Population genetics of the black sharkminnow (*Labeo chrysophekadion*) in the lower Mekong Basin based on mitochondrial DNA segments from RAD-seq

Oanh Truong Thi[1,2,*], Sang Quang Tran[2], Van Ngo Thai Bich[1], Binh Thuy Dang[2,*]

[1]*University of Science and Technology, The University of Danang, 54 Nguyen Luong Bang, Da Nang, Vietnam*

[2]*Nha Trang University, 02 Nguyen Dinh Chieu, Khanh Hoa, Vietnam*

Correspondence

Oanh Truong Thi, University of Science and Technology, The University of Danang, 54 Nguyen Luong Bang, Da Nang, Vietnam

Nha Trang University, 02 Nguyen Dinh Chieu, Khanh Hoa, Vietnam

Email: oanhtt@ntu.edu.vn

Correspondence

Binh Thuy Dang, Nha Trang University, 02 Nguyen Dinh Chieu, Khanh Hoa, Vietnam

Email: binhdt@ntu.edu.vn

Check for updates

ABSTRACT

Background: The *Labeo chrysophekadion* is an economically important cyprinid that migrates short distances seasonally between the mainstream and floodplains of the Mekong River. However, wild populations are threatened by fishing pressure and environmental impacts. Genome-wide studies have been widely applied in molecular ecology to inform fisheries management. This study aimed to assemble and annotate the complete mitogenome from whole-genome restriction site-associated DNA sequencing data and use the assembled mitogenome to identify aligned mitogenome segments and investigate the population genetics of *L. chrysophekadion* in the lower Mekong Basin (LMB). **Methods:** A total of 255 individuals were collected in the lower Mekong Basin. There were six sites on the Mekong mainstem (Paksan, Pakse-Lao PDR, Ubon Ratchathani-Thailand, Kratié-Cambodia, Dong Thap, An Giang-Vietnam), one site at the confluence of the Mekong and 3S Rivers (Stung Treng-Cambodia), and two sites on tributaries in the LMB: the Khan River, Luang Prabang, Lao PDR, and the Chi River, Roi Et, Thailand. High-quality sequence reads were identified and used to assemble, annotate, and visualize the mitogenome using the Mitoz toolkit. Following the RADbarcoder pipeline, aligned DNA segments were identified and used to estimate the genetic diversity and haplotype network. **Results:** The complete mitogenome of *L. chrysophekadion* was 16,600 base pairs in length and exhibited a high identity of 99.8% to a previously published genome (accession number AP011199) derived from an individual fish in Kandal, Cambodia. This genome comprised 13 protein-coding genes, 22 transfer RNA genes, 2 ribosomal RNA genes, and a control region gene. When mapping individual sequence reads to the mitogenome, 757 bp were identified as the aligned mitogenome segment data. A total of 49 haplotypes from 247 individuals were detected, with a haplotype diversity of 0.849 (±0.014) and nucleotide diversity of 0.005 (±0.0008). High connectivity was detected among the sample populations, with three dominant common haplotypes shared among 8-9 populations. Additionally, nine haplotypes were shared by at least two populations, while 35 private haplotypes were distributed among all examined populations. This significant finding provides an overview of population structure, serving as a scientific basis for the conservation and management of aquatic resources. **Conclusion:** This study provides information on the mitogenomic characteristics and spatial genetics of *L. chrysophekadion* in the LMB.

Key words: Aigned mitogenome segments, mitogenome, haplotype network, population genetics, EzRAD

INTRODUCTION

The highly biodiverse Mekong River has been increasingly threatened by climate change and anthropogenic activities[1–3]. This river is characterized by fast-flowing rapids, waterfalls, confluences, tributaries, catchments, seasonal floodplains and deep pools[4]. Along with seasonal flow changes and dynamic hydrological features, 80% of fish species in this area are migratory fish, of which numerous species spawn in the upper reaches[5].

Known as the world's second most productive inland fishery, the Mekong River supplies up to 80% of the protein needs for nearly 65 million people living in the lower Mekong Basin (LMB; [6,7]). Although not as widely known as catfish species, the black sharkminnow *Labeo chrysophekadion* (Bleeker, 1849) (Cypriniformes: Cyprinidae) is an important commercial food fish in Asia. This species is also used in the aquarium trade, and it shows potential for aquaculture development[8,9]. Throughout its distribution range from the Mekong and Chao Phraya Basins to the Malay-Indonesian Archipelago[8], size variations were recorded: 40 cm in Vietnam, 70 cm in Cambodia and Lao PDR, and up to 90 cm in Thailand[10].

VNUHCM PRESS

Currently, the migration pattern of *L. chrysophekadion* is still debated, whether it is a short or long migratory fish, commonly known as "gray" or "white" fish, respectively [11–13]. Although this species has been reported to be migratory [9], information about the migration behavior of *L. chrysophekadion* remains fragmented. Generally, mature adults begin their upstream migration at the end of the rainy season and early dry season (March to August), and spawning occurs at the onset of the rainy season (e.g., June to July in southern Lao PDR) [8]. The fry and adults migrate back to floodplains for feeding and return to mainstems from October to December. The migratory routes may vary due to different habitats, such as between permanent and seasonal water bodies of floodplains, waterfalls (e.g., Khone Falls) to the Mekong Delta, and/or subcatchments and small streams [9,10].

Knowledge gaps exist for most Mekong fish species, including the black sharkminnow, and biological, ecological, and genetic studies are needed. With advancements in next-generation sequencing and bioinformatic tools, we are gradually replacing single-molecule markers with whole-genome sequencing [14]. Mitochondrial DNA is a subset obtained from whole-genome sequencing and a crucial molecular marker used in evolutionary genetics, molecular ecology, species identification, and conservation biology. This marker is characterized by high mutation rates, the absence of recombination, maternal inheritance, and a rapid evolutionary rate [15]. While the complete mitochondrial genome (mitogenome) offers a wealth of information about genetic diversity and evolutionary processes, DNA barcoding provides a valuable tool for identifying fish species in the absence of sufficient morphological data.

Assembly of the complete mitogenome from RAD-seq data is now possible [16–18]. Furthermore, Bird (2021) developed a pipeline to identify aligned mitogenome segments (AMS) from RAD-seq data, which can be applied to investigate the gene flow and population structure of aquatic organisms [19–21].

The lack of information on the genetic diversity and migration routes of exploited species, such as *L. chrysophekadion,* hinders the development of sustainable management policies, which may exacerbate resource depletion. Currently, there is only one complete mitogenome available from an individual collected in Kandal, Cambodia [22]. Multiple distinct populations of *L. chrysophekadion* are believed to exist in the LMB based on fisheries observation data [8]. Moreover, according to Mashyaka and Duong (2021), no significant differences were found between sampling sites from Paksan (Lao PDR) to the Vietnamese Mekong Delta, despite an estimated distance of 1,200 km, as determined by intersimple sequence repeat markers [12].

This study aimed to 1) assemble and annotate *L. chrysophekadion* mitogenomes from the LMB using restriction site-associated DNA sequencing (ezRAD) and 2) identify AMSs to examine the genetic diversity and population structure of *L. chrysophekadion* in the LMB. These significant findings provide valuable information for comprehensive research and effective tools for managing single and multiple species in the Mekong River.

MATERIALS AND METHODS

Fish Sampling

A total of 255 individuals of *L. chrysophekadion* were field-identified [8,23,24] and collected at nine locations along the LMB. The sampling strategy included six mainstem sites spread across four countries (Paksan, Pakse – Lao PDR; Ubon Ratchathani – Thailand; Kratie – Cambodia; Dong Thap, and An Giang – Vietnam); one site at the confluence of the Mekong and 3S Rivers (Stung Treng – Cambodia); and two LMB tributary sites – the Khan River – Luang Prabang, Lao PDR – and the Chi River – Roi Et, Thailand (Table 1, Figure 1)

Muscle tissue (~ 50 mg) from each individual was preserved in 95% molecular grade ethanol and transported to the Molecular Biology Laboratory at Nha Trang University, Vietnam, for further analysis.

Genomic library preparation and sequencing

DNA extraction was performed from preserved tissue samples using the Wizard® SV Genomic DNA Purification System kit (Promega, USA) following the manufacturer's instructions. A minor modification was made in the elution step; the extracted DNA was eluted three separate times, with 100 μL of AE buffer used each time instead of 250 μL of nuclease-free water. Subsequently, all elutions were subjected to electrophoresis on a 1% agarose gel and quantified using a Qubit 2.0 fluorometer with the dsDNA High Sensitivity kit (Invitrogen).

Selected DNA (100 ng, $\geq$ 3 ng/μl) from 255 individuals was used for ezRAD library preparation [16,17]. The implementation process involved randomly fragmenting the genomic DNA, performing end repair, size selection, A-tailing, ligating with Illumina adapters, and PCR amplification. All libraries were then sent to the Genomics Core Laboratory (Texas
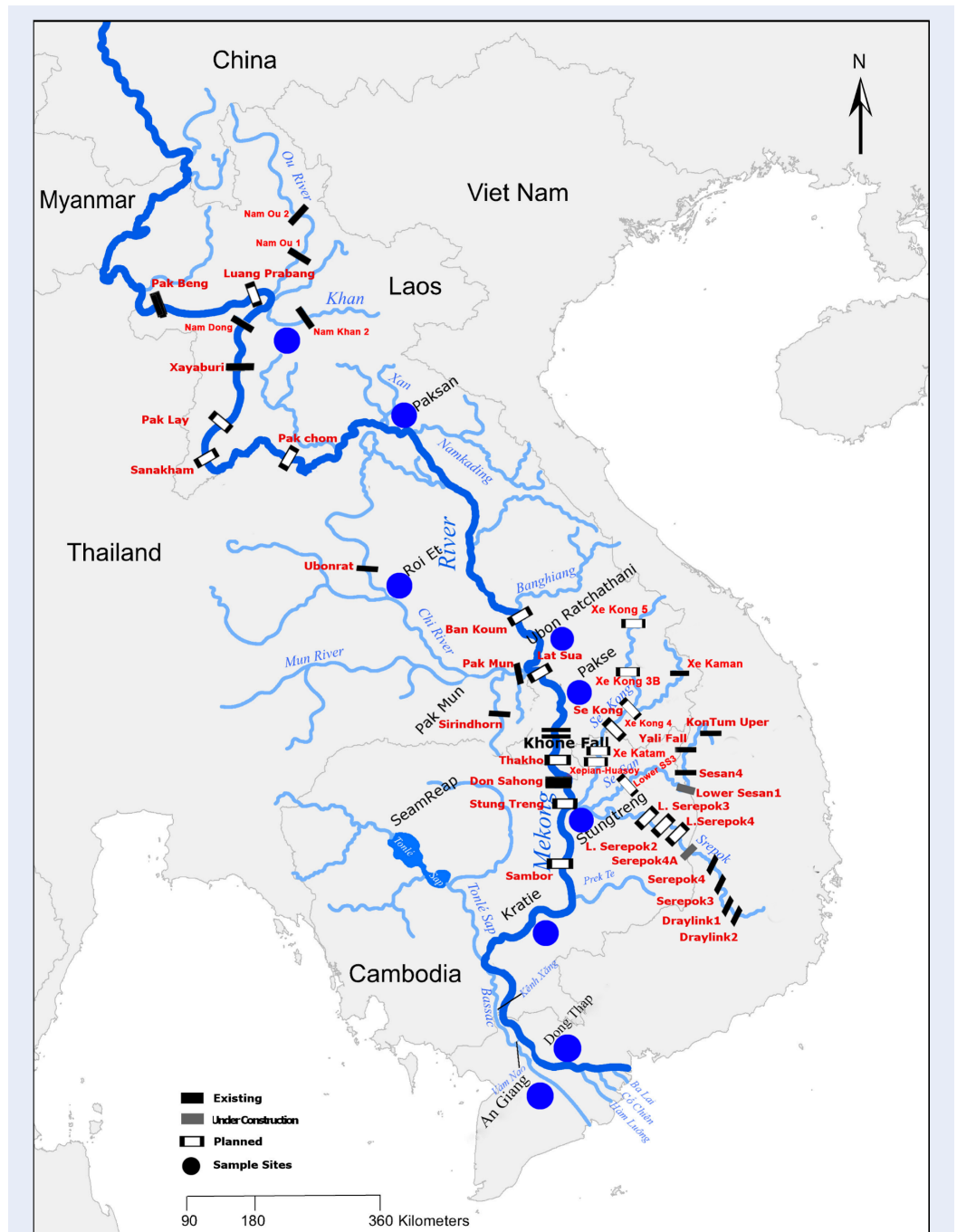
Figure 1: **Sampling sites (blue dots) of Labeo chrysophekadion in the lower Mekong Basin**. The red words represent existing, under construction, or planned hydropower dams (62).
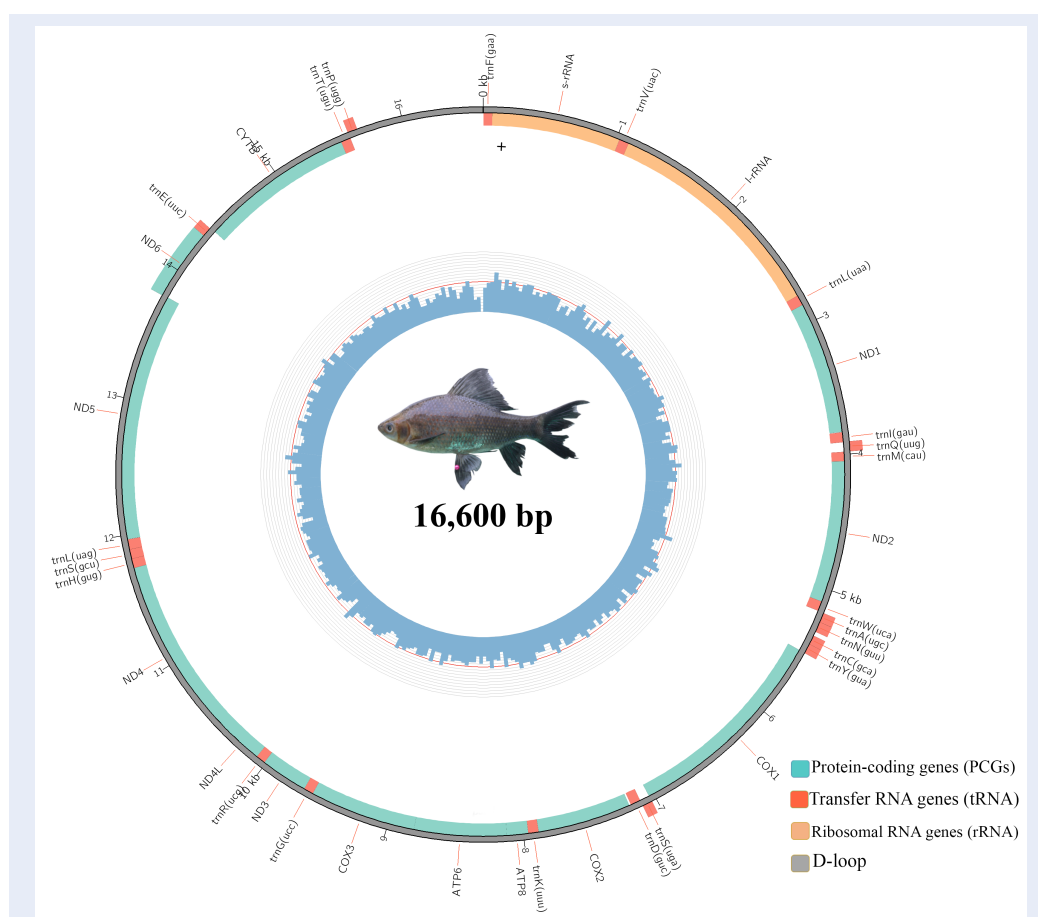
Figure 2: **Haplotype network of Labeo chrysophekadion mitogenomes in the lower Mekong Basin**. Yellow circles represent eight consensus sequences and one mitogenome retrieved from GenBank. Red dots represent median vectors. The numbers in parentheses indicate the number of base step mutations distinguishing the haplotypes.

Table 1: **Information on the sampling sites for *Labeo chrysophekadion* in the lower Mekong Basin**

LMB location	Country	Sampling sites (Code)	Geographic coordinates		No. of individuals
Mainstem	Lao PDR	Paksan (PA)	18°23'40.5"N	103°39'09.1"E	32
		Pakse (PE)	15°07'30.0"N	105°48'47.8"E	32
	Thailand	Ubon Ratchathani (UB)	15°18'48.2"N	105°29'52.6"E	28
	Cambodia	Kratié (KT)	12°49'31"N	106°01'71.5"E	27
	Vietnam	An Giang (AG)	10°41'07.1"N	105°11'59.1"E	24
		Dong Thap (DT)	10°46'59.5"N	105°20'49.6"E	32
Mekong and 3S confluence	Cambodia	Strung Treng (ST)	13°31'46.7"N	105°57'06.6"E	28
Tributary	Lao PDR	Luang Prabang (LP)	19°53'39.2"N	102°08'28.6"E	22
Tributary	Thailand	Roi Et (RE)	15°57'39.8"N	103°59'31.5"E	30
Total					**255**

Figure 3: **Circular map of the mitochondrial genome of Labeo chrysophekasion**. Genes encoded on the H-strand and L-strand are shown inside and outside the circular map, respectively. The GC and AT contents are plotted in the dark and light regions in the inner gray circle, respectively.

A&M University, Corpus Christi, USA) for paired-end 150 bp sequencing on the Illumina HiSeq 4000 platform.

Mitogenome assembly and annotation

The data were analyzed on a server with the following configuration: Intel(R) Xeon(R) Gold 6168 CPU @ 2.40 GHz, 80 CPU, and 187 GB of RAM. The operating system used was Ubuntu 21.10, version X_86 64-bit.

The quality of the raw paired-end reads (FASTQ) of the obtained libraries was analyzed and visualized using FastQC[25] and MultiQC[26], respectively. The mitochondrial genomes were assembled following the MitoZ v3.4 toolkit[27]. Trimmomatic v0.36[28] was used to remove adapters, restriction site sequences, bases with a Phred quality score less than 30, and any reads that were less than 50 base pairs in length. Then, quality-filtered reads were assembled using *de Bruijn*

graph (DBG) algorithms based on Megahit v1.2.9 (quick mode default *k-mer* length of 71)[29]. The output files (FASTA) were assembled contigs and/or scaffolds of both the mitochondrial and nuclear genomes. The FindMitoScaf module was applied for the following steps: *(i)* The genomes were mapped to the profile Hidden Markov Model; *(ii)* All sequences falling outside the database were removed; and *(iii)* The confidence scores were calculated and ranked for protein-coding genes. GeneWise v2.2[30], MiTFi v1.0[31], and infernal v1.1.1[32] were used to annotate protein-coding genes (PCGs), transfer RNA (tRNA), and ribosome RNA (rRNA), respectively.

Consensus genome sequences in FASTA format were aligned to the *L. chrysophekadion* mitogenome from GenBank (AP011199[33]) using pagan2[34]. Based on sequence length and percentage mapping, a haplotype network of eight consensus mitogenomes from LMB and the previously published genome was cre-
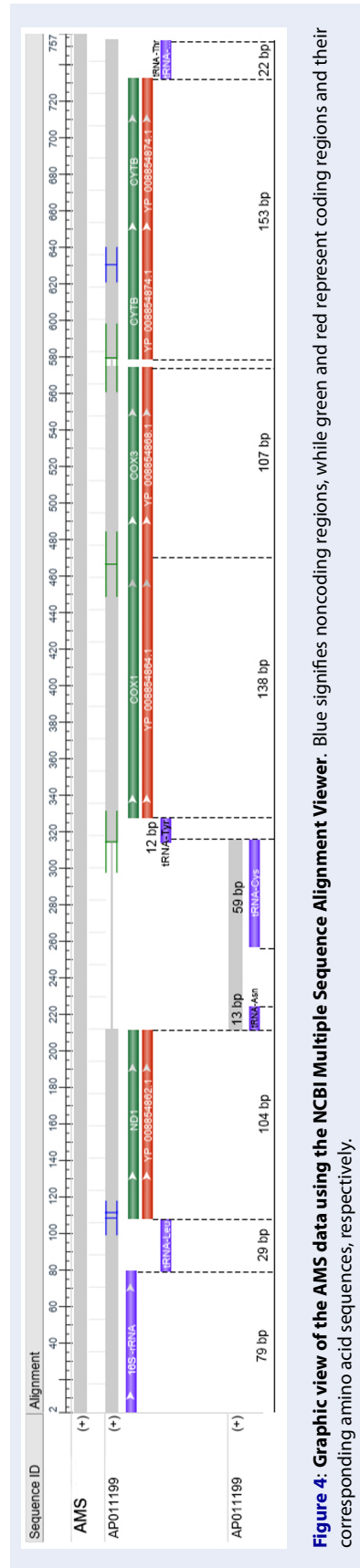
Figure 4: **Graphic view of the AMS data using the NCBI Multiple Sequence Alignment Viewer**. Blue signifies noncoding regions, while green and red represent coding regions and their corresponding amino acid sequences, respectively.
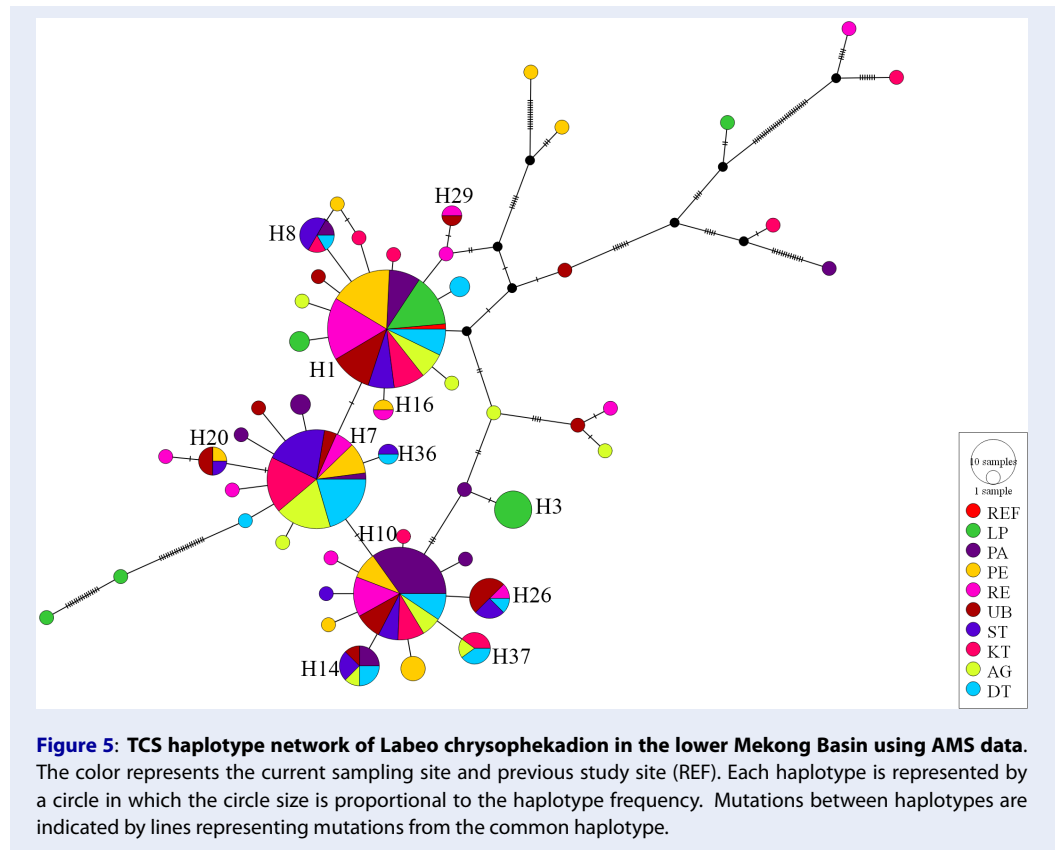
Figure 5: **TCS haplotype network of Labeo chrysophekadion in the lower Mekong Basin using AMS data**. The color represents the current sampling site and previous study site (REF). Each haplotype is represented by a circle in which the circle size is proportional to the haplotype frequency. Mutations between haplotypes are indicated by lines representing mutations from the common haplotype.

ated using POPART v1.7[35]. Additionally, Kimura's two-parameter genetic distance was calculated using BioEdit 7.0.5.3[36] to determine the genetic differences between the LMB *L. chrysophekadion* consensus sequences.

The selected complete mitogenome was rearranged using BWA v0.7.17[37] and SAMtools v1.15.1[38]. A circular map of the *L. chrysophekadion* mitogenome was generated using Circos[39]. The nucleotide composition of the mitogenome was determined using MEGA X[40]. Finally, the complete mitogenome was submitted to GenBank using Bankit (https://submit.ncbi.nl m.nih.gov/about/bankit/).

AMS identification and population genetics

DNA mitogenome processing was implemented using the RADbarcoder pipeline[22]. All reads from each individual that passed quality trimming were mapped to the current mitogenome of *L. chrysophekadion* using BWA v0.7.12[37] with the MEM algorithm[34]. The unmapped reads were filtered using the 'stats' function in SAMtools[38]. The 'bam2GENO' function was used to convert the resulting BAM files to consensus genome sequences in FASTA format, which were

then aligned to two *L. chrysophekadion* mitogenomes (AP011199[33] and the OR637878 in current study) using pagan2[41]. The 'fltrGENOSITES' function was applied to remove sites with missing/ambiguous/indel base calls and individual sequences with low percentage sequence mapping (< 50%). The position on the mitochondrial genome of the collected AMS (FASTA file) was determined using the Basic Logical Alignment Search Tool (BLAST, http://blast.ncbi.nlm.nih .gov/) and viewed by Multiple Sequence Alignment Viewer v1.25.0. Then, the FASTA file of the AMS data was converted to a NEXUS file for further analysis using Seaview[42].

To visualize the relationships between individuals and populations of *L. chrysophekadion* in LMB, a haplotype network was constructed based on the AMS data using the TCS algorithm[43] implemented in POPART v1.7[35]. Genetic diversity indices, including the number of haplotypes (H), number of polymorphism sites (S), haplotype (Hd) and nucleotide (π) diversity, were calculated using DnaSP v5[18]. The genetic differences (F_{ST}) between all pairs of sites were also computed using ARLEQUIN v3.5[44].

RESULTS

Mitogenome structure and composition

In this study, a total of 255 ezRAD libraries of *L. chrysophekadion* from nine sampling sites across the LMB were sequenced. A total of 1,062,049,264 raw sequence reads (151 bp paired-end), ranging from 1,094–37,265,254 reads per individual, were obtained. After adapter trimming and quality filtering, 1,026,721,048 high-quality reads (912–36,260,038 per individual) were passed and used to assemble the mitogenome. With the MitoZ toolkit, 0.06-0.12% of the high-quality reads were successfully mapped to the *L. chrysophekadion* mitogenome. Due to low percentage sequence mapping (< 50%), eight consensus mitogenomes were removed from the dataset. The lengths of the remaining 247 consensus mitogenomes varied from 8,510-16,600 bp, 20-8,069 bp in gap, and 0-152 in missing nucleotides (N). Among these, 46 mitogenomes had lengths greater than 16,000 bp (16,011–16,600) and displayed 96.3–99.8% identity to a previously published genome (AP011199)[33] (Table 2). Furthermore, due to the greatest length and high-quality mapping, eight consensus sequences representing eight sampled populations in the LMB (excluding the An Giang population due to low resolution) were chosen to construct the LMB *L. chrysophekadion* mitogenomes. The sequence differences ranged from 0.2% (Ubon Ratchathani and Stung Treng) to 4.7% (Dong Thap and Kratié) (Table ??). The haplotype network of the eight selected consensus sequences and one previous mitogenome displayed two distinct clades. The six consensus mitogenomes (KT, PE, UB, ST, LP, and RE) clustered with the *L. chrysophekadion* mitogenome retrieved from GenBank (AP011199), differing by 1–10 mutational differences. Moreover, the DT and PA mitogenomes were closely related (10–44 mutational differences) (Figure 2). However, the divergence may be caused by missing information due to the shorter lengths (16,248 bp in DT and 16,571 bp in PA) and the number of gaps (342 and 31 in DT and PA, respectively).

Based on its high similarity (99.8%) to a previously published mitogenome, the selected complete mitogenome of *L. chrysophekadion* from Ubon Ratchathani, Thailand, was chosen for annotation. This mitogenome contains 16,600 bp, with 42.9% GC content. The mitogenome contains 37 typical mitochondrial genes, including 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNA), 2 ribosomal RNA genes (rRNA), and a noncoding control region of the D-loop. Most of the *L. chrysophekadion* mitochondrial genes are encoded on the heavy strand (H-strand), while one PCG (*ND6*) and eight tRNA genes (*tRNA^{Ala}, tRNA^{Asn}, tRNA^{Cys}, tRNA^{Gln}, tRNA^{Glu}, tRNA^{Pro}, tRNA^{Ser},* and *tRNA^{Tyr}*) are encoded on the L-strand (Figure 3, Table ??). There are 14 intergenic spacers totaling 74 bp, of which the longest spacer is 33 bp, located between the tRNAAsn and tRNACys genes. In addition, a total of 41 bp overlaps were identified in 7 genes, with the overlap length of each gene ranging from 1 to 7 bp (Table ??).

As shown in Table ??, all PCGs start with an ATG codon, except for the *COXI* gene, which has GTG as the start codon. In addition, TAA is the most common stop codon, except for *ND2*, which uses TAG. The *ND3, ND4,* and *ND6* genes ended with the incomplete stop codon T–.

AMS identification and population genetics

Following the RADbarcoder pipeline, the mapping of high-quality reads from 247 individuals to the current *L. chrysophekadion* mitogenome and alignment to both the current and previous mitogenomes resulted in the identification of 757 bp of aligned mitogenome segment data (Table 2). Among these, 502 bp were found in PCGs (*ND1, COXI, COXIII,* and *cytb*), 79 bp in rRNA (16S), and the remaining in tRNA genes (*tRNA^{Leu}, tRNA^{Asn}, tRNA^{Cys}, tRNA^{Tyr},* and *tRNA^{Thr}*) (Figure 4).

A total of 49 haplotypes (19.8%) were identified from 247 individuals of *L. chrysophekadion.* The number of haplotypes ranged from 6/22 (LP, 27.3%) to 12/30 (RE, 40%). The average haplotype diversity (Hd) was high (mean±SD = 0.849±0.014), ranging from 0.71±0.071 (LP) to 0.871±0.046 (UB). The average nucleotide diversity (π) and number of polymorphic sites (S) were low (0.005±0.0008 and 113), ranging from 0.002±0.001 and from 9–10 (DT and ST) to 0.016±0.002 and 45 (LP), respectively (Table 3). There was little difference in haplotype diversity between populations in the mainstem, Strung Treng confluence and tributaries of the Mekong River, while the downstream populations (AG, DT in the Mekong Delta) and confluence site (ST) displayed low nucleotide diversity. As shown in Table ??, small F_{ST} values and nonsignificant differences were detected among populations (F_{ST} = 0-0.04, P>0.05), except for Luang Prabang, where significant differentiation was evident (F_{ST} = 0.08-0.17, P<0.05).

The haplotype network of *L. chrysophekadion* revealed high connectivity among the nine defined populations at LMB. Three dominant common haplotypes (H7, H10, and H1, Figure 5) were detected and were

Table 2: Summary of read count, consensus sequence length, and number of individuals after various steps of mitogenome assembly and RADbarcoder pipeline

Parameters	No. of reads/Length	No. of individuals
Raw reads (reads)	1,094 – 37,265,254	255
High-quality reads after trimming (reads)	912 – 36,260,038	255
High-quality reads per individual successfully mapped to mitogenome (%)	0.06 – 0.12	253
Consensus sequence length per individual (bp)	8,510 – 16,600	247
Consensus sequences mapped to reference mitogenome (%)	51.2 – 99.8	247
AMS dataset collected (bp)	757	247

Table 3: Summary statistics of genetic variation in *Labeo chrysophekadion* in the lower Mekong Basin

Sampling sites (Code)	Nse	H	S	Hd (mean±SD)	π (mean±SD)
Paksan (PA)	31	10	37	0.738±0.073	0.005±0.002
Pakse (PE)	30	10	30	0.805±0.058	0.005±0.002
Ubon Ratchathani (UB)	26	11	19	0.871±0.046	0.004±0.002
Kratié (KT)	27	10	55	0.835±0.049	0.009±0.003
An Giang (AG)	24	10	16	0.822±0.061	0.004±0.002
Dong Thap (DT)	29	10	10	0.842±0.049	0.002±0.001
Strung Treng (ST)	28	9	9	0.833±0.051	0.002±0.001
Luang Prabang (LP)	22	6	45	0.71±0.071	0.016±0.002
Roi Et (RE)	30	12	54	0.807±0.06	0.007±0.003
Total	**247**	**49**	**113**	**0.849±0.014**	**0.005±0.0008**

Nse: number of individuals used in analyses; S: number of polymorphic sites; H: number of haplotypes; Hd: haplotype diversity; π: nucleotide diversity

shared among 8 and 9 populations. Additionally, haplotype H14 was shared by 5 populations (PA, UB, ST, AG, and DT); haplotype H8 was found in PA, ST, KT, and DT; and haplotype H26 was found in RE, UB, ST, and DT. Furthermore, six haplotypes (H16, H2O, H29, H36, and H37) were joined by at least 2 populations. The Luang Prabang (LP) population shared only one common haplotype (H1) and was characterized by several private haplotypes (e.g., H3 shared by 7 individuals) spanning mutation steps. A high number of unique haplotypes (i.e., those found at a single location) were detected (35 out of 49) and were distributed among all examined populations (Figure 5).

DISCUSSION

The Mekong River is one of the most biodiverse and productive rivers in the world, supporting more than 1000 fish species and sustaining the livelihoods of millions of people[4,45]. As in other regions of the world, fisheries resources in the Mekong River have experienced declines due to factors such as exploitation, environmental pollution, urban development, habitat fragmentation, and climate change[1,46]. Therefore, understanding the population structure and connectivity of fishes is crucial for mitigating the detrimental effects of these threats and implementing effective multispecies management strategies[17].

In recent years, rapid advances in high-throughput sequencing technologies and available bioinformatic tools have facilitated the successful assembly and annotation of growing numbers of mitogenomes, including those of Mekong fish species[47–50]. Despite its importance and wide distribution range, only one available mitogenome of *L. chrysophekadion* has been generated from an individual fish in Kandal, Cambodia[33]. In this study, based on RAD-seq data, an additional mitogenome (16,600 bp) from a fish individual collected in Ubon Ratchathani, Thailand, was assembled and submitted to GenBank. This genetic information will enrich our understanding of this fish species in the LMB and help fill existing knowledge gaps.

Among the 54 available mitogenomes of *Labeo* species, the sequence length varied from 16,602 bp (*L. chrysophekadion*) to 16,766 bp (*L. pierrei*)[33]. Generally, the minor length variations between closely related species are caused by changes in tandem repeats within the control region, the lengths of intergenic regions, and gene overlaps[51]. In comparison to a previous mitogenome sequence, our study failed to identify two nucleotides (C and A) that were found at the last position of the D-loop. In this case, the cause may be sequence disturbance, rendering it unreliable for detecting these two nucleotides. The circular map also showed that the gene order (13 PCGs, 2 rRNAs, 22 tRNAs and a D-loop) was consistent with that of previously published mitogenomes across various fish species[47–49,52].

Over the past few decades, numerous mitochondrial DNA datasets (later combined with nuclear markers) have been generated across diverse sets of organisms[12,19,21,47–49]. Recently, due to genome-wide analyses, genetic analyses have expanded to include nuclear genomes[53]. AMS is a new approach that utilizes the power of whole-genome sequencing to subset the mitogenome, allowing comparison with the large dataset of mitogenomes that have been produced. The 757 bp AMS present on most coding and noncoding genes allowed a complete analysis of genetic diversity and population connectivity and comparison with analyses from numerous mitogenome studies. Overall, high haplotype (Hd) and low nucleotide (π) diversity of *L. chrysophekadion* were observed based on the categories of genetic diversity suggested by Grant and Bowen (1998)[54]. In comparison to another study conducted in the LMB, one migratory catfish, *Pangasius krempfi* (Siluriformes: Pangasiidae), exhibited similar levels of genetic diversity, with Hd = 0.941 and π = 0.0083 when analyzing the D-loop region.

However, lower genetic diversity was observed for the *cytb* gene, with Hd = 0.381 and π = 0.00063[55].

Low genetic diversity was detected in populations related to the Khan River (LP), confluence site (ST), and delta sites (AG and DT). Interestingly, the RE population, which was separated from the mainstem by the Pak Mun dam, showed high genetic diversity. The haplotype network showed high connectivity among all populations, except for Luang Prabang. *Labeo chrysophekadion* is known to migrate upstream of the Mekong River for spawning and downstream for feeding[9] however, its migration distance has not been documented. Hydropower dams can act as barriers that fragment habitats, block access to spawning grounds, and modify habitats both upstream and downstream. These alterations can result in reduced genetic diversity and increased genetic differences among isolated riverine fish populations[56]. An explanation for the low genetic diversity in the LP population may be the development of hydroelectric dams on the Mekong mainstem and tributaries in Lao PDR (Nam Dong, Nam Khan 2, Nam Khan 3, and Nam Ko), which hinders the population from completing the migration routes.

Thirteen dams have been built in the Sesan and Srepok Rivers in Vietnam (without fish passages), and approximately 7 dams have been proposed to be built in the Sekong River (Lao PDR) **(Figure 1)**. According to the knowledge of fish migration routes in the LMB, the Sesan catchment basin is an important spawning site, as well as a refuge during the dry season[57]. A previous study reported a diminishing effective population size, elevated relatedness, and inbreeding of one catfish species, *Hemibagrus spilopterus*, sampled upstream of dams (Dak Lak) on the Srepok River[58]. A low effective population size (Ne) was also recorded in populations in the 3S tributary of two fish species, *Henicorhynchus lobatus* and *Helicophagus leptorhynchus*[59].

The Mekong Delta is one of the regions that suffers from severe consequences from climate change and human activities. Research has shown that an important food fish species, *Polynemus melanochir*, is at risk of difficulty recovering from environmental changes[17]. A similar situation could occur with *L. chrysophekadion*, a fish species of high economic value that is currently facing overexploitation.

Based on microsatellite markers, Mashyaka and Duong (2019) reported the population connectivity of *L. chrysophekadion* between the Mekong Delta (Can Tho, An Giang, and Dong Thap) and Lao PDR (Paksan) and suggested a long migratory distance of this

species (approximately 1,200 km)[12]. Biological information and studies on the genetic diversity and population structure of Mekong fishes are still limited[12,17,58–62], and additional studies are needed to reliably infer the migratory patterns of *L. chrysophekadion*. Given the ecological diversity and hydrological conditions, multispecies management is essential for the Mekong River. Therefore, more in-depth research on molecular ecology and species biology is needed to develop effective conservation strategies.

CONCLUSIONS

Our study assembled and analyzed the mitogenome sequence of *L. chrysophekadion*, which is 16,600 bp in length. This mitogenome is composed of 37 genes, including 13 PCGs, 22 tRNA genes, 2 rRNA genes, and a D-loop region. An AMS dataset consisting of 757 bp was identified from 247 individuals collected from nine sample sites across the LMB. High connectivity was detected among the sample populations, except for Luang Prabang. This significant finding provides a comprehensive overview of the population structure, serving as a scientific basis for the conservation and management of aquatic resources.

LIST OF ABBREVIATIONS USED

3S: Sekong, Sesan and Srepok

H: number of haplotypes

Hd: haplotype diversity

LMB: lower Mekong Basin

Ne: Effective population size

Nse: number of individuals used in analyses

PCGs: protein-coding genes

RAD: Restriction site-associated DNA

rRNA: ribosome RNA

S: number of polymorphic sites

tRNA: transfer RNA

π: nucleotide diversity

COMPETING INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

1. Campbell I, Barlow C. Hydropower development and the loss of Fisheries in the Mekong River Basin. Front Environ Sci. 2020;8:1-10;Available from: https://doi.org/10.3389/fenvs.2020.566509.

2. Kano Y, Dudgeon D, Nam S, Samejima H, Watanabe K, Grudpan C, et al. Impacts of dams and global warming on fish biodiversity in the Indo-Burma hotspot. PLoS One. 2016;11(8):1-21;Available from: https://doi.org/10.1371/journal.pone.0160151.

3. Stone R. Dam-building threatens Mekong fisheries. Science. 2016;354:1084-5;Available from: https://doi.org/10.1126/science.354.6316.1084.

4. Valbo-Jorgensen J, Coates D, Hortle K. Fish Diversity in the Mekong River Basin. In: Campbell IC, editor. The Mekong biophysical environment of an International river basin. Elsevier, Amsterdam; 2009. p. 161-96;Available from: https://doi.org/10.1016/B978-0-12-374026-7.00008-5.

5. Baran E, Myschowoda C. Have fish catches been declining in the Mekong River Basin? In: Modern Myths of the Mekong - Part II: Making living out of nature. Finland: Helsinki University of Technology; 2008. p. 55-64;.

6. Baran E, Jantunen T, Chong CK. Values of Inland Fisheries in the Mekong River Basin. In: Tropical river fisheries valuation: background papers to a global synthesis. Penang, Malaysia, WorldFish Center; 2007. p. 227-90;.

7. Hortle KG. Chapter 9 - Fisheries of the Mekong River Basin. In: Campbell ICBT-TM, editor. Aquatic Ecology. San Diego: Academic Press; 2009. p. 197-249;Available from: http://www.sciencedirect.com/science/article/pii/B9780123740267000097.

8. Rainboth WJWJ. Fishes of the Cambodian Mekong. FAO Species Identification Field Guide for Fishery Purposes, Rome; 1996. 265, 27 color plates p;.

9. Poulsen AF, Valbo-Jorgensen J, Chan S, Chhuon CK, Hortle KG, Viravong S, et al. Distribution and Ecology of Some Important Riverine Fish Species of the Mekong River Basin. MRC Technical Paper. 2004;.

10. Sokheng C, Chhuon Kim Chhea, Viravong S, Bouakhamvongsa K, Suntornratana UNY, Tung NT, et al. Fish Migrations and Spawning Habits in the Mekong Mainstream - A Survey using Local Knowledge. Assessment of Mekong Fisheries: Fish Migrations and Spawning and the Impact of Water Management Project (AMFC); 1999;.

11. Chan B, Ngor PB, So N, Lek S. Spatial and temporal changes in fish yields and fish communities in the largest tropical floodplain lake in Asia. Int J Limnol. 2017;53:485-93;Available from: https://doi.org/10.1051/limn/2017027.

12. Mashyaka A, Duong T-Y. Genetic diversity analysis revealed possible long migration of Black sharkminnow (Labeo chrysophekadion) along the Mekong River. Songklanakarin J Sci Technol. 2021;43(4):955-60;.

13. Pin K, Nut S, Hogan ZS, Chandra S, Saray S, Touch B, et al. Cambodian freshwater fish assemblage structure and distribution patterns: Using a large-scale monitoring network to understand the dynamics and management implications of species clusters in a global biodiversity hotspot. Water. 2020;12:1-25;Available from: https://doi.org/10.3390/w12092506.

14. Sturk-Andreaggi K. Mitochondrial Genome Analysis Using Next Generation Sequencing for Forensic Applications. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine, Uppsala: Acta Universitatis Upsaliensis; 2022. 73 p;.

15. Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999;27(8):1767-80;PMID: 10101183. Available from: https://doi.org/10.1093/nar/27.8.1767.

16. Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-silva I, Andrews KR, et al. ezRAD: a simplified method for genomic genotyping in nonmodel organisms. PeerJ. 2013;1:e203:1-15;PMID: 24282669. Available from: https://doi.org/10.7717/peerj.203.

17. Dang BT, Vu QHD, Biesack EE, Doan T V, Truong OT, Tran TL, et al. Population genomics of the peripheral freshwater fish Polynemus melanochir (Perciformes, Polynemidae) in a changing Mekong Delta. Conserv Genet. 2019;20:961-72;Available from: https://doi.org/10.1007/s10592-019-01189-x.

18. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003;19(18):2496-7;Available from: https://doi.org/10.1093/bioinformatics/btg359.

19. Pante E, Abdelkrim J, Viricel A, Gey D, France S, Boisselier M, et al. Use of RAD sequencing for delimiting species. Heredity (Edinb). 2015;114:450-9;PMID: 25407078. Available from: https://doi.org/10.1038/hdy.2014.105.

20. Stobie CS, Cunningham MJ, Oosthuizen CJ, Bloomer P. Finding stories in noise: Mitochondrial portraits from RAD data. Mol Ecol Evol. 2019;19(1):191-205;PMID: 30312000. Available from: https://doi.org/10.1111/1755-0998.12953.

21. Rue CR, Selwyn JD, Cockett PM, Gillis B, Gurski L, Jose P, et al. Genetic diversity across the mitochondrial genome of eastern oysters (Crassostrea virginica) in the northern Gulf of Mexico. PeerJ. 2021;9:e12205;PMID: 34692250. Available from: https://doi.org/10.7717/peerj.12205.

22. Bird CE. radBARCODER: A pipeline to assemble, align, and filter plastid genomes from nucleotide data in binary alignment maps. 2021;Available from: https://github.com/cbirdlab/radBARCODER.

23. Tran DD, Shibukawa K, Nguyen TP, Ha PH, Tran XL, Mai VH, et al. Fishes of the Mekong Delta, Vietnam. Can Tho University Publishing House, Can Tho; 2013. 174 p;.

24. Nagao Natural Environment Foundation. Fishes of the Indochinese Mekong. Nagao Natural Environment Foundation, Tokyo; 2021. xii+546;.

25. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010;Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

26. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8;PMID: 27312411. Available from: https://doi.org/10.1093/bioinformatics/btw354.

27. Meng G, Li Y, Yang C, Liu S. MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 2019;47(11);PMID: 30864657. Available from: https://doi.org/10.1093/nar/gkz173.

28. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20;PMID: 24695404. Available from: https://doi.org/10.1093/bioinformatics/btu170.

29. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: An ultrafast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674-6;PMID: 25609793. Available from: https://doi.org/10.1093/bioinformatics/btv033.

30. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14(5):988-95;PMID: 15123596. Available from: https://doi.org/10.1101/gr.1865504.

31. Jühling F, Pütz J, Bernt M, Donath A, Middendorf M, Florentz C, et al. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 2012;40(7):2833-45;Available from: https://doi.org/10.1093/nar/gkr1131.

32. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29(22):2933-5;Available from: https://doi.org/10.1093/bioinformatics/btt509.

33. Yang L, Arunachalam M, Sado T, Levin BA, Golubtsov AS, Freyhof J, et al. Molecular phylogeny of the cyprinid tribe Labeonini (Teleostei: Cypriniformes). Mol Phylogenet Evol. 2012;65(2):362-79;Available from: https://doi.org/10.1016/j.ympev.2012.06.007.

34. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics. 2013;3;Available from: https://doi.org/10.6084/M9.FIGSHARE.963153.V1.

35. Leigh JW, Bryant D. POPART: full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6:1110-6;Available from: https://doi.org/10.1111/2041-210X.12410.

36. Hall TATA. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows95/98/NT. Nucleic Acids Symp Ser. 1999;41:95-8;.

37. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60;Available from: https://doi.org/10.1093/bioinformatics/btp324.

38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9;PMID: 19505943. Available from: https://doi.org/10.1093/bioinformatics/btp352.

39. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45;Available from: https://doi.org/10.1101/gr.092759.109.

40. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547-9;Available from: https://doi.org/10.1093/molbev/msy096.

41. Löytynoja A, Vilella AJ, Goldman N. Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. Bioinformatics. 2012;28(13):1684-91;PMID: 22531217. Available from: https://doi.org/10.1093/bioinformatics/bts198.

42. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27(2):221-4;PMID: 19854763. Available from: https://doi.org/10.1093/molbev/msp259.

43. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. Mol Ecol. 2000;9(10):1657-9;Available from: https://doi.org/10.1046/j.1365-294x.2000.01020.x.

44. Excoffier L, Lischer HEL. An Integrated Software Package for Population Genetics Data Analysis. Mol Ecol Resour. 2010;10:564-7;Available from: https://doi.org/10.1111/j.1755-0998.2010.02847.x.

45. Baumgartner LJ, Boys CA, Barlow C, Roy M. Lower Mekong Fish Passage Conference: Applying innovation to secure fisheries productivity. Ecol Manag Restor. 2017;18(3):E8-12;Available from: https://doi.org/10.1111/emr.12272.

46. Ziv G, Baran E, Nam S, Rodríguez-iturbe I, Levin SA. Trading-off fish biodiversity, food security, and hydropower in the Mekong River Basin. Proc Natl Acad Sci. 2012;109(15):5609-14;Available from: https://doi.org/10.1073/pnas.1201423109.

47. Duong TY, Pham LTK, Le XTK, Nguyen NTT, Nor SAM, Le TH. Mitophylogeny of Pangasiid Catfishes and its Taxonomic Implications for Pangasiidae and the Suborder Siluroidei. Zool Stud. 2023;48;Available from: https://doi.org/10.22541/au.168084560.06743241/v1.

48. Zhang R, Deng L, Lv X, Tang Q. Complete mitochondrial genomes of two catfishes (Siluriformes, Bagridae) and their phylogenetic implications. Zookeys. 2022;1115:103-

16;Available from: https://doi.org/10.3897/ZOOKEYS.1115.85249.

49. Laopichienpong N, Ahmad SF, Singchat W, Suntronpong A, Pongsanarm T, Jangtarwan K, et al. Complete mitochondrial genome of Mekong fighting fish, Betta smaragdina (Teleostei: Osphronemidae). Mitochondrial DNA Part B Resour. 2021;6(3):776-8;Available from: https://doi.org/10.1080/23802359.2021.1882893.

50. Jondeung A, Sangthong P, Zardoya R. The complete mitochondrial DNA sequence of the Mekong giant catfish (Pangasianodon gigas), and the phylogenetic relationships among Siluriformes. Gene. 2007;387(1-2):49-57;Available from: http://dx.doi.org/10.1016/j.gene.2006.08.001.

51. Liu ZQ, Shen Y zhong, Zhao M, Wang W, Chen W, Ma CY, et al. Characterization of the complete mitochondrial DNA sequence of the Ariosoma meeki (Anguilliformes, Congridae). Mitochondrial DNA Part B Resour. 2020;5(3):2539-41;Available from: https://doi.org/10.1080/23802359.2020.1781557.

52. Islam MN, Sultana S, Alam MJ. Sequencing and annotation of the complete mitochondrial genome of a threatened labeonine fish, Cirrhinus reba. Genomics and Informatics. 2020;18(3):1-7;Available from: https://doi.org/10.5808/GI.2020.18.3.e32.

53. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17(2):81-92;Available from: https://doi.org/10.1038/nrg.2015.28.

54. Grant WS, Bowen BW. Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. J Hered. 1998;89(5):415-426;Available from: https://doi.org/10.1093/jhered/89.5.415.

55. Duong TY, Nguyen NTT, Tran DD, Le TH, Nor SAM. Multiple genetic lineages of anadromous migratory Mekong catfish Pangasius krempfi revealed by mtDNA control region and cytochrome b. Ecol Evol. 2023;13(2):1-14;Available from: https://doi.org/10.1002/ece3.9845.

56. Reid SM, Wilson CC, Mandrak NE, Carl LM. Population structure and genetic diversity of black redhorse (Moxostoma duquesnei) in a highly fragmented watershed. Conserv Genet. 2008;9(3):531-46;Available from: https://doi.org/10.1007/s10592-007-9367-2.

57. Baran E, Samadee S, Teoh SJ, Tran TC. Fish and fisheries in the Sesan river basin: catchment baseline, fisheries section. Project report. Mekong Challenge Program project MK3 "Optimizing the management of a cascade of reservoirs at the catchment level." The WorldFish Center; 2011. 61 p;.

58. Ackiss AS, Dang BT, Bird CE, Biesack EE, Chheng P, Phounvisouk L, et al. Cryptic Lineages and a Population Dammed to Incipient Extinction' Insights into the Genetic Structure of a Mekong River Catfish. J Hered. 2019;110(5):535-47;Available from: https://doi.org/10.1093/jhered/esz016.

59. Biesack EE, Dang BT, Ackiss AS, Bird CE, Chheng P, Phounvisouk L, et al. Evidence for population genetic structure in two exploited Mekong River fishes across a natural riverine barrier. J Fish Biol. 2020;97(3):696-707;Available from: https://doi.org/10.1111/jfb.14424.

60. Takagi APAP, Ishikawa S, Nao T, Song SLSL, Hort S, Thammavong K, et al. Genetic differentiation and distribution routes of the bronze featherback Notopterus notopterus (Osteoglossiformes: Notopteridae) in Indochina. Biol J Linn Soc. 2010;101(3):575-82;.

61. Takagi AP, Ishikawa S, Nao T, Song SL, Hort S, Thammavong K, et al. Genetic differentiation of Macrognathus siamensis within the Mekong River between Laos and Cambodia. J Appl Ichthyol. 2011;27(5):1150-4;Available from: https://doi.org/10.1111/j.1439-0426.2011.01774.x.

62. Stimson Center. Mekong Dam Monitor, Sharing Data, Empowering People. 2023;Available from: https://monitor.mekongwater.org/home/?v=1680811357217.