

**THE UNIVERSITY OF DANANG  
UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**TRUONG THI OANH**

**IDENTIFICATION AND APPLICATION OF SNP  
MARKERS USING EzRAD TECHNIQUE FOR  
THREE TYPICAL FISH SPECIES IN THE  
LOWER MEKONG RIVER BASIN**

**Specialization: Biotechnology  
Code: 9420201**

**TECHNICAL DOCTORAL THESIS SUMMARY**

**Da Nang – 2026**

The work was completed at:  
**UNIVERSITY OF SCIENCE AND TECHNOLOGY**

Supervisors:

- 1. ASSOC. PROF. DR. DANG THUY BINH**
- 2. DR. NGO THAI BICH VAN**

Reviewer 1: .....

Reviewer 2: .....

The thesis will be defended at the doctoral thesis evaluation committee meeting at the University of Science and Technology at date: .....

The thesis can be found at:

- Vietnam National Library
- Center for Learning Resources and Communication at Polytechnic University

# INTRODUCTION

## 1. Reasons for choosing the thesis

The Mekong River Basin (MRB), a biodiversity hotspot, is facing increasing threats from environmental degradation and human activities, particularly hydropower dam construction. These impacts not only create physical barriers that hinder migration, alter or fragment natural fish habitats, and affect the behavior of many aquatic species, but also change natural flow regimes, flood pulses, and seasonal water inputs. These changes have led to declining fisheries resources and an increased risk of extinction for fish species. Therefore, in-depth research on genetic diversity and population structure of fish species representing characteristic ecosystems and life history traits plays a crucial role in fisheries management, stock restoration, and the pursuit of sustainable ecosystem maintenance.

To date, a variety of molecular markers (mitochondrial DNA, Random Amplified Polymorphic DNA-RAPD, microsatellites, Single Nucleotide Polymorphisms-SNPs) have been used in population genetic studies of fish species in the MRB. Among these, the discovery of thousands of genome-wide SNPs from Restriction-site Associated DNA sequencing (RAD-seq) data has contributed to elucidating the population structure of organisms, especially species with limited genomic information.

Given the diverse migratory characteristics of Mekong River fish species in a dynamic and changing environment, three species with migratory patterns and different life cycles were collected, including *Macrognathus siamensis* (black-fish, non-migratory), *Labeo chrysophekadion* (grey-fish, short-distance migration and facultative), and *Pangasius larnaudii* (white-fish, long-distance migration). This

study addresses two questions: 1) How do migratory patterns and life cycles interact to shape genetic variation and gene flow among species? 2) Do existing genetic structures and diversity loss reflect the cumulative effects of landscape barriers and environmental changes in the LMB? These findings will help elucidate the mechanisms of genetic maintenance and differentiation in Mekong fish species, while providing a critical scientific basis for conservation strategies and the sustainable management of fishery resources.

## **2. Research Objectives**

By employing advanced genetic techniques, this study evaluates and compares the genetic diversity and population connectivity of three typical fish species in the LMB. The findings aim to establish a genetic database to support the management, conservation, and sustainable exploitation of fishery resources in the region.

## **3. Research Content**

1. *De novo* assembly and determination of SNP molecular markers for three typical fish species in the LMB.
2. Investigation and compare of genetic diversity, population structure, estimation of effective population size, and prediction of migration patterns of three typical fish species in the LMB.
3. Complete mitogenome assembly and annotation, identification of aligned mitochondrial DNA segments from RAD-seq data, and population structure analysis of *L. chrysophekadion*.

## **4. Scientific and Practical Significance of the Thesis**

**Scientific significance:** The study has elucidated the genetic structure and migratory patterns of three typical fish species in the Mekong River through the EzRAD technique. The results not only overcome the limitations of traditional methods regarding data

resolution but also provide a valuable reference framework for biodiversity conservation in major river systems. Notably, the integration of mitochondrial genome analysis and AMS from RAD-seq data has established a novel approach in studying maternal genetics and the adaptive capacity of aquatic species.

**Practical significance:** The findings of this research hold significant practical implications for the conservation and management of fisheries resources within the MRB. The elucidation of population structure and migration patterns of fish species enables resource managers to implement appropriate conservation strategies, ensuring the sustainability of fish populations. The SNPs and mitogenome data generated from this study constitute a valuable genetic resource, which can be utilized to monitor temporal genetic changes in fish populations, assess the impacts of environmental factors, and support population restoration programs. Furthermore, the assembly and analysis of the mitogenome of *L. chrysophekadion* can be applied in the development of species identification methods and seafood product traceability.

## **5. New contributions of the thesis**

**Content:** This study utilizes the EzRAD technique to determine and select highly efficient and accurate population-specific SNP markers. This approach provides a novel framework for population genetics research on three typical Mekong fish species, characterized by varying life histories and migratory patterns: the non-migratory *M. siamensis* (representative of local populations), the short-distance and facultative migrant *L. chrysophekadion* (demonstrating regional connectivity and environmental adaptability), and the long-distance migratory *P. larnaudii* (reflecting broader ecological linkages).

### **New points and highlights of the research:**

– This study offers a comprehensive analysis of genetic diversity and population structure in three typical Mekong fish species, using a large SNP dataset generated by EzRAD. It significantly improves upon previous studies that used traditional markers or single-species SNP analysis.

– This study, for the first time, employed SNP data and diverse algorithms to accurately predict migration patterns of *L. chrysophekadion* and *P. larnaudii* within the LMB, elucidating migration routes and population connectivity that are consistent with their historical development.

– This study utilizes EzRAD data from multiple individuals to both reconstruct the mitochondrial genome and simultaneously screen high-density polymorphic sites in *L. chrysophekadion*. This data source was effectively applied to assess the genetic diversity and population structure of the target species, opening a new approach for mitochondrial genome research in fish species.

### **❖ Scientific and practical significance:**

This research significantly advances our understanding of population genetics, migration patterns, and mitogenomes in Mekong fish species, providing a robust scientific foundation for effective conservation and sustainable management of regional fish resources.

## **6. Structure of the thesis**

The dissertation consists of 130 pages (excluding appendices and references), including 17 tables and 29 figures and graphs. Introduction 5 pages; conclusion and recommendations 2 pages; list of published research works 1 page; references 22 pages, including 265 English-language references consisting of scientific journal articles and books.

The main content of the thesis is divided into 03 chapters: Chapter 1: Overview Literature 36 pages; Chapter 2: Materials and methods 23 pages; Chapter 3: Results and Discussion 42 pages.

## **CHAPTER 1: OVERVIEW LITERATURE**

### **1.1. Characteristics of the Lower Mekong River Basin**

The Mekong River, with a length of 4,909 km, flows through the territories of China and Myanmar (the Upper Mekong Basin, UMB) and four countries: Laos, Thailand, Cambodia, and Vietnam (the Lower Mekong Basin, LMB). The Mekong River Basin (MRB) is one of the global biodiversity hotspots and the third most biodiverse ecosystem in the world, after the Amazon and Congo rivers. A total of approximately 1,393 known fish species in MBR, of which 293 species (21%) are identified as migratory. According to ecological characteristics and migratory patterns, fish species in the MRB are divided into three groups: whitefish, blackfish, and greyfish. Many Mekong fish species are known to migrate vertically (upstream-downstream migration) and/or horizontally (migration from rivers to floodplains) and between habitats. Migration often occurs at all developmental stages and is related to sheltering during the dry season, foraging during the flood season, and migrating for spawning as well as avoiding adverse environmental conditions. Therefore, three long-distance migration systems (upstream, midstream, and downstream) and three lateral migration systems (Tonle Sap, Sekong-Sesan-Srepok (3S), and Korat) are identified in the LMB. Hydrological factors are considered as cues for fish migration.

However, MRB is facing significant impacts from human activities (hydropower dam construction, overexploitation, population growth, urbanization, and land use changes, environmental pollution) and the

effects of climate change (temperature and rainfall variations, saltwater intrusion). These impacts not only create physical barriers that hinder migration, alter or fragment natural fish habitats, and affect the behavior of many aquatic species, but also change natural flow regimes, flood pulses, and seasonal water inputs. As a result of these changes, fisheries resources are declining and the risk of extinction is increasing. Therefore, in-depth research on genetic diversity and population structure of fish species representing characteristic ecosystems and diverse life history traits plays a crucial role in fisheries management, stock restoration, the pursuit of sustainable ecosystem maintenance, and monitoring the adaptive responses of organisms to changing environments.

## **1.2. Biological, reproductive, and migration characteristics of the fish target species**

### **1.2.1. Spiny eel *Macrognathus siamensis* Günther, 1861**

*M. siamensis* is native to the MRB, Mae Klong, Chao Phraya (Thailand), and the Malay Peninsula (Malaysia), and is an invasive species in the United States, and Singapore. *M. siamensis* belongs to the blackfish group – non-migratory and bottom-dwelling.

### **1.2.2. Black sharkminnow *Labeo chrysophekadion* Bleeker, 1850**

*L. chrysophekadion* is distributed in Southeast Asia, including MRB, Chao Phraya, Malay Peninsula, Sumatra, Java, and Borneo (Indonesia). This species typically lives in schools, concentrating in the mid-water and bottom layers. They are facultative migrants, possibly exhibiting short-distance (greyfish) or long-distance (white-fish) migration patterns. There are many populations of this species distributed in the MRB and some reservoirs in Thailand.



### ***1.2.3. Black ear catfish *Pangasius larnaudii* Bocourt, 1866***

*P. larnaudii* is distributed in medium and large rivers and floodplains in the MRB and the Chao Phraya. They belong to the whitefish and typically migrate in April–July for feeding and spawning. A downstream population of this species has been recorded, ranging from Pakse (Laos) to the Mekong Delta (Vietnam), and there is no information on the population structure in the upstream region.

## **1.3. Overview of research methodology**

### ***1.3.1. Restriction-site associated DNA sequencing (RAD-seq)***

RAD-seq technology was developed to reduce genome complexity and size by using restriction enzymes to cut the genome into short DNA fragments and perform next-generation sequencing. From this, thousands of SNPs molecular markers-DNA sequence differences between two individuals of a species at the level of a single base pair, where a single nucleotide in the genome is replaced - are identified. Among RAD-seq techniques, ezRAD has been successfully applied in population genetics studies of organisms with limited genomic information. Optimized library preparation through flexible use of restriction enzymes and commercial kits led to the selection of this technique.

### ***1.3.2. Algorithms and tools for de novo assembly***

For non-model species, *de novo* assembly in RAD-seq is utilized to generate a reduced reference genome rather than a whole genome reconstruction. This set of sequences, flanking the restriction sites, functions as a scaffold for read mapping and SNP calling, which are essential for population genetic analysis. Current bioinformatics pipelines for *de novo* assembly typically employ either *graph-based* or *greedy-based* algorithms.

### ***1.3.3. Methods for assessing genetic diversity, genetic differentiation, and population structure***

Genetic diversity is assessed through parameters such as number of alleles per locus, effective number of alleles, observed and expected heterozygosity, and inbreeding coefficient. The genetic variation of a population is determined by allele frequency and heterozygosity rate. Identification of population structure groups—clusters of individuals within the same species that exhibit genetic differentiation from other individuals in a population—is fundamental for establishing strategies for fisheries resource management.

### ***1.3.4. Methods for studying fish migration***

For fisheries-dependent methods, surveying local ecological knowledge and monitoring fishing activities are two common approaches to record population trends, assess species diversity and infer migration timing.

For fisheries-independent methods, fish migration can be determined through studies of reproductive biology, captive breeding under experimental conditions, chemical composition analysis in otoliths, tagging, the use of modern devices and genetic tools.

### ***1.3.5. Mitogenome assembly and AMS from RAD-seq data identification methods***

To sequence and assemble mitogenomes, four methods are employed, including (i) amplifying DNA fragments using PCR and Sanger sequencing, (ii) amplifying long-range DNA fragments (from 2kb) and Sanger sequencing, (iii) amplifying long-range DNA fragments and NGS, (iv) genome library preparation using restriction enzymes and NGS.

Aligned mitochondrial DNA segments derived from RAD-seq data are short, identifiable fragments within an organism's mitogenome, obtained through alignment of RAD-seq reads to the reference mitogenome sequence. These segments are used for organism classification and/or the determination of geographic variation.

#### **1.4. Research overview concerning the thesis contents**

The MRB is an ecosystem with high biodiversity, however, genomic information on fish species remains limited. With the development of NGS techniques and the application of bioinformatics tools, the genomes of several fish species in the MRB have been successfully assembled (*Pn. hypophthalmus*). Additionally, *de novo* assembly of some fish species have also been conducted in population genetics studies.

Studies on the diversity and population genetics of Mekong fish species have been conducted. In non-migratory fish groups, high genetic diversity and population structure with separation corresponding to geographic areas in the Mekong main stems in Cambodia and Laos (such as *A. testudineus* and *M. siamensis*) and in the Mekong Delta region (such as *A. testudineus*, *H. temminckii*, *P. fasciata*, and *P. melanocheilus*) have been recorded. In short-distance migratory fish groups, Ackiss et al. (2019) documented population separation of *H. spilopterus* above (Paske) and below the Khone Falls (Kratie, Stung Treng) and between the 3S river branches in Laos (Sekong), Cambodia (Sesan), and Vietnam (Serepok). High inbreeding rates and low effective population sizes in the Dak Lak (Serepok River) were recorded, with the predicted cause being the development of hydropower dams. Based on microsatellite markers, high genetic diversity and low genetic distance between populations

of *L. chrysophekadion* in Laos and Vietnam, Mashyaka and Duong (2021) hypothesized that this is a long-distance migratory fish species, approximately 1200 km. In long-distance migratory fish groups, population separation depends on the study subject, distribution area, and/or molecular markers used. For *Pn. hypophthalmus*, no differences were found between populations collected in Cambodia when using RFLP markers, or three population groups were detected in Cambodia and Vietnam using microsatellite markers, or two population groups above and below the Khone Falls using SNPs markers. Population connectivity was also observed in several other catfish species of the Pangasiidae (*Pn. gigas*, *P. bocourti*, *P. krempfi*, and *P. macronema*). For *Henicorhynchus siamensis*, population above and below the Khone Falls was detected using mitochondrial DNA markers, however, no differences were found between these two regions using microsatellite markers. For *H. lobatus*, population connectivity in the main stem was determined using mitochondrial DNA and microsatellite markers, while SNPs markers showed differentiation of populations above and below the Khone Falls.

A review of the literature shows that research on the diversity and population genetics of Mekong fish species has been conducted either separately or in combination, but only on a small scale with limited sample sizes. These studies also reveal different levels of fish population structure in the Mekong mainstreams, as well as between the mainstream and its tributaries. Simultaneously, a decline in genetic diversity, high inbreeding rates, and the isolation of populations of Mekong fish species have been documented, with overexploitation, hydropower dam development, and climate change identified as likely

causes. Additionally, a variety of molecular markers such as mitochondrial DNA, microsatellites, and SNPs have been used.

In a complex and multi-species ecosystem like the MRB, single-species management is not feasible. Comparative phylogeography and population genomics are employed to analyze population structures using genetic data across their distributions at both the gene and genome levels. These analyses contribute to clarifying the impacts leading to genetic variation, as well as evolutionary lineages according to the characteristics of geographical distribution, developmental history, and habitat fluctuations of biological species.

To date, only one mitogenome sequence of *L. chrysophekadion* has been published and used to investigate phylogenetic relationships within the Labeonini (family Cyprinidae). To expand scientific knowledge, this study assembles and annotates an additional mitogenome for *L. chrysophekadion*. Aligned mitochondrial DNA segments from RAD-seq data are identified to assess genetic diversity and population structure of *L. chrysophekadion* in the LMB.

## **CHAPTER 2. MATERIALS AND METHODS**

### **2.1. Fish target species, scope, and sampling methods**

Three target species (*M. siamensis*, *L. chrysophekadion*, and *P. larnaudii*) were collected from fishermen, local markets, or fish harbours in the MRB from 2017–2021 (**Figure 2.2**). Fish specimens were identified at the field based on morphological characteristics. A total of 272 samples of *M. siamensis* (9 populations), 263 samples of *L. chrysophekadion* (10 populations), and 193 samples of *P. larnaudii* (8 populations) were collected (**Table 2.1**). Due to insufficient sample sizes of *L. chrysophekadion* and *P. larnaudii* collected at Tacheilek, those samples were excluded from analysis.

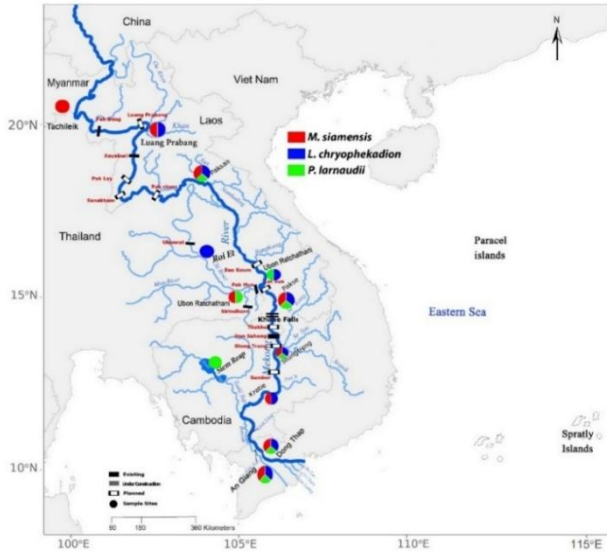


Figure 2.2. Map of sampling locations for *M. siamensis* (red color), *L. chrysophekadion* (blue), and *P. larnaudii* (green) across the MRB. “=” indicates the location of Khone Falls.

Table 2.1. Number of fish individuals collected in the MRB

Country	MRB location		Sampling sites (Code)	<i>M. siamensis</i>	<i>L. chrysophekadion</i>	<i>P. larnaudii</i>
Myamnar	UMB	Tributary	Tacheilek (TK)	32	8	1
Laos	Upper-LMB	Tributary	Luang Prabang (LP)	32	22	--
	Middle-LMB	Mainstem	Paksan (PA)	34	32	21
		Mainstem	Pakse (PE)	27	32	--
Mainstem		Ubon Ratchathani (UB-MK)	--	28	29	
Tributary		Ubon Ratchathani (UB-MR)	32	--	24	
Tributary		Roi Et (RE)	--	30	--	
Campuchia	Lower-LMB	Confluence Mekong-3S	Strung Treng (ST)	32	28	32
		Mainstem	Kratié (KT)	28	27	--
		Tributary	Siem Reap (SR)	--	--	30
Vietnam		Mainstem	Dong Thap (DT)	25	32	32
		Mainstem	An Giang (AG)	30	24	24
Total				272	263	193

## **2.2. *De novo* assembly and SNPs detection**

### **2.2.1. *EzRAD* library preparation and sequencing**

Genomic DNA was extracted from preserved tissue samples using Wizard® Genomic DNA Purification kit (Promega, USA). DNA concentrations were measured using Qubit® 2.0 Fluorometer (Invitrogen). EzRAD libraries were prepared using the TruSeq Nano HT Library Preparation kit (Illumina), following the protocols of Toonen et al. (2013) and Dang et al. (2019). The libraries were sent to the Genomics Core Laboratory (Texas A&M University, USA) for paired-end 150 bp sequencing on the Illumina HiSeq 4000 platform. Raw read sequences were uploaded to the GenBank database.

### **2.2.2. *De novo* assembly and SNPs detection**

Raw read sequences (FASTQ files) of three fish species were quality-assessed using FastQC and MultiQC. Low-quality reads and bases were then trimmed with Trimmomatic v0.33. High-quality reads were merged based on overlapping regions using zcat v1.10 to generate contigs for *de novo* assembly.

Three *de novo* assemblers - dDocent 4.5, ipyrad 0.9.84 (both greedy algorithms), and Stacks 2.4 (graph-based) - were evaluated using a *P. larnaudii* contig dataset. Assembly metrics, including the number of contigs, N50, largest cluster size, genome size, and read alignment percentages, were calculated using awk v5.1.0, grep v3.4, and Samtools v1.9, respectively, to evaluate *de novo* assembly results. The optimal assembler was then used for *de novo* assembly of *M. siamensis* and *L. chrysophekadion*.

Raw SNPs were detected with FreeBayes, and subsequent filtering was performed using vcftools and vcfilter. Outlier loci, identified using Lositan Selection Workbench and BayeScan v2.1, were removed to

obtain neutral SNPs. Neutral and adaptive SNP datasets (in .vcf format) were used to analyze genetic diversity and population structure.

### **2.3. Genetic diversity, population structure, and migration pattern prediction of three typical fish species in the LMB**

#### **2.3.1. Genetic diversity**

Genetic diversity was assessed using GenAlEx v6.5.2, calculating observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity for each population and all individuals combined.

#### **2.3.2. Population structure**

Genetic differentiation was assessed using Arlequin v3.5, calculating overall ( $G_{ST}$ ) and pairwise ( $F_{ST}$ ) differentiation between populations, and performing AMOVA to determine differentiation within and among populations and subgroups. Population genetic structure was then analyzed using STRUCTURE and PCA. Inbreeding coefficients ( $G_{IS}$ ) were then determined using GenoDive v2.0b27, based on population clustering from STRUCTURE and PCA analyses. To test correlations between genetic differentiation and geographic distance, Mantel and db-MEM analyses were applied.

#### **2.3.3. Migration pattern prediction**

To test hypotheses regarding population structure based on recorded migration systems, migration patterns of two fish species with different migratory behaviors-*L. chrysophekadion* (short-distance, facultative) and *P. larnaudii* (long-distance) - were predicted using divMigrate and TreeMix v1.13.

### **2.4. AMS from RAD-seq data identification of *Labeo chrysophekadion***

#### **2.4.1. Mitochondrial genome assembly and annotation**

Using high-quality reads, the mitochondrial genome of *L. chrysophekadion* was assembled and annotated according to the



MitoZ v3.4 pipeline. Protein-coding genes, transfer RNAs, and ribosomal RNAs were annotated using GeneWise v2.2, MiTFi v1.0, and infernal v1.1.1, respectively. BWA v0.7.17 and Samtools v1.15.1 were employed to organize the genes according to the mitochondrial genome structure, and the circular structure was visualized using Circos. The mitochondrial genome was deposited in GenBank.

#### ***2.4.2. AMS from RAD-seq data identification and population genetics analysis in *L. chrysophekadion****

AMS from RAD-seq data of *L. chrysophekadion* were determined using the radBARCODER pipeline. AMS data were then used to assess genetic diversity of *L. chrysophekadion* in the LMB by total number of haplotypes (Nh), haplotype diversity (Hd), nucleotide diversity ( $\pi$ ), and the number of polymorphic sites (S) using DnaSP v5. Pairwise genetic differentiation ( $F_{ST}$ ) between populations was determined using Arlequin v3.5. A haplotype network was constructed using the Templeton Crandall and Sing algorithm on PopART software.

### **CHAPTER 3. RESULTS AND DISCUSSION**

#### **3.1. *De novo* assembly and SNPs detection**

##### ***3.1.1. EzRAD library preparation and sequencing***

A total of **272** DNA libraries from *M. siamensis*, **255** libraries from *L. chrysophekadion*, and **192** libraries from *P. larnaudii* were sequenced, yielding **36,447,827**; **1,062,049,264**; and **313,556,27 raw reads**, respectively. After removing low-quality nucleotides, adapters, and reads shorter than 50 bp, the number of high-quality reads obtained were **28,733,914** (78.8%) for *M. siamensis*; **1,026,721,048** (96.6%) for *L. chrysophekadion*; and **259,912,528** (82.9%) for *P. larnaudii*.

### **3.2.2. De novo assembly and quality evaluation**

The dDocent tool, utilizing a greedy algorithm, was selected for *de novo* assembly. The resulting assembled genome sizes were **30,372,557** bp (36.1% GC content) for *M. siamensis*, **32,975,632** bp (35.9% GC) for *L. chrysophekadion*, and **35,729,239** bp (35.1% GC) for *P. larnaudii*.

### **3.2.3. SNPs detection**

Following detection and filtering, the study selected **4,237** SNPs from 239 individuals (9 populations) of *M. siamensis*, **825** SNPs from 232 individuals (9 populations) of *L. chrysophekadion*, and **1,270** SNPs from 160 individuals (7 populations) of *P. larnaudii*. Employing BayeScan and Lositan to test for outlier loci, the research yielded a neutral SNP dataset comprising **3,736** SNPs in *M. siamensis*, **760** SNPs in *L. chrysophekadion*, and **1,176** SNPs in *P. larnaudii*. Datasets of adaptive loci, containing **189**, **23**, and **11** loci for *M. siamensis*, *L. chrysophekadion*, and *P. larnaudii*, respectively, were also obtained.

## **3.3. Genetic diversity, population structure, and migration pattern prediction of three typical fish species in the LMB**

This study analyzed genetic diversity, population structure, and AMOVA in three fish species using both neutral SNPs and adaptive loci. However, due to the lack of clear evidence for environmental adaptation, results from adaptive loci were considered non-informative. Therefore, the results section presents only analyses based on neutral SNP data.

### **3.3.1. Genetic diversity**

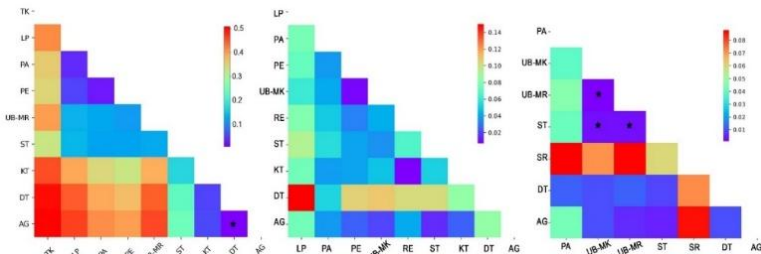
Genetic diversity analysis showed that  $H_o$  was consistently equal to or lower than  $H_e$  across all fish populations. *P. larnaudii* exhibited the highest genetic diversity ( $H_o/H_e=0.220/0.266$ ), followed by *L. chrysophekadion* ( $H_o/H_e = 0.185/0.233$ ), and *M. siamensis* ( $H_o/H_e = 0.140/0.176$ ) (**Table 3.4**).

### 3.3.2. Population structure of fish species in the LMB

Overall genetic differentiation ( $G_{ST}$ ) was highest in *M. siamensis* ( $G_{ST} = 0.303$ ), followed by *L. chrysophekadion* ( $G_{ST} = 0.038$ ), and lowest in *P. larnaudii* ( $G_{ST} = 0.003$ ), and significant ( $P < 0.001$ ). Pairwise genetic differentiation ( $F_{ST}$ ) showed statistically significant differences in 27/28 *M. siamensis* population pairs, 36/36 *L. chrysophekadion* pairs, and 18/21 *P. larnaudii* pairs.

Regarding inbreeding coefficients ( $G_{IS}$ ), the *M. siamensis* subpopulation in UB-MR had the highest value ( $G_{IS} = 0.61$ ), while the lower-LMB region had the lowest ( $G_{IS} = 0.40$ ). *L. chrysophekadion* subpopulations showed similar  $G_{IS}$ , ranging from 0.225 (upper-LMB) to 0.299 (lower-LMB), with 0.245 in the middle-LMB. The overall  $G_{IS}$  for *P. larnaudii* was 0.173.

For *M. siamensis*, populations downstream of the Khone falls had higher genetic differentiation than upstream. The *L. chrysophekadion* from Dong Thap and Luang Prabang exhibited considerable genetic differentiation relative to other pairs. *P. larnaudii* in Siem Reap showed the greatest differentiation, followed by Paksan, compared to the remaining population pairs (**Figure 3.3**).



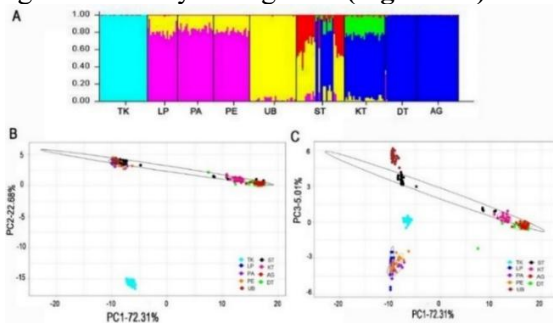
**Figure 3.3. Heatmap depicting pairwise population genetic differentiation ( $F_{ST}$ ) for *M. siamensis* (A), *L. chrysophekadion* (B), and *P. larnaudii* (C). \* Indicates non-significant differences after FDR correction**

**Table 3.4. Genetic diversity parameters in the populations of three typical fish species in the LMB**

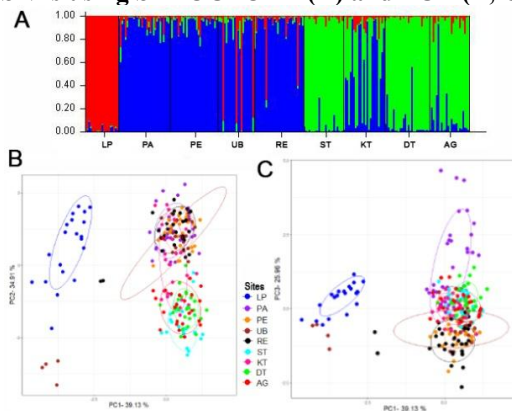
Mekong location			Sampling sites	<i>M. siamensis</i>			<i>L. chrysophekadion</i>			<i>P. larnaudii</i>		
				Nse	Ho	He	Nse	Ho	He	Nse	Ho	He
UMB		Tributary	TK	32	0.140	0.140						
LMB	Upper-LMB	Tributary (Khan River)	LP	20	0.128	0.141	20	0.193	0.208			
	Middle-LMB	Mainsteam	PK	24	0.110	0.142	31	0.243	0.248	20	0.263	0.298
			PE	24	0.108	0.146	29	0.217	0.245			
			UB-MK				22	0.214	0.245	26	0.259	0.271
		Tributary (Mun River)	UB-MR	31	0.106	0.125				15	0.268	0.266
		Tributary (Chi River)	RE				30	0.140	0.234			
	Lower-LMB	Confluence Mekong-3S	ST	32	0.146	0.236	24	0.238	0.238	28	0.230	0.270
		Tributary	KT	27	0.206	0.239	25	0.143	0.237			
		Biển Hồ	SR							22	0.225	0.261
		Mainsteam	AG	28	0.153	0.202	24	0.185	0.228	19	0.235	0.265
	Mainsteam	DT	21	0.158	0.210	27	0.139	0.222	30	0.223	0.269	
Total/Mean				239	0.140	0.176	232	0.185	0.233	160	0.220	0.266

Population coded are shown in Table 3.1; Nse: Number of analyzed samples, Ho: Observed heterozygosity, He: Expected heterozygosity; Gis: Inbreeding coefficient.

Based on STRUCTURE and PCA analyses, the *M. siamensis* populations were divided into 5 groups: (1) Tachileik, (2) Luang Prabang, Paksan, and Pakse, (3) Ubon Ratchathani (Mun River), (4) a mixture of genetic information at Strung Treng, and (5) Kratié to the Mekong Delta (**Figure 3.4**). Three subgroups of *L. chrysophekadion* were identified: (1) Luang Prabang, (2) Paksan, Pakse, Ubon Ratchathani, Roi Et, and (3) Stung Treng, Kratié, Dong Thap, and An Giang (**Figure 3.5**). In *P. larnaudii*, the indistinct separation of populations was observed, indicating high connectivity among them (**Figure 3.6**).

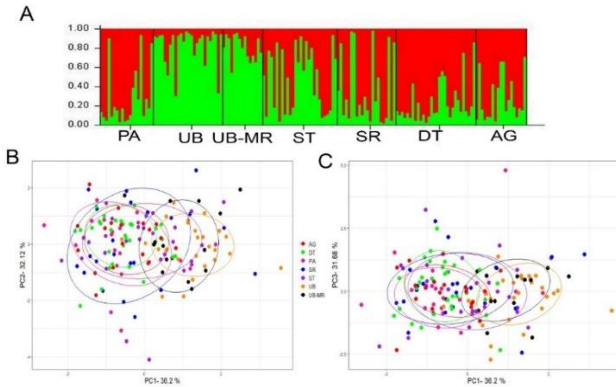


**Figure 3.4. Population structure of *M. siamensis* based on 1,936 neutral SNPs using STRUCTURE (A) and PCA (B, C)**



**Figure 3.5. Population structure of *L. chrysophekadion* based on 760 neutral SNPs using STRUCTURE (A) and PCA (B, C)**

Mantel test analysis revealed a statistically significant correlation between increased geographic distance and genetic differentiation in *M. siamensis* populations and *L. chrysophekadion* populations, whereas *P. larnaudii* populations did not show this correlation. For dbMEM analysis, the MEM-1 axis indicated spatial structure separation of populations in all three fish species, while MEM 2-6 showed no significant.

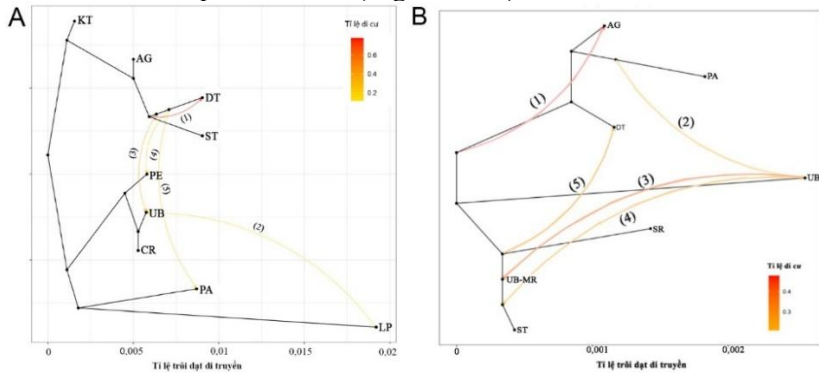


**Figure 3.6. Population structure of *P. larnaudii* based on 1,176 neutral SNPs using STRUCTURE (A) and PCA (B, C)**

AMOVA analysis of *M. siamensis* revealed significant genetic differentiation between populations above and below Khone Falls (31.5%;  $P = 0.03$ ). Variation at other levels was also significant, with 49.1% occurring among individuals, 9.5% among individuals within populations, and 9.8% among populations within groups. For *L. chrysophekadion*, the primary source of variation was among individuals (50% and 84.49%;  $P < 0.001$ ). *P. larnaudii* showed 57% variation among individuals, 39% among individuals within populations, and only 3% among populations. No significant genetic differentiation was observed between groups in *P. larnaudii* (0.4%;  $P > 0.05$ ).

### 3.3.3. Predicting fish migration patterns in the LMB

DivMigrate analysis revealed bidirectional migration (both downstream and upstream) in *L. chrysophekadion* and *P. larnaudii*, with *P. larnaudii* exhibiting a higher migration rate. Notably, upstream migration across Khon Falls was observed in both species. Treemix analysis indicated that the optimal number of migration events for both species was 5 (Figure 3.10).



**Figure 3.10. Treemix-generated phylogenetic trees showing population divergence and admixture in *L. chrysophekadion* (A) and *P. larnaudii* (B). Arrows indicate the direction of gene flow, and the color and thickness of the lines represent migration rates**

In *L. chrysophekadion*, (1) a downstream occurred from an intermediate point upstream of AG to DT, (2) a downstream from LP to UB, and three upstream from an intermediate point between DT and ST to populations upstream of Khone Falls, including UB (3), PE (4), and PA (5). In *P. larnaudii*, (1) downstream occurred from an intermediate point above and below Khone Falls to AG, (2) downstream from an intermediate point at PA to UB; (3) lateral and upstream from UB to UB-MR; (4) lateral and downstream from UB to an intermediate point

between UB-MR and ST; and (5) downstream from an intermediate point above and below Khone Falls to DT (**Figure 3.10**).

### **3.4. Mitogenome assembly and annotation, AMS from RAD-seq data, population structure investigation of *L. chrysophekadion***

#### **3.4.1. Mitogenome assembly and annotation**

The mitochondrial genome of *L. chrysophekadion*, collected in Ubon Ratchathani, was successfully assembled (OR637878). The assembled genome is 16,600 base pairs in length (42.9% GC), and exhibits 99.8% similarity to other genomes of the same species in GenBank. The annotated genome includes 37 genes: 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNA), 2 ribosomal RNA genes (rRNA), and 1 non-coding region (D-loop).

#### **3.4.2. AMS from RAD-seq data identification, and population structure investigation of *Labeo chrysophekadion***

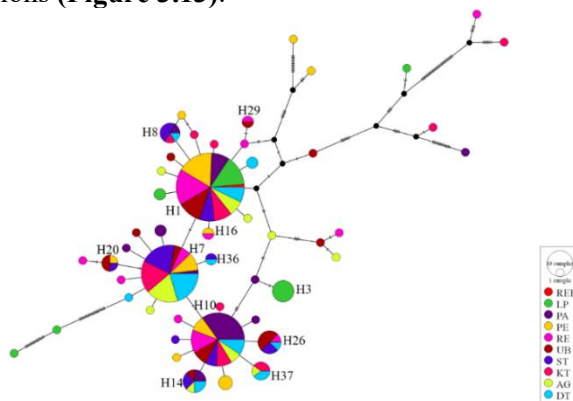
This study identified a 757 bp of aligned mitogenome segment data. Among these, 502 bp (66.2%) were found in PCGs (*NDI*, *COXI*, *COXIII*, and *cytb*), 79 bp in rRNA (*16S*), and the remaining in tRNA genes (*tRNA<sup>Leu</sup>*, *tRNA<sup>Asn</sup>*, *tRNA<sup>Cys</sup>*, *tRNA<sup>Tyr</sup>*, and *tRNA<sup>Thr</sup>*).

A total of 49 haplotypes (19.8%) were identified using AMS data (757 bp) from 247 *L. chrysophekadion* individuals, ranging from 6 haplotypes/22 individuals (27.3%) in Luang Prabang to 12 haplotypes/30 individuals (40%) in Roi Et. The results indicated high haplotype diversity ( $H_d=0.849 \pm 0.014$ ). Low pairwise genetic differentiation ( $F_{ST}=0-0.17$ ) was observed among populations, with no significant differences, except for the Luang Prabang.

The haplotype network revealed high genetic connectivity among *L. chrysophekadion* populations in the LMB. Notably, haplotype H1 was shared by all populations, haplotypes H7 and H10 were found in



8 out of 9 populations (excluding Luang Prabang), and haplotype H14 was shared by 5 out of 9 populations. Haplotypes H8 and H26 were shared by 4 out of 9 populations, and haplotypes H16, H20, H29, H36, and H37 were observed in sequences from 2-3 populations. A large number of unique haplotypes (35/49) were found across all populations (**Figure 3.13**).



**Figure 3.13. Haplotype network of *L. chrysophekadion* populations in the LMB using AMS data. Lines connecting circles represent one mutational step. Dashed lines indicate additional mutational steps. The size of the circles represents the number of sequences. Colors correspond to each sampling population and reference sequences from GenBank.**

## CONCLUSIONS AND RECOMMENDATIONS

### Conclusions

1. This study successfully collected and prepared genomic libraries for three typical fish species of the MBR: *M. siamensis* (272 samples, 9 populations), *L. chrysophekadion* (255 samples, 9 populations), and *P. larnaudii* (192 samples, 7 populations). *De novo* genome assemblies were generated for each species. SNP analysis revealed 4,237 species-specific SNPs for *M. siamensis*, 825 for *L. chrysophekadion*; and 1,270 for *P. larnaudii*.

2. The genetic diversity, genetic differentiation, and population structure of three typical fish species in the Mekong River were determined. Additionally, the migration patterns of *L. chrysophekadion* and *P. larnaudii* in the LMB were predicted using SNP markers. Specifically:

- Genetic diversity positively correlated with fish migration, while inbreeding coefficient negatively correlated. Non-migratory *M. siamensis* showed the lowest diversity ( $H_o/H_e=0.140/0.176$ ) and highest inbreeding ( $G_{IS}=0.40-0.61$ ); long-distance migratory *P. larnaudii*, the highest diversity ( $H_o/H_e=0.220/0.266$ ) and lowest inbreeding ( $G_{IS}=0.173$ ); and short-distance/facultative migratory *L. chrysophekadion*, intermediate levels ( $H_o/H_e=0.185/0.233$ ;  $G_{IS}=0.225-0.299$ ).

- Analysis of genetic differentiation and population structure revealed distinct population differentiation among the fish species. *M. siamensis* showed four population groups corresponding to the Upper Mekong Basin, the upper and middle LMB, tributaries in the middle LMB, and the lower LMB. *L. chrysophekadion* exhibited three population groups, corresponding to the upper, middle, and lower LMB. In contrast, *P. larnaudii* populations showed high genetic connectivity.

- + Migration model predictions for both *L. chrysophekadion* and *P. larnaudii* indicated significant migration capabilities, including downstream, upstream, and passage over the Khone Falls. However, a notable observation was that short-distance/facultative migratory fish exhibited longer-than-expected migration routes, while long-distance migratory fish showed discontinuous migration, mainly limited between the lower and middle LMB.

3. The mitogenome of *L. chrysophekadion* (Genbank accession number OR637878) was assembled and annotated, with a length of

16,600 bp. Using AMS data (757 bp), the results showed high genetic connectivity among populations, with high haplotype diversity and low nucleotide diversity. No significant genetic differentiation was observed among populations, except Luang Prabang.

### **Recommendations**

#### **❖ Management and conservation**

- Transition to genetic-Based management: Establish conservation boundaries based on population clusters (04 regions for *Acantopsis* sp., 03 regions for *Labeo chrysophekadion*, and 01 region for *Pangasius larnaudii*) to ensure the protection of endemic diversity within each geographical area.

- Maintain connectivity at ecological functions: Prioritize the protection of river segments surrounding the Khone Falls and the middle-to-lower corridors to maintain transboundary gene flow, particularly for white and grey fish species.

#### **❖ Future research**

- Integrate methodologies and scale expansion: Combine SNP data with otolith analysis and electronic tagging, while expanding sampling sites across diverse ecological zones (e.g., Siem Reap, Luang Prabang, and the 3S river system). This will allow for a comparison between “historical gene flow” and “actual migratory behavior” to clarify existing contradictions regarding identified migration routes.

- Impact assessment and long-term monitoring: Establish periodic monitoring programs for effective population size using eDNA or SNPs to provide early warnings for populations at risk of extinction due to inbreeding. Additionally, conduct in-depth research on the impacts of hydroelectric dams and climate change on population structures.

## LIST OF PUBLISHED PUBLICATIONS

1. **Truong, O.T.**, Tran, S.Q., Carpenter, K.E., Vu, Q.D.H., Duong, T.-Y., Kyaw, M.M., Grudpan, C., Thai Bich, V.N., Dang, B.T (2025). Population genetics of *Macrognathus siamensis* (Synbranchiformes: Mastacembelidae): Implications for non-migratory fishery resources in the Mekong River basin. **Fisheries Research**, **281**:107210. <https://doi.org/10.1016/j.fishres.2024.107210>.

2. **Oanh Truong Thi**, Sang Quang Tran, Kent E. Carpenter, Ut Ngoc Vu, Sophorn Uy, Chaiwut Grudpan, Phounvisouk Latsamy, Binh Thuy Dang (2025). Population genetics and the importance of migration in the facultative migratory fish, *Labeo chrysophekadion*, in the Lower Mekong Basin. **Conservation Genetics**, **26**, 307–318. <https://doi.org/10.1007/s10592-024-01668-w>.

3. **Oanh Thi Truong**, Sang Quang Tran, Van Ngo Thai Bich, Binh Thuy Dang (2024). Population genetics of the Black Sharkminnow (*Labeo chrysophekadion*) in the Lower Mekong Basin based on mitochondrial DNA segments from RAD-seq. **VNUHCM Journal of Science and Technology Development**, **26** (SI), 25-37. <https://doi.org/10.32508/stdj.v26iSI.4199>.

4. **Oanh Thi Truong**, Sang Quang Tran, Quyen Dang Ha Vu, Van Ngo Thai Bich, Binh Thuy Dang (2022). Comparative tools for *de novo* genome assembly: Apply in population genetics of Mekong fish species, *Pangasius larnaudii* (Siluriformes: Pangasiidae). **Proceeding books in 7<sup>th</sup> Asia Pacific International Modern Sciences Congress**, Jakarta, Indonesia, 363 – 372. ISBN 978-625-8246-59-9.

5. **Trương Thị Oanh**, Ngô Thái Bích Vân, Đặng Thủy Bình (2024). Dự đoán mô hình di cư cá vồ đém *Pangasius larnaudii* (Siluriformes: Pangasiidae) ở Hạ lưu sông Mekong. **Kỷ yếu Hội nghị Công nghệ Sinh học toàn quốc**, 860 – 866. ISBN 978-604-489-393-8.

6. **Oanh Thi Truong**, Sang Quang Tran, Quyen Vu Dang Ha, Van Ngo Thai Bich, Binh Thuy Dang. Genetic Differentiation and Isolation by Distance in Mekong River Fishes with Typical Migration Patterns. **Asian Fisheries Science**, **38(3)**, 131-142.