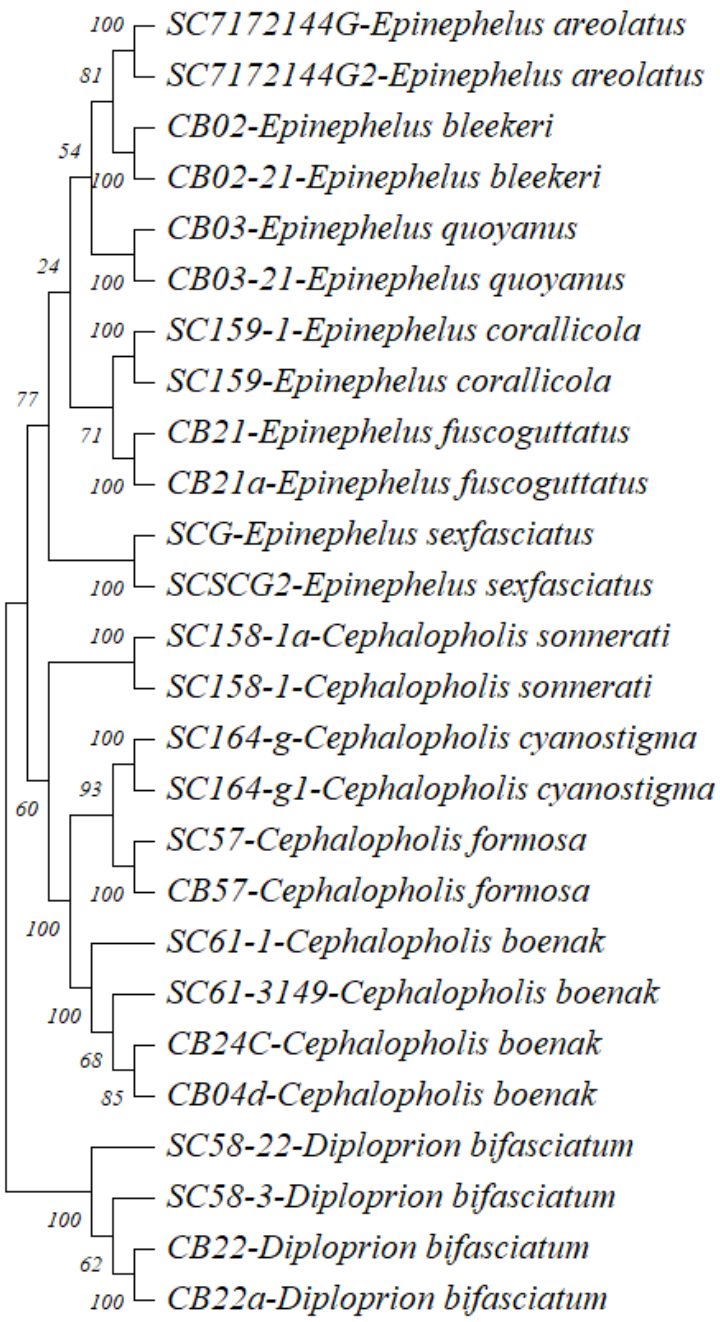

Figure 4. NJ tree on the concatenated dataset (COI) used in the present study. Bootstrap (1000) values were indicated for each nodes.

A study by Vu et al (2022) examined *Johnius carouna* otoliths and revealed significant morphological differences among populations in north and south regions, attributed to variations in geographical location, environmental conditions, and climate. Genetic drift emerges as an additional factor driving genetic differentiation. This understanding could serve as a basis for further molecular and population studies, providing important information for the management and conservation of grouper resources in Vietnam. To gain a deeper understanding of the evolutionary process and phylogenetic relationships of these species, more complex molecular evolutionary studies are needed, incorporating ecological and geographical factors to explain their genetic divergence.

This phylogenetic tree exhibits similarities to the phylogenetic tree in a previous study conducted in Indonesia, wherein the two genera, *Epinephelus* and *Cephalopholis*, demonstrated close proximity by sharing a common branch, distinct from the remaining branches. In this phylogenetic tree, a similar pattern emerges to the one observed in the

Indonesian study, as *E. corallicola* and *E. fuscoguttatus* are grouped together on the same branch, indicating genetic closeness. Additionally, *E. sexfasciatus* forms a separate branch, diverging from the other species within the *Epinephelus* genus. Moreover, in line with a study conducted by Hassanien & Al-Rashada (2021), our findings corroborate the close genetic relationship between *E. bleekeri* and *E. areolatus*, as compared to the other species under investigation. These findings align with earlier research by Craig et al (2001) and Ding et al (2006). Ding et al (2006) proposed the possibility of a divergence event within the *Epinephelus* genus resulting in the formation of two distinct clades during the early stages of evolution. However, it is imperative to conduct additional molecular evolutionary studies to address this issue conclusively.

Conclusions. Based on the analysis of base haplotypes, polymorphic sites (S), total mutations (Eta), haplotypes (h), haplotype diversity (Hd), nucleotide diversity (Pi), Theta (per site) derived from Eta, and Theta (per site) derived from S (Theta-W), considering the three genera (*Epinephelus*, *Cephalopholis*, *Diploprion*) collectively, our study reveals a notably high genetic diversity across the entire grouper sub-family Epinephelinae. *Epinephelus* and *Cephalopholis* exhibit high genetic diversity and haplotype diversity. Our haplotype network analysis of *Epinephelus*, *Cephalopholis*, and *Diploprion* revealed diverse haplotype distributions and genetic lineages. Phylogenetic classification analysis based on COI has revealed distinct separations among the genera *Epinephelus* and *Cephalopholis* compared to the genus *Diploprion*. In investigating genetic differentiation based on geographical regions of sample collection, we observed minimal differences between the species collected in SC and CB.

Acknowledgements. This work was funded by the grant in aids from Vietnam-Russia Tropical Center "ST.Đ1.09/23" and partially funded by the Vietnam Academy of Science and Technology's grant in aid (ĐL0000.01/23-24)

Conflict of interest. The authors declare that there is no conflict of interest.

References

Andriyono S., Damora A., Hidayani A. A., 2020 Genetic diversity and phylogenetic reconstruction of grouper (Serranidae) from Sunda Land, Indonesia. Egyptian Journal of Aquatic Biology and Fisheries 24(3):403-415.

Baharum S. N., Nurdalila A. A., 2011 Phylogenetic relationships of *Epinephelus fuscoguttatus* and *Epinephelus hexagonatus* inferred from mitochondrial cytochrome b gene sequences using bioinformatics tools. International Journal of Bioscience, Biochemicals and Bioinformatics 1(1):47-52.

Craig M. T., Hastings P. A., 2007 A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini. Ichthyological Research 54:1-17.

Craig M. T., Pondella II D. J., Franck J. P., Hafner J. C., 2001 On the status of the serranid fish genus *Epinephelus*: evidence for paraphyly based upon 16S rDNA sequence. Molecular Phylogenetics and Evolution 19(1):121-130.

Desalle R., Schierwater B., Hadrys H., 2017 MtDNA: the small workhorse of evolutionary studies. Frontiers in Bioscience 22(5):873-887.

Ding S., Zhuang X., Guo F., Wang J., Su Y., Zhang Q., Li Q., 2006 Molecular phylogenetic relationship of China Seas groupers based on cytochrome b gene fragment sequences. Science in China. Series C, Life Sciences 49(3):235-242.

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R., 1994 DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3(5):294-299.

Halasan L. C., Geraldino P. J. L., Lin H. C., 2021 First evidence of cryptic species diversity and population structuring of *Selaroides leptolepis* in the Tropical Western Pacific. Frontiers in Marine Science 8:756163.

Halim A., 2001 Grouper culture: an option for grouper management in Indonesia. Coastal Management 29(4):319-326.
Hassanien H. A., Al-Rashada Y., 2021 Assessment of genetic diversity and phylogenetic relationship among grouper species *Epinephelus spp.* from the Saudi waters of the Arabian Gulf. Saudi Journal of Biological Sciences 28(3):1779-1786.
Heemstra P. C., Randall J. E., 1993 FAO species catalogue. Vol. 16. Groupers of the world (Family Serranidae, Subfamily Epinephelinae). An annotated and illustrated catalogue of the grouper, rockcod, hind, coral grouper and lyretail species known to date. FAO Fisheries Synopsis No. 125, Vol. 16, FAO, Rome, 382 pp.
Kimura M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120.
Kumar S., Stecher G., Li M., Knyaz C., Tamura K., 2018 MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35(6):1547-1549.
Le T. T. T., Vo V. Q., Nguyen P. U. V., 2011 A check list of grouper species (Serranidae) in Vietnam sea. Proceedings of the National Conference on Marine Science and Technology (5th), Ha Noi, Vietnam, pp. 145-153.
Nagelkerken W. P., 1979 Biology of the graysby, *Epinephelus cruentatus*, of the coral reef of Curacao. Studies of the Fauna of Curacao and other Caribbean Islands 186:1-118.
Nor Rahim N. S., Esa Y., Arshad A., 2016 Molecular phylogeny of some Malaysian groupers (Subfamily: Epinephelinae, Family: Serranidae) inferred from mitochondrial and nuclear gene sequences. Journal of Sustainability Science and Management 11(2):1-10.
Pomeroy R. S., 2002 The status of grouper culture in Southeast Asia. SPC Live Reef Fish Information Bulletin 10:22-26.
Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J. C., Guirao-Rico S., Librado P., Ramos-Onsins S. E., Sánchez-Gracia A., 2017 DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution 34(12):3299-3302.
Shamblin B. M., Bagley D. A., Ehrhart L. M., Desjardin N. A., Martin R. E., Hart K. M., et al., 2015 Genetic structure of Florida green turtle rookeries as indicated by mitochondrial DNA control region sequences. Conservation Genetics 16:673-685.
Sharina S. N., Kartavtsev Y. P., 2010 Phylogenetic and taxonomic analysis of flatfish species (Teleostei, Pleuronectiformes) inferred from the primary nucleotide sequence of cytochrome oxidase 1 gene (Co-1). Russian Journal of Genetics 46: 356-361.
Tamura K., Stecher G., Kumar S., 2021 MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution 38(7):3022-3027.
Tran V. D., Vu Q. T., Tran C. T., Savinkin O. V., 2022 [Species diversity of grouper (Serranidae) and new records for the grouper checklist of Vietnamese waters]. Journal of Tropical Science and Technology 30:114-123. [in Vietnamese]
Tupper M., Sheriff N., 2008 Capture-based aquaculture of groupers. In: Capture-based aquaculture: global overview. Lovatelli A., Holthus P. F. (eds), FAO Fisheries Technical Paper No. 508, FAO, Rome, pp. 217-253.
Vu Q. T., Tran V. D., Nguyen V. Q., 2022 Otolith shape utilization for the stock identification of Caroun croaker, *Johnius carouna* (Cuvier, 1830) (Perciformes: Sciaenidae), from the east Vietnam sea. Iranian Journal of Fisheries Sciences 21(4):864-879.
Zhang Q., Sun C., Zhu Y., Xu N., Liu H., 2020 Genetic diversity and structure of the round-tailed paradise fish (*Macropodus ocellatus*): implications for population management. Global Ecology and Conservation 21:e00876.
Zhu Z. Y., Yue G. H., 2008 The complete mitochondrial genome of red grouper (*Plectropomus leopardus*) and its applications in identification of grouper species. Aquaculture 276(1-4):44-49.

Received: 04 July 2024. Accepted: 12 August 2024. Published online: 27 October 2024.
Authors:
Quyet Thanh Vu, Joint Vietnam-Russia Tropical Science and Technology Research Center, Cau Giay, 100000 Hanoi, Vietnam, e-mail: vuquyetthanh@gmail.com
Van Quan Nguyen, Institute of Marine Environment and Resources, Vietnam Academy of Science and Technology, Cau Giay, 100000 Hanoi, Vietnam, e-mail: fishhio@gmail.com

How to cite this article:
Vu Q. T., Nguyen V. Q., 2024 Phylogenetic and genetic diversity of Serranidae species from the South China Sea based on COI sequences. AACL Bioflux 17(5):2175-2188.