

ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC BÁCH KHOA
📖

TRƯƠNG THỊ OANH

PHÁT HIỆN VÀ NGHIÊN CỨU ỨNG DỤNG
CÁC CHỈ THỊ ĐA HÌNH NUCLEOTIDE ĐƠN (SNPs)
BẰNG KỸ THUẬT EzRAD CỦA BA LOÀI CÁ
ĐIỂN HÌNH Ở LƯU VỰC HẠ LƯU SÔNG MEKONG

LUẬN ÁN TIẾN SĨ KỸ THUẬT

ĐÀ NẴNG – 2026

ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC BÁCH KHOA


TRƯỜNG THỊ OANH

**PHÁT HIỆN VÀ NGHIÊN CỨU ỨNG DỤNG
CÁC CHỈ THỊ ĐA HÌNH NUCLEOTIDE ĐƠN (SNPs)
BẰNG KỸ THUẬT EzRAD CỦA BA LOÀI CÁ
ĐIỂN HÌNH Ở LƯU VỰC HẠ LƯU SÔNG MEKONG**

Chuyên ngành: Công nghệ Sinh học

Mã số: 9420201

LUẬN ÁN TIẾN SĨ KỸ THUẬT

NGƯỜI HƯỚNG DẪN KHOA HỌC:

PGS. TS. ĐẶNG THÚY BÌNH

TS. NGÔ THÁI BÍCH VÂN

ĐÀ NẴNG – 2026

LỜI CAM ĐOAN

Tôi xin cam đoan luận án là công trình khoa học do tôi thực hiện dưới sự hướng dẫn của tập thể giáo viên hướng dẫn. Luận án sử dụng thông tin trích dẫn từ nhiều nguồn tham khảo khác nhau và các thông tin trích dẫn được ghi rõ nguồn gốc. Các kết quả nghiên cứu được công bố cùng với các đồng tác giả và được sự nhất trí của các tác giả khi đưa vào luận án. Các số liệu, kết quả được trình bày trong luận án là hoàn toàn trung thực và chính xác và là một phần trong dự án PEER 6-435 “***Di truyền học tổng quan về lịch sử tự nhiên của các loài cá khai thác ở lưu vực sông Mekong***” do PGS. TS. Đặng Thúy Bình chủ trì.

Tôi xin chịu hoàn toàn trách nhiệm về tính xác thực của lời cam đoan này.

Đà Nẵng, ngày tháng năm 2026

Nghiên cứu sinh

Trương Thị Oanh

LỜI CẢM ƠN

Trong suốt quá trình học tập tại Trường Đại học Bách khoa - Đại học Đà Nẵng và thực hiện luận án tại Trường Đại học Nha Trang, em đã nhận được sự quan tâm và hướng dẫn tận tình của nhà trường, các thầy cô giáo, các đồng nghiệp, bạn bè và gia đình.

Nhân đây, cho phép em được bày tỏ lòng kính trọng và biết ơn sâu sắc tới PGS. TS. Đặng Thúy Bình và TS. Ngô Thái Bích Vân đã tận tình hướng dẫn, dành nhiều công sức, thời gian và tạo điều kiện cho em trong suốt quá trình học tập, thực hiện luận án và công bố các công trình nghiên cứu.

Em cảm ơn TS. Vũ Đặng Hạ Nguyên (Giảng viên Viện Công nghệ Sinh học và Môi trường) và ThS. Trần Quang Sáng (Thành viên nhóm nghiên cứu “Đa dạng Sinh học và Bảo tồn”) tại Trường Đại học Nha Trang đã giúp đỡ và hỗ trợ về mọi mặt, từ những đóng góp về mặt chuyên môn học thuật cho đến những chia sẻ về cuộc sống hằng ngày.

Em xin gửi lời cảm ơn đến GS. Kent Carpenter (Đại học Old Dominion, Mỹ) đã hướng dẫn và đồng hành cùng em trong các chuyến thu mẫu thực địa, hỗ trợ em trong việc phân tích dữ liệu và công bố các công trình nghiên cứu.

Em xin cảm ơn quý thầy cô trong các hội đồng bảo vệ tổng quan, các chuyên đề nghiên cứu và đánh giá ở đơn vị chuyên môn đã dành thời gian quý báu để đọc và đóng góp những ý kiến đánh giá để luận án này của em được hoàn thiện hơn.

Em xin bày tỏ lòng biết ơn chân thành tới Ban Giám hiệu, các thầy cô ở Phòng Đào tạo, Khoa Hóa, Bộ môn Công nghệ Sinh học trường Đại học Bách Khoa, Đại học Đà Nẵng đã tận tình giúp đỡ, tạo điều kiện và hỗ trợ em trong quá trình học tập, thực hiện và hoàn thành luận án.

Em cũng xin gửi lời cảm ơn đến các đối tác của dự án PEER, GS. TS. Vũ Ngọc Út, PGS. TS. Dương Thúy Yên, ThS. Nguyễn Thị Ngọc Trân (Trường Đại học Cần Thơ), TS. Phounvisouk Latsamy (Lào), ThS. Sophorn Uy và ThS. Samol Chhuoy (Đại học Hoàng gia, Campuchia), TS. Chaiwut Grudpan (Đại học Ubon Ratchathani, Thái Lan), TS. Mie Mie Kyaw (Đại học Mandalay, Myanmar), TS. Ellen E. Biesack và TS. Amanda S. Ackiss (Đại học Old Dominion, Mỹ), TS. Christopher Bird và

Sharon F. Magnuson (Đại học Texas A&M, Corpus Christi, Mỹ) đã giúp đỡ, chia sẻ kinh nghiệm và hỗ trợ em trong quá trình nghiên cứu từ thực địa đến thực hiện thư viện DNA và xử lý số liệu.

Gửi lời cảm ơn ba mẹ, gia đình và bạn bè thân thiết đã luôn ủng hộ và động viên trên con đường học tập và nghiên cứu khoa học.

Và cuối cùng, em xin chân thành cảm ơn đến nguồn tài trợ kinh phí học tập và nghiên cứu của tổ chức USAID thông qua dự án PEER và Quỹ Đổi mới sáng tạo Vingroup (VINIF).

Trong quá trình thực hiện luận án, do kiến thức và thời gian nghiên cứu của em còn hạn chế nên không thể tránh khỏi những thiếu sót. Em mong nhận được những đóng góp ý kiến quý báu của thầy cô để luận án của em được hoàn thiện hơn.

Đà Nẵng, ngày tháng năm 2026

Nghiên cứu sinh

Trương Thị Oanh

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC	iv
DANH MỤC HÌNH ẢNH	vii
DANH MỤC BẢNG	ix
DANH MỤC CHỮ VIẾT TẮT	x
GIẢI THÍCH CÁC THUẬT NGỮ	xii
MỞ ĐẦU	1
1. Lý do chọn đề tài	1
2. Mục tiêu nghiên cứu	3
2.1. Mục tiêu chung	3
2.2. Mục tiêu cụ thể	3
3. Nội dung nghiên cứu	3
4. Những đóng góp mới của luận án	4
5. Ý nghĩa khoa học và thực tiễn của luận án	5
6. Cấu trúc của luận án	5
CHƯƠNG 1: TỔNG QUAN TÀI LIỆU	6
1.1. Tổng quan về lưu vực hạ lưu sông Mekong	6
1.1.1. Đặc điểm địa lý, khí hậu và thủy văn của lưu vực hạ lưu sông Mekong ..	6
1.1.2. Các hệ thống di cư ở hạ lưu sông Mekong	8
1.1.3. Nguồn lợi thủy sản và các yếu tố ảnh hưởng đến đa dạng sinh học ở lưu vực hạ lưu sông Mekong	10
1.2. Đặc điểm sinh học, sinh sản và di cư của các đối tượng nghiên cứu	12
1.2.1. Cá chạch lá tre <i>Macrognathus siamensis</i> Günther, 1861	12
1.2.2. Cá ét mội <i>Labeo chrysophekadion</i> Bleeker, 1850	13
1.2.3. Cá vồ đém <i>Pangasius larnaudii</i> Bocourt, 1866	16
1.3. Tổng quan về phương pháp nghiên cứu	18
1.3.1. Kỹ thuật giải trình tự DNA dựa trên vị trí cắt hạn chế	18

1.3.2. Các công cụ lắp ráp <i>de novo</i> hệ gen sử dụng dữ liệu RAD-seq	19
1.3.3. Chỉ thị đa hình nucleotide đơn (SNPs)	22
1.3.4. Phương pháp đánh giá đa dạng di truyền, khác biệt di truyền và cấu trúc quần thể	23
1.3.5. Các phương pháp nghiên cứu sự di cư của cá	25
1.3.6. Lắp ráp, chú giải hệ gen ti thể và xác định mã vạch RAD	28
1.4. Tình hình nghiên cứu liên quan đến các nội dung của luận án	32
1.4.1. Lắp ráp <i>de novo</i> hệ gen và xác định chỉ thị phân tử SNPs của các loài cá phân bố ở lưu vực sông Mekong.....	32
1.4.2. Nghiên cứu về đa dạng di truyền quần thể và mô hình di cư của các loài cá ở lưu vực hạ lưu sông Mekong.....	34
1.4.3. Các nghiên cứu về lắp ráp hệ gen ti thể và xác định mã vạch RAD của các loài cá thuộc giống <i>Labeo</i>	39
1.5. Cơ sở khoa học của luận án.....	40
CHƯƠNG 2: ĐỐI TƯỢNG, PHẠM VI VÀ PHƯƠNG PHÁP NGHIÊN CỨU ..	42
2.1. Đối tượng, phạm vi nghiên cứu và phương pháp thu mẫu.....	42
2.2. Lắp ráp <i>de novo</i> hệ gen của các loài cá và xác định chỉ thị SNPs	47
2.2.1. Xây dựng thư viện DNA các loài cá sử dụng kỹ thuật EzRAD và giải trình tự	47
2.2.2. Lắp ráp <i>de novo</i> hệ gen của các loài cá và xác định chỉ thị SNPs	51
2.2.3. Xác định chỉ thị SNPs của các loài cá.....	54
2.3. Khảo sát sự đa dạng di truyền, cấu trúc quần thể và dự đoán mô hình di cư của ba loài cá điển hình ở lưu vực hạ lưu sông Mekong	58
2.3.1. Khảo sát sự đa dạng di truyền của các quần thể cá.....	58
2.3.2. Xây dựng cấu trúc di truyền quần thể các loài cá ở hạ lưu sông Mekong...	58
2.3.3. Dự đoán mô hình di cư của các loài cá ở hạ lưu sông Mekong.....	61
2.4. Lắp ráp và chú giải hệ gen ti thể, xác định mã vạch RAD và khảo sát cấu trúc di truyền quần thể của cá ét mọi <i>Labeo chrysophekadion</i>	63
2.4.1. Lắp ráp và chú giải hệ gen ti thể cá ét mọi	63

2.4.2. Xác định mã vạch RAD thuộc hệ gen ti thể và khảo sát cấu trúc di truyền quần thể của cá ét mọi ở hạ lưu sông Mekong.....	64
CHƯƠNG 3: KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN.....	66
3.1. Lắp ráp <i>de novo</i> hệ gen của ba loài cá điển hình ở lưu vực hạ lưu sông Mekong và xác định các chỉ thị phân tử SNPs.....	66
3.1.1. Tạo thư viện hệ gen của ba loài cá và giải trình tự thế hệ mới.....	66
3.1.2. Lắp ráp <i>de novo</i> và đánh giá chất lượng hệ gen của ba loài cá	67
3.1.3. Xác định chỉ thị phân tử SNPs của ba loài cá điển hình ở lưu vực hạ lưu sông Mekong	71
3.2. Khảo sát sự đa dạng di truyền, cấu trúc quần thể và dự đoán mô hình di cư của các loài cá điển hình ở lưu vực hạ lưu sông Mekong	75
3.2.1. Sự đa dạng di truyền của các quần thể cá.....	75
3.2.2. Cấu trúc di truyền quần thể của các loài cá ở LMB	78
3.2.3. Dự đoán mô hình di cư của các loài cá ở LMB.....	88
3.2.4. Thảo luận	91
3.3. Lắp ráp và chú giải hệ gen ti thể, xác định mã vạch RAD và khảo sát cấu trúc di truyền quần thể cá ét mọi <i>Labeo chrysophekadion</i>	101
3.3.1. Lắp ráp và chú giải hệ gen ti thể của cá ét mọi	101
3.3.2. Xác định mã vạch RAD thuộc hệ gen ti thể và khảo sát di truyền quần thể cá ét mọi ở LMB	104
KẾT LUẬN VÀ KIẾN NGHỊ	109
Kết luận	109
Kiến nghị	110
DANH SÁCH CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ	111
TÀI LIỆU THAM KHẢO	112

DANH MỤC HÌNH ẢNH

Hình 1.1. Hệ thống lưu vực sông Mekong.....	7
Hình 1.2. Các hệ thống di cư (A) [34] và các đập thủy điện trên dòng chính và dòng nhánh (B) [35] ở sông Mekong.....	9
Hình 1.3. Sơ đồ phân bố vực sâu và số lượng loài (A), vùng ngập (màu vàng, B) và số lượng loài cá mang trứng (màu vàng, C) ở dòng chính sông Mekong [7].....	9
Hình 1.4. Hình thái ngoài (A) [66] và bản đồ phân bố (chấm màu vàng, B) [58] của cá chạch lá tre <i>Macrognathus siamensis</i>	13
Hình 1.5. Hình thái ngoài (A), bản đồ phân bố (chấm màu tím, B) [72] và hướng di cư theo mùa (C) [6] của cá ét mọi <i>Labeo chrysophekadion</i>	15
Hình 1.6. Hình thái ngoài (A), bản đồ phân bố (chấm màu tím, B) [76] và hướng di cư theo mùa (C) [6] của cá vồ đém <i>Pangasius larnaudii</i>	17
Hình 1.7. Mô phỏng các bước thực hiện của thuật toán dựa trên đồ thị trong lắp ráp <i>de novo</i> hệ gen.	20
Hình 1.8. Mô phỏng các bước thực hiện của thuật toán tham lam trong lắp ráp <i>de novo</i> hệ gen	21
Hình 1.9. Hình minh họa về chỉ thị đa hình nucleotide đơn (SNP).....	22
Hình 1.10. Tỷ lệ Sr:Ca từ lúc cá mới đẻ (tâm đá tai) đến lúc thu mẫu (rìa bên ngoài)	26
Hình 2.1. Sơ đồ khối thực hiện ba nội dung nghiên cứu của luận án	42
Hình 2.2. Bản đồ các địa điểm thu mẫu các loài cá trong nghiên cứu hiện tại.....	45
Hình 2.3. Quy trình tạo thư viện hệ gen bằng kỹ thuật EzRAD	48
Hình 2.4. Cấu trúc của adapter.....	50
Hình 2.5. Hình ảnh minh họa file trình tự với định dạng FASTQ.....	51
Hình 2.6. Sơ đồ khối quy trình xử lý số liệu.....	52
Hình 3.1. Hình ảnh minh họa kết quả điện di DNA tổng số (A) và thư viện hệ gen (B) của một số mẫu cá trên gel agarose 1,5%	66
Hình 3.2. Đồ thị thể hiện kết quả kiểm định loci ngoại vi của cá chạch lá tre (A, B), ét mọi (C, D) và vồ đém (E, F) theo BayeScan (A, C, E) và Lositan (B, D, F)	74

Hình 3.3. Biểu đồ heatmap thể hiện giá trị khác biệt di truyền theo cặp quần thể (F_{ST}) của cá chạch lá tre (A), cá ét mọi (B) và cá vô đém (C). * chỉ ra sự khác biệt không có ý nghĩa thống kê sau khi hiệu chỉnh FDR	79
Hình 3.4. Cấu trúc quần thể cá chạch lá tre dựa trên phân tích nhóm Structure (A) và thành phần chính PCA (B, C) sử dụng 1.936 SNPs trung tính	80
Hình 3.5. Cấu trúc quần thể cá ét mọi dựa trên phân tích nhóm Structure (A) và thành phần chính PCA (B, C) sử dụng 760 SNPs trung tính	81
Hình 3.6. Cấu trúc quần thể cá vô đém dựa trên phân tích nhóm Structure (A) và thành phần chính PCA (B, C) sử dụng 1.176 SNPs trung tính	82
Hình 3.7. Kết quả kiểm định Mantel về mối tương quan giữa khoảng cách địa lý và sự khác biệt di truyền của cá chạch lá tre (A), ét mọi (B) và vô đém (C)	82
Hình 3.8. Các trục MEM-1 giữa sai khác di truyền và khoảng cách địa lý từ phân tích dbMEM của cá chạch lá tre (A), cá ét mọi (B) và cá vô đém (C).	83
Hình 3.9. Mô hình di cư hiện tại của cá ét mọi (A) và cá vô đém (B) ở hạ lưu sông Mekong.....	88
Hình 3.10. Sơ đồ hình cây về lịch sử phân tách và pha trộn quần thể cá ét mọi (A) và vô đém (B) sử dụng Treemix.	90
Hình 3.11. Hệ gen ti thể của cá ét mọi <i>Labeo chrysophekadion</i>	102
Hình 3.12. Giao diện đồ họa thể hiện vị trí của vùng mã vạch RAD trên hệ gen ti thể sử dụng tính năng BLAST	104

DANH MỤC BẢNG

Bảng 1.1. So sánh ưu và nhược điểm của các loại chỉ thị phân tử phổ biến.....	23
Bảng 1.2. Ưu và nhược điểm của các phương pháp nghiên cứu di cư của cá	27
Bảng 2.1. Thông tin vị trí địa lý, vùng sinh thái và số lượng cá thể của các loài cá thu thập tại các địa điểm nghiên cứu ở lưu vực sông Mekong	46
Bảng 2.2. Các thông số được cài đặt để lắp ráp <i>de novo</i> hệ gen.....	53
Bảng 2.3. Các thông số và giá trị lọc được sử dụng trong nghiên cứu hiện tại	56
Bảng 3.1. Các thông số lắp ráp và sắp xếp hệ gen cá vồ đếm từ ba bộ công cụ.....	68
Bảng 3.2. Các thông số về hệ gen của cá chạch lá tre và cá ét mọi.....	70
Bảng 3.3. Số lượng cá thể và SNPs qua các bước lọc của ba loài cá	71
Bảng 3.4. Các thông số đa dạng di truyền ở các quần thể của ba loài cá điển hình ở lưu vực sông Mekong.....	77
Bảng 3.5. Số nhóm quần thể được sử dụng để phân tích phương sai phân tử.....	83
Bảng 3.6. Kết quả ước lượng mức độ thay đổi phân tử AMOVA của các quần thể cá chạch lá tre <i>Macrognathus siamensis</i>	84
Bảng 3.7. Kết quả ước lượng mức độ thay đổi phân tử AMOVA của các quần thể cá ét mọi <i>Labeo chrysophekadion</i>	85
Bảng 3.8. Kết quả ước lượng mức độ thay đổi phân tử AMOVA của các quần thể cá vồ đếm <i>Pangasius larnaudii</i>	86
Bảng 3.9. Hệ số cận huyết và kích thước quần thể hiệu quả của ba loài cá	87
Bảng 3.10. Sự tương đồng trình tự hệ gen ti thể của UB-MK503 với dữ liệu từ Ngân hàng Gen	101
Bảng 3.11. Đa dạng di truyền của các quần thể cá ét mọi ở hạ lưu sông Mekong dựa trên dữ liệu mã vạch RAD.....	105
Bảng 3.12. Sự khác biệt di truyền (F_{ST}) giữa các quần thể cá ét mọi ở hạ lưu sông Mekong.....	106

DANH MỤC CHỮ VIẾT TẮT

Từ viết tắt	Từ đầy đủ	Nghĩa tiếng Việt sử dụng trong luận án
2b-RAD	2-enzyme-based Restriction site Associated DNA sequencing	Kỹ thuật sử dụng enzyme cắt giới hạn thuộc nhóm IIB
AFLP	Amplified Fragment Length Polymorphism	Đa hình các đoạn cắt khuếch đại
AG		An Giang
AMOVA	Analysis of molecular variance	Phân tích phương sai phân tử
<i>cyt b</i>		Gen cytochrome b
DBG	de Bruijn graph	Đồ thị <i>de Bruijn</i>
ĐBSCL		Đồng bằng sông Cửu Long
ddRAD	double digest RAD	Kỹ thuật sử dụng 2 enzyme cắt giới hạn
DT		Đồng Tháp
EzRAD	enzyme-reduced representation sequencing	Kỹ thuật sử dụng một hay nhiều enzyme cắt giới hạn
GBS	genotype by sequencing	Xác định kiểu gen bằng giải trình tự
He	expected heterozygosity	Dị hợp tử mong đợi
Ho	observed heterozygosity	Dị hợp tử quan sát
HWE	Hardy-Weinberg Equilibrium	Cân bằng Hardy-Weinberg
ISSR	Inter-simple Sequence Repeat	Trình tự lặp lại đơn giản
KT		Kratié
LD	Linkage disequilibrium	Mất cân bằng liên kết
LMB	Lower Mekong River Basin	Hạ lưu sông Mekong
MAC	Minor Allele Count	Số lượng Alen thay thế
mAF	minor Allele Frequency	Tần số alen thay thế
minMeanDP	minimum Mean Depth	Độ sâu bao phủ
ND2		NADH dehydrogenase 2
Ne	effective population size	Kích thước quần thể hiệu quả
NGS	Next Generation Sequencing	Công nghệ giải trình tự thế hệ mới
OLC	Overlap-Layout-Consensus	Vùng lặp-Sắp xếp-Trình tự đồng thuận
PA		Paksan

PE		Pakse
Q	Mean quality score	Chất lượng trung bình
RAD-seq	Restriction site Associated DNA Sequencing	Kỹ thuật giải trình tự DNA dựa trên vị trí cắt giới hạn
RAPD	Random amplification of polymorphic DNA	Đa hình nhân bản DNA ngẫu nhiên
RFLP	Restriction fragment length polymorphism	Đa hình độ dài các đoạn cắt giới hạn
SBL	Sequencing By Ligation	Giải trình tự bằng phương pháp gắn
SBS	Sequencing by Synthesis	Giải trình tự bằng phương pháp tổng hợp
SNP	Single Nucleotide Polymorphism	Đa hình nucleotide đơn
ST		Stung Treng
UB		Ubon Ratchathani
UMB	Upper Mekong River Basin	Thượng lưu sông Mekong

GIẢI THÍCH CÁC THUẬT NGỮ

1. Đoạn đọc (*read*): đoạn trình tự DNA có kích thước ngắn (<300 bp) hoặc dài (vài kilobase) tùy theo chiến lược giải trình tự.
2. Contigs là các đoạn trình tự dài, liên tục, không có khoảng trống thu được từ việc lắp ráp các đoạn đọc chồng lấp (*overlap*) lên nhau.
3. Scaffolds là tập các contigs được sắp xếp theo trật tự và có thể có khoảng trống giữa chúng.
4. Độ sâu bao phủ (*Depth of Coverage*, thường gọi là "Depth") là số lần một nucleotide cụ thể trong trình tự DNA được đọc trong quá trình giải trình tự.
5. Độ sâu bao phủ trung bình tối thiểu (*Minimum Mean Depth*) là độ sâu bao phủ trung bình thấp nhất được yêu cầu hoặc quan sát trong một tập dữ liệu giải trình tự, và thường được sử dụng để đánh giá chất lượng dữ liệu và độ tin cậy của kết quả phân tích. Ví dụ độ sâu bao phủ $\geq 30x$, nghĩa là mỗi nucleotide nên được đọc ít nhất 30 lần để đảm bảo độ chính xác cao.
6. Alen thay thế (*Minor Allele Count*) là số lần xuất hiện của alen ít phổ biến hơn trong một tập hợp dữ liệu di truyền.
7. Tần số alen thay thế (*Minor Allele Frequency*) là tỉ lệ xuất hiện của alen ít phổ biến hơn (alen thay thế) trong một quần thể.
8. Loci trung tính (*neutral loci*) là các vị trí (loci) trong hệ gen không chịu tác động trực tiếp của chọn lọc tự nhiên.
9. Loci ngoại vi (*outlier loci*) là những vị trí trong hệ gen có sự khác biệt đáng kể về tần số alen hoặc mức độ đa dạng di truyền so với phần lớn các loci khác trong quần thể.
10. Loci đáp ứng (*adaptive loci*) là các vị trí trong hệ gen có liên quan đến sự thích nghi của sinh vật với môi trường.
11. Loci cân bằng (*balance loci*) là những vị trí trong hệ gen mà tại đó tính đa dạng di truyền được duy trì nhờ sự tác động của các cơ chế chọn lọc đặc biệt, chẳng hạn như chọn lọc cân bằng (*balancing selection*). Các loci này có sự tồn tại song song

của nhiều alen, vì các alen này không bị loại bỏ hoặc cố định một cách triệt để, mà thay vào đó chúng duy trì sự cân bằng trong quần thể qua nhiều thế hệ.

12. Phân tích phương sai phân tử (*Analysis of Molecular Variance* – AMOVA) là phương pháp thống kê được sử dụng để phân tích và đo lường sự phân bố của biến dị di truyền trong các nhóm quần thể sinh vật. Phương pháp này phân chia tổng biến dị di truyền thành các phần tử tương ứng với các mức phân lớp (giữa các quần thể, giữa các nhóm trong quần thể, và trong các cá thể của từng nhóm.) và xác định tỉ lệ biến dị giữa các nhóm và trong từng nhóm. AMOVA đặc biệt hữu ích trong việc nghiên cứu cấu trúc quần thể và mức độ phân hóa di truyền giữa các nhóm hoặc quần thể.

13. *Markov chain Monte Carlo (MCMC) sampling*: Một kỹ thuật thống kê để tích hợp một hàm bằng cách lấy các mẫu ngẫu nhiên (“Monte Carlo”) từ hàm, dựa trên mỗi mẫu của mẫu trước đó (“Markov chain”). Kỹ thuật ngẫu nhiên này rất hữu ích khi hàm không thể được tích hợp trực tiếp, nhưng có thể thất bại nếu mẫu được rút ra không đủ lớn hoặc không khám phá tất cả các vùng quan trọng của hàm.

14. Dóng hàng (*alignment*) là quá trình sắp xếp hai hoặc nhiều trình tự để xác định các vùng tương đồng giữa chúng.

15. Sắp xếp (*mapping*) là quá trình gán các đoạn trình tự ngắn lên một trình tự tham chiếu đã biết nhằm xác định vị trí chính xác của các đoạn đọc trên trình tự tham chiếu.

MỞ ĐẦU

1. Lý do chọn đề tài

Sông Mekong bắt nguồn từ cao nguyên Tây Tạng (Trung Quốc) với chiều dài 4.909 km chảy qua địa phận của 6 quốc gia châu Á trước khi đổ ra Biển Đông, gồm Trung Quốc và Myanmar thuộc phần thượng lưu (*Upper Mekong Basin* – UMB); Lào, Thái Lan, Campuchia và Việt Nam thuộc phần hạ lưu (*Lower Mekong Basin* – LMB). Lưu vực sông Mekong được biết đến là trung tâm đa dạng sinh học thủy sản toàn cầu, chỉ xếp sau Amazon và Congo, với hơn 1.300 loài cá [1–4]. Nguồn lợi thủy sản khổng lồ này đóng vai trò then chốt trong an ninh lương thực và sinh kế của hơn 70 triệu người dân và được dự báo sẽ tăng lên 100 triệu vào năm 2050 [5]. Đặc trưng của hệ sinh thái này là sự tồn tại của các nhóm cá với tập tính di cư phức tạp và đa dạng, từ cá đen (không di cư) đến cá trắng (di cư chặng dài), được điều phối chặt chẽ bởi các yếu tố thủy văn, đặc biệt là chu kỳ nhịp lũ hàng năm [6, 7].

Tuy nhiên, hệ sinh thái Mekong đang đối mặt với những thách thức nghiêm trọng do tác động tổng hợp của con người và biến đổi khí hậu. Việc khai thác quá mức, xây dựng đập thủy điện, phát triển nông nghiệp và đô thị hóa đã tạo ra những rào cản vật lý, làm thay đổi dòng chảy tự nhiên và chia cắt môi trường sống của cá, dẫn đến sự suy giảm nguồn lợi thủy sản và gia tăng nguy cơ tuyệt chủng cho nhiều loài [4, 8]. Trong bối cảnh đó, việc nghiên cứu sâu về đa dạng di truyền và cấu trúc quần thể của các loài cá ở lưu vực sông Mekong trở nên cấp thiết, nhằm cung cấp cơ sở khoa học cho công tác bảo tồn, phục hồi và quản lý bền vững nguồn lợi thủy sản.

Cho đến nay, một số nghiên cứu về đa dạng di truyền và cấu trúc quần thể cá sông Mekong đã được thực hiện, sử dụng các chỉ thị phân tử khác nhau như DNA ti thể [9–12], microsatellite [11, 13, 14] và đa hình nucleotide đơn (*Single Nucleotide Polymorphisms* – SNPs) [15–18]. Kết quả của các nghiên cứu này cho thấy mức độ phân hóa di truyền giữa các quần thể khác nhau dọc theo dòng chính và các dòng nhánh, phản ánh mối liên hệ giữa cấu trúc di truyền với tập tính di cư và lịch sử tiến hóa của loài. Các loài cá di cư chặng dài thường duy trì mức độ đa dạng di truyền cao

và có sự kết nối quần thể rộng hơn so với các loài cá không di cư [10, 11, 13, 14, 19, 20]. Gần đây, những nghiên cứu ứng dụng chỉ thị SNPs đã chỉ ra dấu hiệu suy giảm đa dạng di truyền, hệ số cận huyết cao và kích thước quần thể hiệu quả thấp ở một số loài cá sông Mekong, được cho là hệ quả của khai thác quá mức, phát triển thủy điện và biến đổi khí hậu [15, 17].

Mặc dù vậy, một khoảng trống kiến thức lớn vẫn tồn tại khi các nghiên cứu hiện tại chủ yếu tập trung vào từng loài đơn lẻ hoặc quy mô mẫu hạn chế, dẫn đến những dự báo về khả năng thích nghi của các nhóm cá trước biến động môi trường thiếu tính thống nhất. Sự thiếu hụt các phân tích so sánh đồng thời giữa các loài đại diện cho các chiến lược sống khác nhau trong cùng một bối cảnh thủy văn đã cản trở việc xây dựng các mô hình bảo tồn tổng thể cho toàn lưu vực.

Để khắc phục hạn chế về độ phân giải của các chỉ thị truyền thống, nghiên cứu này sử dụng chỉ thị SNPs với ưu điểm vượt trội về mật độ phân bố toàn hệ gen, cho phép phát hiện các biến dị di truyền ở mức độ nucleotide [21, 22]. Kỹ thuật EzRAD (*Enzyme Restriction-site Associated DNA*) được áp dụng như một chiến lược tối ưu để thu nhận bộ dữ liệu SNPs quy mô lớn cho các loài chưa có hệ gen tham chiếu, giúp đánh giá chi tiết và tin cậy hơn về mức độ đa dạng di truyền và cấu trúc quần thể trong bối cảnh môi trường bị phân mảnh tại lưu vực sông Mekong [23].

Trên cơ sở đó, nghiên cứu “**Phát hiện và nghiên cứu ứng dụng các chỉ thị đa hình nucleotide đơn (SNPs) bằng kỹ thuật EzRAD của ba loài cá điển hình ở lưu vực hạ lưu sông Mekong**” được thực hiện, tập trung vào 03 loài đại diện cho các tập tính di cư và vòng đời phát triển khác nhau, gồm cá chạch lá tre *Macrognathus siamensis* (cá đen, không di cư), cá ét mòi *Labeo chrysophekadion* (cá xám, di cư chặng ngắn và tùy nghi) và cá vồ đém *Pangasius larnaudii* (cá trắng, di cư chặng dài). Nghiên cứu hướng tới giải quyết các vấn đề then chốt thông qua hai câu hỏi: 1) Sự tương tác giữa tập tính di cư và vòng đời phát triển hình thành nên các biến thiên di truyền và khả năng duy trì dòng gen giữa các loài như thế nào?; 2) Liệu các bằng chứng về cấu trúc di truyền và sự suy giảm đa dạng hiện nay có phản ánh những tác

động tích lũy từ các rào cản cảnh quan và biến đổi điều kiện môi trường tại hạ lưu sông Mekong hay không? Những kết quả này sẽ góp phần làm sáng tỏ cơ chế duy trì và phân hóa di truyền của các loài cá sông Mekong, đồng thời cung cấp luận cứ khoa học quan trọng cho chiến lược bảo tồn và quản lý bền vững nguồn lợi thủy sản.

2. Mục tiêu nghiên cứu

2.1. Mục tiêu chung

Nghiên cứu này được thực hiện nhằm **đánh giá và so sánh** mức độ đa dạng di truyền và sự **kết nối quần thể** của 03 loài cá điển hình ở sông Mekong thông qua việc sử dụng kỹ thuật di truyền tiên tiến, từ đó xây dựng cơ sở dữ liệu di truyền phục vụ công tác quản lý, bảo tồn và khai thác bền vững nguồn lợi thủy sản trong khu vực.

2.2. Mục tiêu cụ thể

- **Lắp ráp *de novo* hệ gen** của ba loài cá điển hình (cá chạch lá tre *M. siamensis*, cá ét mòi *L. chrysophekadion* và cá vồ đém *P. larraudii*) ở lưu vực hạ lưu sông Mekong, đồng thời **phát hiện các SNPs** phục vụ phân tích di truyền quần thể.

- **Đánh giá và so sánh thông tin di truyền quần thể** của ba loài cá, gồm mức độ đa dạng di truyền trong và giữa quần thể, kích thước quần thể hiệu quả, mô hình di cư và mức độ kết nối di truyền giữa các quần thể.

- **Lắp ráp và chú giải hệ gen ti thể** của cá ét mòi, xác định mã vạch RAD và phân tích cấu trúc di truyền quần thể nhằm **bổ trợ và kiểm chứng** kết quả phân tích dựa trên dữ liệu SNPs thuộc hệ gen nhân.

3. Nội dung nghiên cứu

- 1) Lắp ráp *de novo* hệ gen của 03 loài cá điển hình ở lưu vực hạ lưu sông Mekong và xác định các chỉ thị phân tử SNPs.

- 2) Khảo sát và so sánh đa dạng di truyền, cấu trúc di truyền quần thể, ước lượng kích thước quần thể hiệu quả và dự đoán mô hình di cư của 03 loài cá nghiên cứu ở lưu vực hạ lưu sông Mekong.

- 3) Lắp ráp và chú giải hệ gen ti thể, xác định mã vạch RAD và khảo sát cấu trúc di truyền quần thể của cá ét mòi *Labeo chrysophekadion*.

4. Những đóng góp mới của luận án

❖ *Về nội dung nghiên cứu:*

Nghiên cứu ứng dụng kỹ thuật EzRAD để phát hiện và tuyển chọn các chỉ thị SNPs đặc trưng quần thể phân bố rộng trên toàn hệ gen, mở ra hướng tiếp cận mới trong nghiên cứu di truyền quần thể cho 03 loài cá điển hình ở sông Mekong có tập tính di cư và vòng đời phát triển khác nhau, gồm cá chạch lá tre (không di cư, đại diện cho quần thể cá địa phương), cá ét mọi (di cư chặng ngắn và tùy nghi, thể hiện sự liên kết giữa các vùng sinh thái và thích nghi linh hoạt với môi trường) và cá vô đém (di cư chặng dài, phản ánh sự kết nối giữa các vùng sinh thái và các tuyến đường di cư).

❖ *Về điểm mới và tính nổi bật:*

– *Phân tích đa dạng di truyền và cấu trúc quần thể toàn diện:* Sử dụng số lượng lớn chỉ thị SNPs được tuyển chọn từ kỹ thuật EzRAD, nghiên cứu cung cấp cái nhìn tổng quan về mức độ đa dạng di truyền và cấu trúc quần thể của cả 03 loài cá, vượt trội so với các nghiên cứu trước đây sử dụng các chỉ thị truyền thống, hoặc/và sử dụng chỉ thị SNPs trong nghiên cứu đơn loài.

– *Dự đoán mô hình di cư phù hợp với lịch sử phát triển:* Lần đầu tiên dữ liệu SNPs được ứng dụng và sử dụng các thuật toán đa dạng để dự đoán mô hình di cư của cá ét mọi và cá vô đém ở LMB, gợi mở thông tin về các tuyến đường di cư và sự kết nối giữa các quần thể cá.

– *Ứng dụng dữ liệu EzRAD trong phân tích hệ gen ti thể và cấu trúc quần thể:* Nghiên cứu khai thác dữ liệu EzRAD từ nhiều cá thể để vừa tái cấu trúc hệ gen ti thể, vừa sàng lọc đồng thời các vị trí đa hình mật độ cao ở cá ét mọi. Nguồn dữ liệu này được ứng dụng hiệu quả vào việc đánh giá đa dạng di truyền và cấu trúc quần thể của loài nghiên cứu, mở ra hướng tiếp cận mới trong nghiên cứu hệ gen ti thể của các loài cá.

5. Ý nghĩa khoa học và thực tiễn của luận án

Ý nghĩa khoa học: Nghiên cứu đã làm sáng tỏ cấu trúc di truyền và mô hình di cư của ba loài cá điển hình tại sông Mekong thông qua kỹ thuật EzRAD. Kết quả không chỉ khắc phục hạn chế của các phương pháp truyền thống về độ phân giải dữ liệu mà còn cung cấp khung tham chiếu giá trị cho công tác bảo tồn đa dạng sinh học tại các hệ thống sông lớn. Đặc biệt, việc kết hợp phân tích hệ gen ti thể và mã vạch RAD đã thiết lập hướng tiếp cận mới trong nghiên cứu di truyền dòng mẹ và khả năng thích nghi của các loài thủy sinh.

Ý nghĩa thực tiễn: Kết quả nghiên cứu này có ý nghĩa thực tiễn quan trọng trong việc bảo tồn và quản lý nguồn lợi thủy sản ở sông Mekong. Việc xác định cấu trúc quần thể và mô hình di cư của các loài cá giúp các nhà quản lý đưa ra các biện pháp bảo tồn phù hợp, đảm bảo sự bền vững của các quần thể cá. Dữ liệu SNPs và hệ gen ti thể được tạo ra từ nghiên cứu này là nguồn tài nguyên di truyền quý giá, có thể được sử dụng để theo dõi sự thay đổi di truyền của các quần thể cá theo thời gian, đánh giá tác động của các yếu tố môi trường và hỗ trợ các chương trình phục hồi quần thể. Ngoài ra, việc giải mã hệ gen ti thể của cá ét mọi được ứng dụng trong việc phát triển các phương pháp xác định loài và theo dõi nguồn gốc sản phẩm thủy sản.

6. Cấu trúc của luận án

Luận án gồm 130 trang (không kể bìa, mục lục, danh mục các hình, bảng biểu và phần phụ lục), bao gồm:

Mở đầu có 5 trang trình bày tính cấp thiết, mục tiêu, nội dung, tính mới, ý nghĩa khoa học và thực tiễn của luận án.

Nội dung chính gồm 3 chương:

Chương 1: Tổng quan tài liệu gồm 36 trang;

Chương 2: Đối tượng, phạm vi và phương pháp nghiên cứu, gồm có 24 trang;

Chương 3: Kết quả nghiên cứu và thảo luận gồm 42 trang;

Phân kết luận và kiến nghị gồm 2 trang;

Phần các công trình công bố và tài liệu tham khảo gồm 22 trang.

Trong luận án, tổng cộng có 17 bảng, 28 hình ảnh, 265 tài liệu tham khảo tiếng Việt và tiếng Anh.

CHƯƠNG 1:

TỔNG QUAN TÀI LIỆU

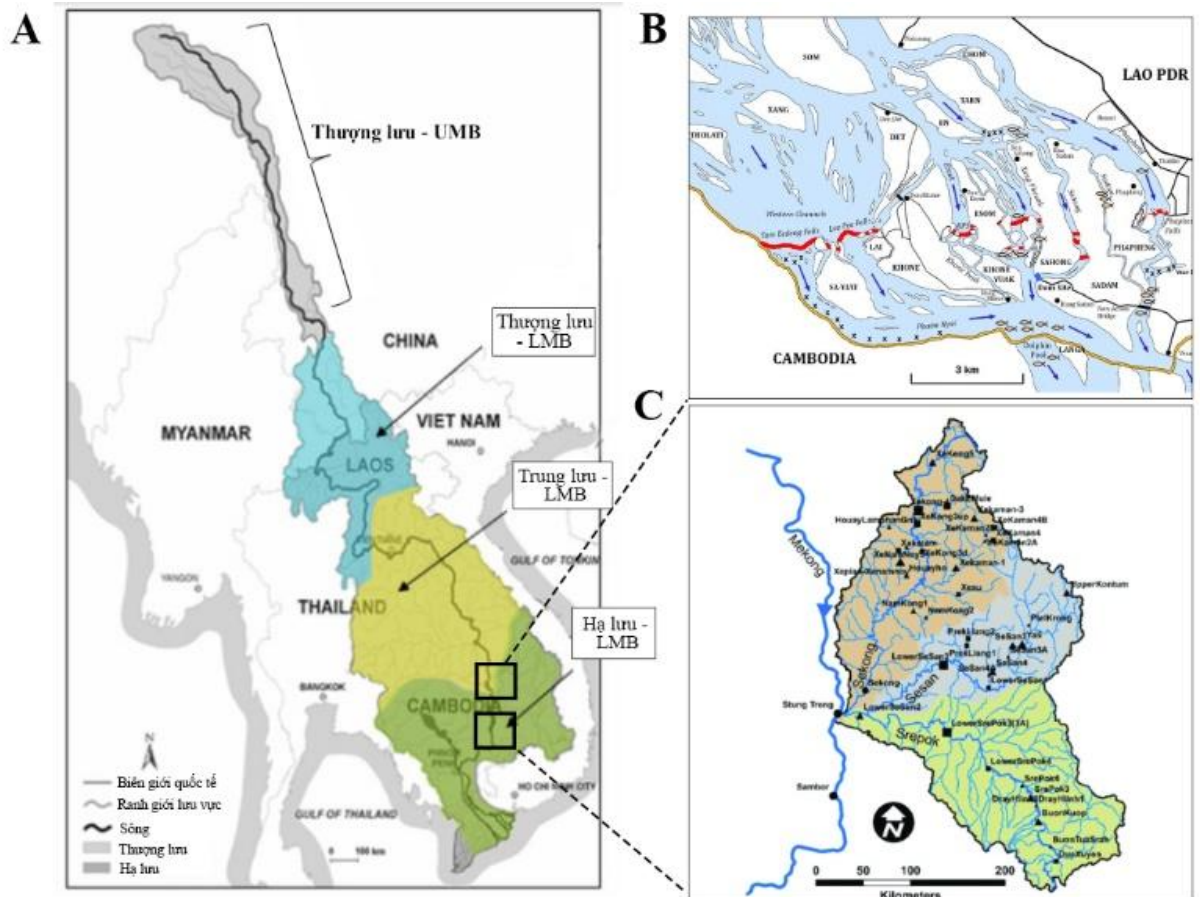
1.1. Tổng quan về lưu vực hạ lưu sông Mekong

1.1.1. Đặc điểm địa lý, khí hậu và thủy văn của lưu vực hạ lưu sông Mekong

Sông Mekong là một trong những hệ thống sông lớn nhất thế giới, bắt nguồn từ cao nguyên Tây Tạng (Trung Quốc) và chảy qua 6 quốc gia Châu Á, đạt tổng chiều dài 4.909 km [24], diện tích lưu vực 795.000 km², với lưu lượng trung bình hàng năm 14.500 m³/s [25]. Về mặt địa lý, lưu vực phân chia thành Thượng lưu (*Upper Mekong River Basin* – UMB) chảy qua Trung Quốc, Myanmar (chiếm 24% diện tích lưu vực) và Hạ lưu (*Lower Mekong River Basin* – LMB) nằm trên địa phận Lào, Thái Lan, Campuchia và Việt Nam (chiếm 76%). Lưu vực LMB lại chia tiếp thành thượng lưu-LMB, trung lưu-LMB và hạ lưu-LMB (**Hình 1.1A**), với ranh giới tự nhiên giữa trung lưu-LMB và hạ lưu-LMB là thác Khôn - quần thể thác ghềnh dài hơn 10 km, vốn là rào cản tự nhiên quan trọng đối với sự di cư của loài cá [26] (**Hình 1.1B**). Xuôi dòng, sông hợp lưu với hệ thống 3S (Sesan, Srepok, Sekong) tại Campuchia (**Hình 1.1C**), sau đó kết nối với Biển Hồ qua sông Tonle Sap. Khi vào Việt Nam, dòng chính chia thành sông Tiền và sông Hậu, tạo nên Đồng bằng sông Cửu Long (ĐBSCL). Xét về khu hệ sinh thái, ngoại trừ vùng Thượng Lan Thương thuộc Trung Quốc, bốn vùng sinh thái còn lại (Hạ Lan Thương, cao nguyên Khorat, Kratié-Stung Treng và châu thổ Mekong) đều thuộc lưu vực LMB [27].

Thủy văn của lưu vực sông Mekong chịu sự chi phối mạnh mẽ của khí hậu gió mùa, tạo nên hai mùa rõ rệt: mùa mưa (từ giữa tháng 6 đến đầu tháng 11) và mùa khô (từ tháng 12 đến tháng 5), cùng hai giai đoạn chuyển tiếp. Sự thay đổi này dẫn đến biến động lưu lượng dòng chảy đáng kể. Mùa mưa cung cấp khoảng 75% tổng lượng nước hàng năm, thường gây ra lũ lụt trên diện rộng ở hạ lưu và đóng vai trò thiết yếu cho chu trình tưới tiêu, sản xuất và đánh bắt cá [28]. Lưu lượng nước đạt đỉnh điểm (khoảng 40.000 m³/s) vào tháng 9 – 10 [29]. Ngược lại, mùa khô là giai đoạn lưu lượng nước thấp nhất (khoảng 2.000 m³/s vào đầu tháng 4) [29], với khoảng 35%

tổng lưu lượng nước đến từ thượng nguồn Trung Quốc [30]. Do đó, cơ sở hạ tầng cấp nước ở thượng lưu đóng vai trò quan trọng trong việc điều tiết thủy văn, đặc biệt trong giai đoạn khô hạn. Lưu lượng nước bắt đầu tăng mạnh vào cuối tháng 5, báo hiệu sự khởi đầu của chu kỳ gió mùa mới.



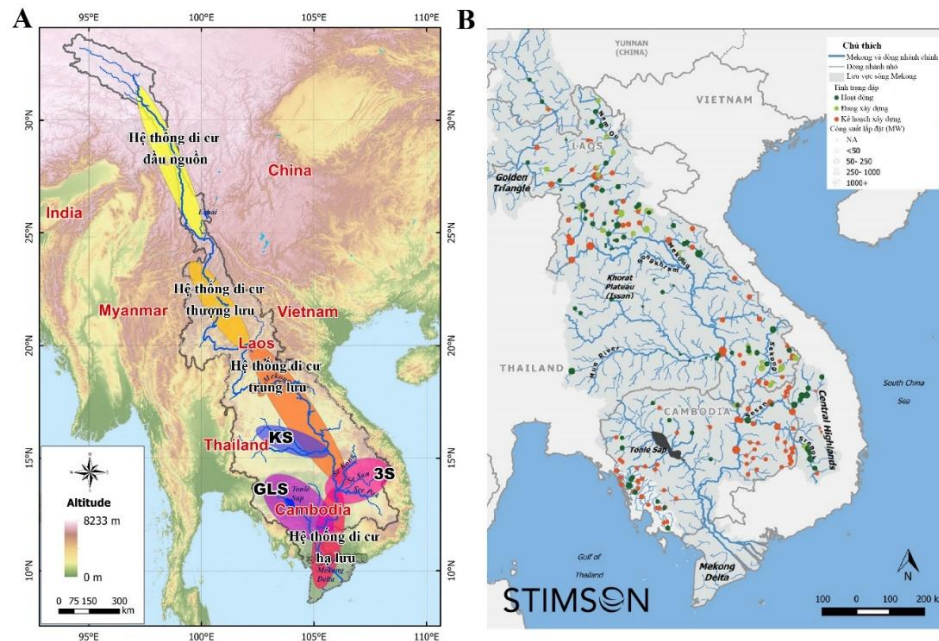
Hình 1.1. Hệ thống lưu vực sông Mekong

A. Lưu vực sông Mekong phân chia theo địa lý (màu xám thuộc khu vực thượng lưu-UMB, xanh dương là thượng lưu-LMB, vàng là trung lưu-LMB và xanh lá là hạ lưu-LMB) [31]; B. Khu vực thác Khôn (chữ in nghiêng là các kênh chính, chữ in hoa là các đảo, màu đỏ là ghềnh thác và mũi tên màu xanh là hướng dòng chảy) [32], C. Khu vực Stung Treng và các đập thủy điện đã (▲) và sẽ (■) xây dựng [33]

1.1.2. Các hệ thống di cư ở hạ lưu sông Mekong

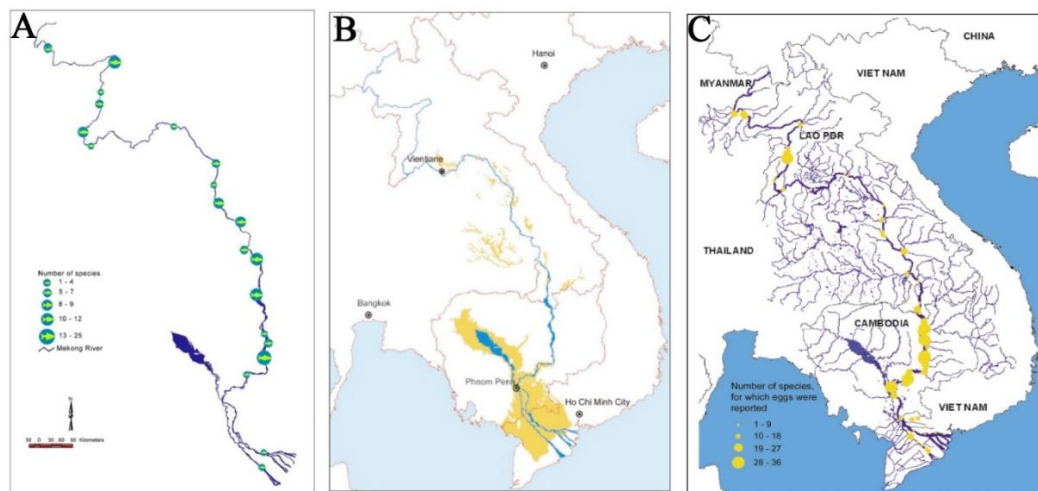
Ở hạ lưu sông Mekong, ba hệ thống di cư đường dài được xác định [7, 34] (**Hình 1.2A**). (1) *Hệ thống di cư thượng lưu-LMB*: trải dài từ điểm hợp lưu giữa dòng nhánh Xiaohai (Trung Quốc) và dòng chính đến Luang Prang (Lào). Đoạn sông đặc trưng bởi việc thiếu các nhánh sông lớn và ít vùng ngập. Tại hệ thống này, các loài di cư từ nơi trú ẩn mùa khô ở dòng chính, ngược dòng lên khu vực sinh sản (như Sông Buyuan) khi mùa mưa bắt đầu. (2) *Hệ thống di cư trung lưu-LMB*: từ Luang Prang xuống đến thác Khôn, đặc trưng với địa hình gồ ghề, thung lũng hẹp và nhiều sông nhánh lớn. Ở hệ thống này, cá sinh sản ở các nhánh sông có dòng chảy mạnh, sau đó di cư về dòng chính để trưởng thành, trú ẩn trong các vực sâu vào mùa khô. (3) *Hệ thống di cư hạ lưu-LMB*: Từ Thác Khôn (Campuchia) xuống đến Đồng bằng sông Cửu Long (ĐBSCL) của Việt Nam. Vào đầu mùa mưa, cá di cư từ nơi trú ẩn mùa khô (vực sâu quanh Kratié – Stung Treng **Hình 1.3A**) đến các vùng ngập phía Nam Campuchia và ĐBSCL để kiếm ăn (**Hình 1.3B**).

Bên cạnh đó, Kang & Huang (2021) [34] mô tả thêm ba hệ thống di cư ngang (**Hình 1.2A**), gồm (1) *Hệ thống Biển Hồ (GLS)*: vùng ngập lũ lớn nhất ở Campuchia, kết nối với Sông Mekong qua Sông Tonle Sap, nơi dòng chảy đảo chiều hai lần mỗi năm. Vùng này cung cấp môi trường sống và sinh sản quan trọng, cho phép các loài (như cá linh *Henicorhynchus* spp.) di cư ngược dòng chính để sinh sản, sau đó ấu trùng trôi ngược về vùng hạ lưu để phát triển [7]. (2) *Hệ thống sông Sekong-Sesan-Srepok (3S)*: Hợp lưu của ba nhánh sông lớn ở Stung Treng (Campuchia). Ở hệ thống này, nhiều loài cá như cá linh *Henicorhynchus siamensis*, các loài cá chép kích thước lớn (*Catlocarpio siamensis*, *Cirrhinus microlepis*, *Cyclocheilichthys enoplos*, *Probarbus jullieni*) và một số loài cá da trơn thực hiện di cư từ dòng chính sông Mekong vào hệ thống 3S trong mùa khô để tìm nơi trú ẩn hoặc sinh sản. (3) *Hệ thống Korat (KS)* nằm ở phía đông bắc của Thái Lan, gồm 2 nhánh sông lớn là sông Chi và sông Mun, đóng vai trò là một trong những bãi đẻ quan trọng.



Hình 1.2. Các hệ thống di cư (A) [34] và các đập thủy điện trên dòng chính và dòng nhánh (B) [35] ở sông Mekong. (KS: hệ thống Korat, GLS: Hệ thống Biển Hồ, 3S: Hệ thống sông Sekong-Sesan-Srepok)

Các hệ thống di cư trung lưu và hạ lưu-LMB, bao gồm cả hệ thống ngang (GLS, 3S, KS), có sự liên kết mật thiết, trong khi hệ thống thượng lưu-LMB có xu hướng tách biệt hơn. Điều này cũng được thể hiện qua số lượng vực sâu ít hơn (Hình 1.3A) và ít bắt gặp cá ở giai đoạn mang trứng (Hình 1.3C) [7].



Hình 1.3. Sơ đồ phân bố vực sâu và số lượng loài (A), vùng ngập (màu vàng, B) và số lượng loài cá mang trứng (màu vàng, C) ở dòng chính sông Mekong [7]

1.1.3. Nguồn lợi thủy sản và các yếu tố ảnh hưởng đến đa dạng sinh học ở lưu vực hạ lưu sông Mekong

1.1.3.1. Nguồn lợi thủy sản ở lưu vực hạ lưu sông Mekong

Lưu vực sông Mekong, được đặc trưng bởi đa dạng các vùng sinh thái, các môi trường sống khác nhau (dòng chính, dòng nhánh, hồ, hồ chứa, vực sâu, vùng ngập, hang động và cửa sông) và chu kỳ lũ hàng năm, là một trong ba lưu vực giàu nguồn lợi cá nước ngọt nhất thế giới, chỉ đứng sau Amazon và Congo [4]. Hệ thống sông hỗ trợ một ngành thủy sản nội địa khổng lồ, cung cấp nguồn protein chính cho hơn 70 triệu người [36, 37].

Nguồn tài nguyên cá của Mekong vô cùng đa dạng với 1.393 loài cá được ghi nhận, trong đó có ít nhất 293 loài (21%) là loài di cư [38]. Sản lượng thu hoạch thủy sản (bao gồm cá và các loài khác) đạt đỉnh khoảng 3 triệu tấn (năm 2010), với tổng giá trị kinh tế ước tính lên đến 11 tỷ USD; tuy nhiên, sản lượng này đã suy giảm đáng kể vào năm 2020. Sản lượng thủy sản hàng năm được khai thác ở LMB ước tính chiếm khoảng 19,3% tổng sản lượng thủy sản nước ngọt toàn cầu [39].

Dựa trên tập tính sinh thái và sử dụng môi trường sống [31, 40], các loài cá Mekong được phân thành ba nhóm chính:

(1) Nhóm cá đen: Các loài sống chủ yếu ở vùng nước tĩnh (hồ, đầm lầy) và di chuyển ngắn theo chu kỳ giữa vùng nước và vùng ngập lũ (ví dụ: như cá rô *Anabas testudineus*, cá trê *Clarias batrachus*, cá lóc đen *Channa striata*, cá chạch lá tre *Macrognathus siamensis*). Nhóm này chiếm khoảng 50% tổng sản lượng cá LMB [41];

(2) Nhóm cá trắng: Các loài di cư đường dài, thường di chuyển lên thượng nguồn các nhánh sông vào mùa mưa để sinh sản và trở về nơi sinh sống vào cuối mùa. Ví dụ như các loài thuộc họ cá chép Cyprinidae như cá cóc (*Cyclocheilichthys enoplos*) và cá trôi vảy nhỏ (*Cirrhinus microlepis*), cũng như các loài cá da trơn thuộc họ Pangasiidae như cá tra *Pn. hypophthalmus*, cá ba sa *P. bocourti*, cá bông lau *P. krempfi*, cá hú *P. conchophilus* và cá vô đém *P. larnaudii*, với sản lượng khai thác của nhóm này chiếm 36% [41];

(3) Nhóm cá xám: Nhóm trung gian, thực hiện các cuộc di cư ngắn hơn giữa vùng ngập và sông lân cận (ví dụ cá trên bầu *Ompok bimaculatus*, cá thác lác *Notopterus notopterus*, cá ét mọi *Labeo chrysophekadion*). Tỷ lệ sản lượng của nhóm này là 16% [41].

1.1.3.2. Các yếu tố ảnh hưởng đến đa dạng sinh học ở hạ lưu sông Mekong

Đa dạng sinh học ở LMB đang đối mặt với những thách thức đáng kể từ tác động của con người và biến đổi khí hậu (BĐKH). Các tác động của con người được phân loại thành bốn nhóm chính: (1) chuyển nước giữa các lưu vực; (2) thay đổi sử dụng nước (phát triển nông nghiệp, đô thị hóa, phá rừng); (3) xây dựng công trình (đập thủy điện, đập dâng, đào kênh, nạo vét); và (4) tác động trực tiếp lên dòng chảy (ô nhiễm, khai thác loài bản địa, du nhập loài ngoại lai) [42]. Trong số này, việc phân mảnh môi trường sống là mối đe dọa sinh thái lớn nhất [43]. Trên toàn hệ thống, hơn 400 đập đã hoàn thành hoặc dự kiến xây dựng; riêng 136 đập đã được xác định gây tác động tiêu cực đến tính toàn vẹn sinh thái (**Hình 1.2B**) [44]. Các đập này tạo ra rào cản vật lý ngăn chặn sự di cư của cá, làm thay đổi chế độ dòng chảy và môi trường sống, gây ảnh hưởng tiêu cực đến sinh sản và dẫn đến suy giảm đáng kể số lượng loài cá (ước tính từ 26% đến 42%) và hơn 100 loài có thể biến mất do những thay đổi trong hệ sinh thái và chế độ dòng chảy [36]. Ví dụ, việc hoàn thành đập Pak Mun (Thái Lan) đã làm suy giảm sản lượng cá lên tới 60–80% ở thượng nguồn [45]. Ngoài ra, việc giữ lại trầm tích tại đập và khai thác cát quá mức làm thay đổi hình thái lòng sông và gây xói lở. Những yếu tố này, kết hợp với các phương pháp đánh bắt hủy diệt, đã thay đổi cấu trúc sản lượng đánh bắt, chuyển từ cá lớn, có giá trị sang cá nhỏ hơn và ít giá trị kinh tế hơn [34].

BĐKH đang làm gia tăng áp lực thông qua thay đổi chế độ thủy văn và môi trường sống. Các nghiên cứu dự báo nhiệt độ trung bình toàn lưu vực sẽ tăng khoảng 0,79°C và lượng mưa hàng năm tăng 200 mm vào giai đoạn 2026–2041, dẫn đến lũ lụt trên diện rộng và hạn hán kéo dài [36]. Sự gia tăng nhiệt độ nước ảnh hưởng trực tiếp đến sinh lý và hành vi của các loài biến nhiệt như cá, làm tăng tỷ lệ trao đổi chất,

giảm kích thước tối đa, và tạo điều kiện cho các tác nhân gây bệnh phát triển [46]. Đồng thời, BĐKH làm gia tăng tần suất và cường độ của các hiện tượng thời tiết cực đoan; mùa khô kéo dài gây căng thẳng cho loài thủy sản do thiếu thức ăn, trong khi lũ lụt gia tăng cản trở quá trình di cư sinh sản [29]. Thêm vào đó, nước biển dâng và xâm nhập mặn ở ĐBSCL đang làm thay đổi môi trường sống ven biển, đe dọa các loài sinh vật đặc hữu, làm thay đổi phạm vi môi trường sống của các loài và thúc đẩy sự xuất hiện của các loài ngoại lai [47–49].

1.2. Đặc điểm sinh học, sinh sản và di cư của các đối tượng nghiên cứu

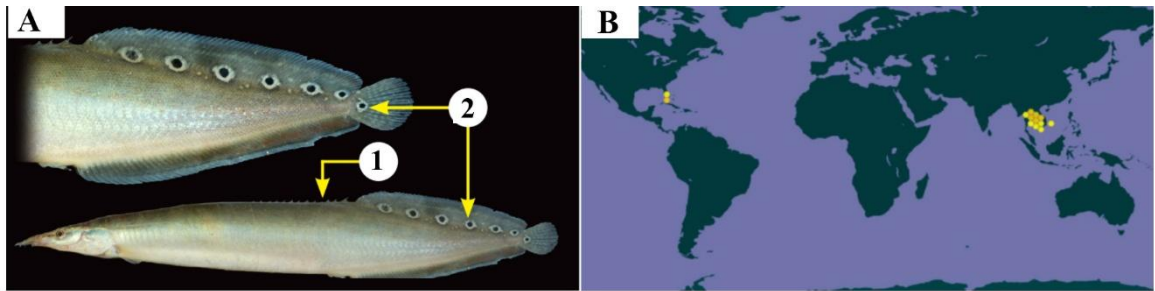
1.2.1. Cá chạch lá tre *Macroglyphus siamensis* Günther, 1861

1.2.1.1. Đặc điểm hình thái

Cá chạch lá tre *M. siamensis* (Synbranchiformes: Mastacembelidae) có cơ thể dài và kích thước tối đa khoảng 30 cm. Toàn bộ cơ thể cá được bao phủ bởi lớp vảy nhỏ. Đặc điểm đặc trưng của chúng là có miệng nhỏ và kéo dài thành vòi. Vây lưng kéo dài dọc theo thân và chia thành 2 phần: Phần trước gồm các gai cứng nhọn (13 – 19 gai), trong đó gai cuối cùng có kích thước to và dài nhất (**Hình 1.4A**); và phần sau là các tia vây mềm. Trên phần tia vây mềm của vây lưng, cá có 3 – 6 đốm tròn lớn, màu đen và viền trắng. Vây đuôi cũng có một đốm tương tự, nhưng kích thước nhỏ hơn [50–52] (**Hình 1.4A**).

1.2.1.2. Đặc điểm phân bố, dinh dưỡng và sinh sản

Cá chạch lá tre là loài bản địa phân bố ở lưu vực sông Mekong, sông Mae Klong, Chao Phraya (Thái Lan) và bán đảo Mã Lai (Malaysia) [53]. Ở lưu vực sông Mekong, cá chạch được ghi nhận ở Myanmar, Lào, Thái Lan, Campuchia, và Việt Nam [54]. Ngoài ra, cá chạch lá tre cũng được ghi nhận là loài ngoại lai ở Florida, Hoa Kỳ [55, 56], Singapore [57] (**Hình 1.4B**).



Hình 1.4. Hình thái ngoài (A) [66] và bản đồ phân bố (chấm màu vàng, B) [58] của cá chạch lá tre *Macrognathus siamensis*

(1) gai cứng nhọn phía trước vây lưng, (2) các đốm tròn lớn, màu đen, viền trắng

Cá chạch lá tre thuộc nhóm cá đen [59–61], thường sống chủ yếu ở các môi trường nước tĩnh như ao, hồ, đầm, ruộng lúa và các vùng ngập nước ven sông. Chúng di chuyển đến các vùng ngập theo mùa, nơi có nguồn dinh dưỡng và thức ăn phong phú, đồng thời cung cấp nơi trú ẩn an toàn trong mùa mưa như khu vực phía nam của Campuchia (hệ thống sông Tonle Sáp và Biển Hồ), Việt Nam (ĐBSCL) và những vùng ngập nhỏ kết nối với các dòng nhánh ở phía trung và thượng lưu thuộc Thái Lan và Lào [7]. Cá chạch lá tre là loài sống đáy, ưa thích các khu vực có dòng chảy yếu hoặc nước tĩnh. Chúng thường vùi mình trong các lớp bùn, cát, hoặc sỏi nhỏ, cũng như ở những vùng nước sâu và chỉ để lộ phần đầu. Đây là loài ăn tạp, thường kiếm ăn vào ban đêm và sáng sớm. Chúng ăn nhiều loại thức ăn khác nhau, bao gồm ấu trùng, côn trùng, giáp xác, cá nhỏ, giun, trứng và artemia [68].

Cá chạch lá tre có vòng đời ngắn và khả năng sinh sản sớm. Chúng bắt đầu sinh sản từ năm thứ hai và mùa sinh sản diễn ra vào đầu và giữa mùa mưa (từ tháng 5 đến 10 hàng năm). Số lượng trứng trung bình là 3.339 trứng/ cá thể mẹ (sức sinh sản tuyệt đối) hoặc 136 trứng trên mỗi gram trọng lượng cơ thể (sức sinh sản tương đối) [62]. Thời gian phát tán ấu trùng của cá chạch lá tre là khoảng 22 ngày [63].

1.2.2. Cá ét mọi *Labeo chrysophekadion* Bleeker, 1850

1.2.2.1. Đặc điểm hình thái

Cá ét mọi *L. chrysophekadion* (Cypriniformes: Cyprinidae) có màu xám đen, với phần lưng sẫm màu hơn so với bụng và hai bên thân. Toàn thân cá được bao phủ

bởi vây tròn, ngoại trừ phần đầu. Thân cá cao và có nhiều chấm đen và đỏ. Đầu của chúng nhỏ và hơi dẹp. Môi cá rất dày và có nhiều gai thịt (**Hình 1.5A**). Cá ét mọi là một trong những loài có kích thước lớn trong họ cá chép Cyprinidae, với kích thước tối đa khác nhau tùy theo khu vực: 40 cm ở Việt Nam, 60 cm ở Campuchia và Lào, và đạt 90 cm ở Thái Lan [6].

1.2.2.2. Đặc điểm phân bố, dinh dưỡng và sinh sản

Cá ét mọi là loài có tập tính sống theo đàn, thường tập trung ở tầng giữa và đáy của các thủy vực có dòng chảy chậm và nhiều giá thể [64]. Chúng phân bố rộng khắp khu vực Đông Nam Á, gồm sông Mekong, sông Chao Phraya (Thái Lan), bán đảo Mã Lai, cũng như các đảo Sumatra, Java và Borneo (Indonesia) (**Hình 1.5B**). Chúng thường được bắt gặp ở các sông suối, kênh rạch và vùng ngập lũ, đôi khi cũng được tìm thấy trong các ao nhưng với số lượng không nhiều. Cá ét mọi là loài có giá trị kinh tế quan trọng và được khai thác với số lượng lớn ở sông Mekong, với tỉ lệ khai thác cao ghi nhận ở sông Mun, Thái Lan [65]. Tuy nhiên, nguồn lợi cá ét mọi đang suy giảm và có nguy cơ tuyệt chủng, chủ yếu do khai thác quá mức [65], suy thoái môi trường sống và các tác động tiêu cực khác từ hoạt động của con người [3, 66].

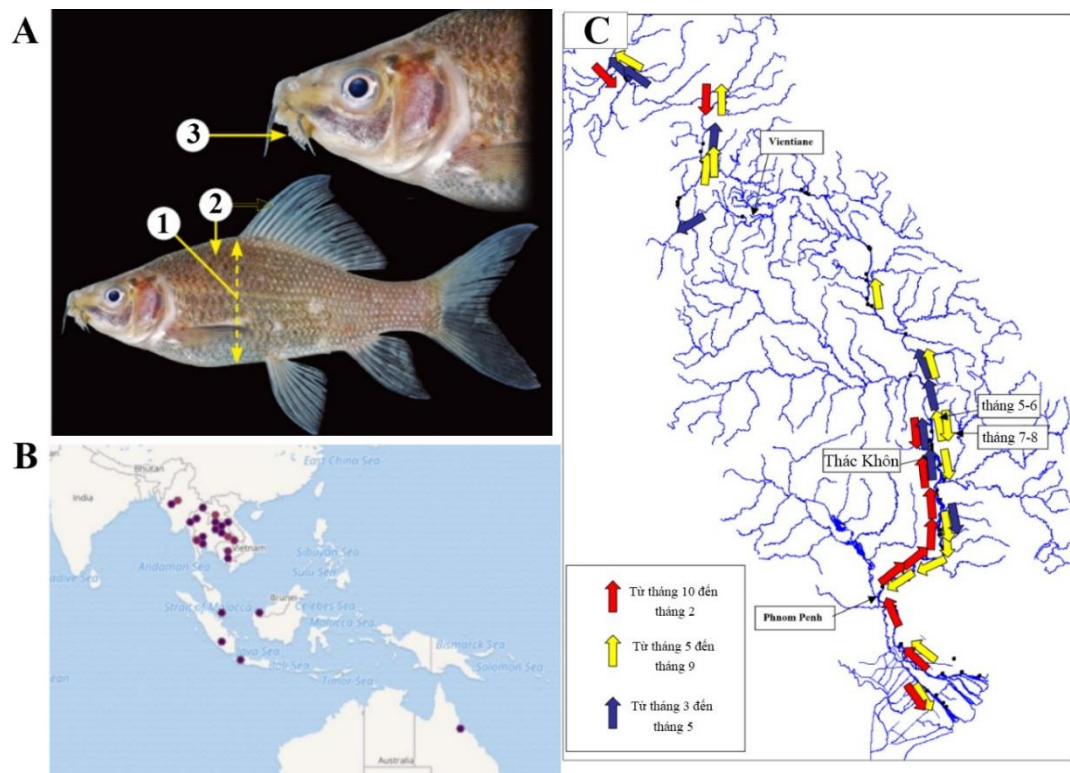
Cá ét mọi có tuổi thọ trung bình khoảng 28,7 năm [67]. Loài cá này có đặc điểm thành thực sinh dục muộn, thường sau hơn 2 năm tuổi và những cá thể đạt trọng lượng trên 2kg và kích thước lớn hơn 50cm mới bắt đầu quá trình sinh sản [64]. Cá ét mọi có khả năng sinh sản linh hoạt ở đa dạng các môi trường sống, như các khu vực ngập nước, đầm lầy và cả các hồ chứa nhân tạo [31]. Trứng và ấu trùng của cá ét mọi có thời gian trôi nổi trong nước kéo dài từ 30 đến 50 ngày [68].

1.2.2.3. Đặc điểm di cư

Cá ét mọi có tập tính di cư theo mùa. Cá trưởng thành thường di chuyển ngược dòng từ các vực sâu ở dòng chính vào các vùng ngập nước để sinh sản vào giai đoạn đầu mùa mưa (tháng 5 – 6). Cá bột và cá con thường ở lại các vùng ngập trong suốt mùa mưa. Khi nước bắt đầu rút, cả cá con và cá trưởng thành đều quay trở lại dòng chính để trú ẩn trong suốt mùa khô (tháng 10 – 12) [31]. Ở Thái Lan và Lào, quá trình

di cư ngược dòng của cá diễn ra từ tháng 3 – 8, và di cư xuôi dòng vào tháng 10 (chỉ được ghi nhận ở Thái Lan). Tại Campuchia, cá ét mọi di cư ngược dòng (từ Biển Hồ đến Stung Treng) được quan sát thấy từ tháng 10 – 3, và di cư xuôi dòng (từ Stung Treng đến ĐBSCL) từ tháng 3 – 8 (**Hình 1.5C**). Ở Việt Nam, thông tin về con đường di cư của cá ét mọi vẫn còn hạn chế [6].

Cho đến nay, chưa có thông tin về khoảng cách di cư của cá ét mọi, do đó, chúng có thể thuộc nhóm cá xám – di cư chặng ngắn [69] hoặc cá trắng – di cư chặng dài [70, 71]. Cá ét mọi cũng được biết đến là loài di cư tùy nghi (*facultative migration*), có thể thực hiện di cư hoặc không tùy thuộc vào điều kiện môi trường và nguồn thức ăn [31]. Theo Sokheng *et al.* (1999) và Poulsen *et al.* (2004), có nhiều quần đàn cá ét mọi phân bố ở các dòng nhánh (như sông Mun) và một số hồ chứa ở Thái Lan (như Bung Borape, Ubon Ratana) [72].



Hình 1.5. Hình thái ngoài (A), bản đồ phân bố (chấm màu tím, B) [72] và hướng di cư theo mùa (C) [6] của cá ét mọi *Labeo chrysophekadion* (1) Thân cao, (2) chấm đen và đỏ trên thân, (3) mô dày và có nhiều gai thịt

1.2.3. Cá vồ đém *Pangasius larnaudii* Bocourt, 1866

1.2.3.1. Đặc điểm hình thái

Cá vồ đém *P. larnaudii* (Siluriformes: Pangasiidae) có thân dài. Phần thân phía trước có dạng tròn, trong khi phần thân phía sau dẹp dần sang hai bên. Chúng có màu xám đen ánh xanh lục ở lưng và đầu, chuyển dần sang màu trắng ở bụng. Chúng có 2 cặp râu, trong đó râu hàm trên kéo dài đến gốc vây ngực và râu cằm ngắn hơn, không vượt quá màng mang. Một đốm đen lớn có thể thấy ở phía trên gốc vây ngực (**Hình 1.6A**). Vây lưng và vây ngực có gai cứng, với răng cưa hướng về gốc ở phía sau. Tia vây đơn của vây bụng dài hơn cả tia vây hậu môn. Ngoài ra, cá vồ đém còn có vây mỡ nhỏ, vây hậu môn khá dài và vây đuôi phân thùy sâu (**Hình 1.6A**).

1.2.3.2. Đặc điểm phân bố, dinh dưỡng và sinh sản

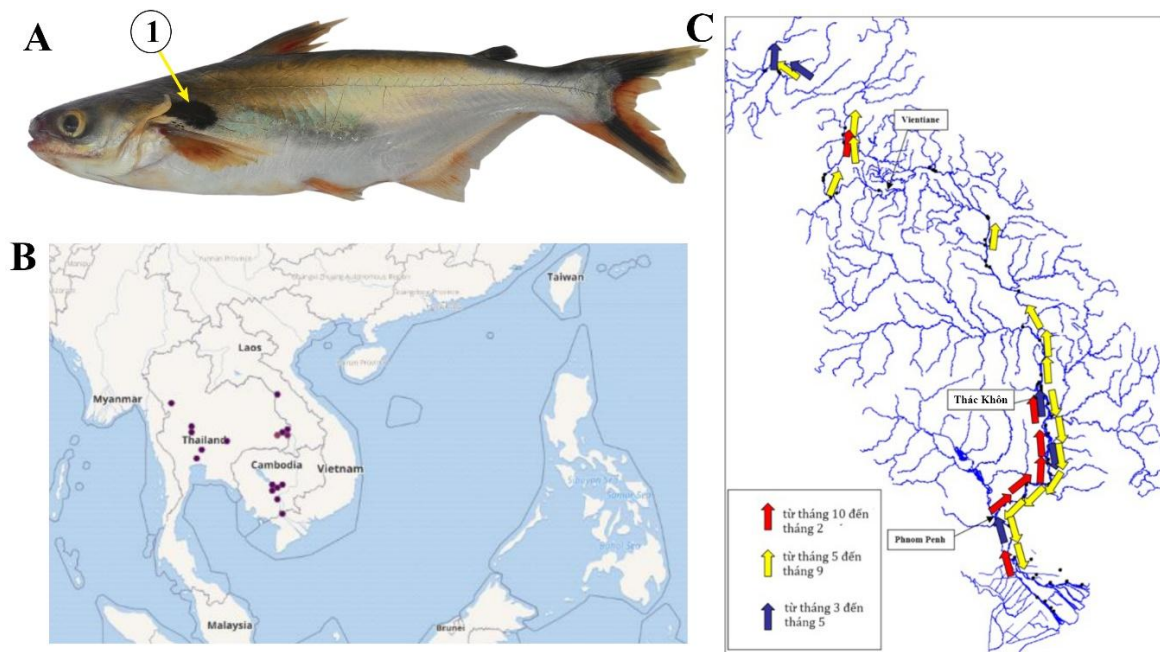
Cá vồ đém thường được tìm thấy ở các sông vừa và lớn, các vùng ngập ở sông Mekong và sông Chao Phraya (Thái Lan) (**Hình 1.6B**). Loài cá này có kích thước lớn, với chiều dài tối đa đạt tới 150 cm và phổ biến là từ 90 – 100 cm [31]. Cá vồ đém có giá trị về mặt kinh tế và được xem là đối tượng tiềm năng cho nuôi trồng thủy sản. Trong điều kiện tự nhiên, thông tin về tuổi thành thực sinh dục của cá vồ đém còn hạn chế, tuy nhiên, trong điều kiện nuôi nhốt, chúng được ghi nhận là đạt đến độ tuổi này ở khoảng 4 – 5 năm [73]. Một quần đàn cá vồ đém ở hạ lưu được ghi nhận, bắt đầu từ Pakse (Lào) đến ĐBSCL (Việt Nam), kể cả hệ thống sông Tonlé Sap [31].

1.2.3.3. Đặc điểm di cư

Cá vồ đém thuộc nhóm cá trắng có tập tính di cư trong sông (*potamodromy*) [31, 69, 71]. Chúng thường di cư vào tháng 4 – 7 để tìm kiếm thức ăn và sinh sản [74]. Theo Poulsen and Valbo-Jørgensen (2001), sự thay đổi mực nước và màu nước là yếu tố khởi động cho quá trình di cư của cá vồ đém [6]. Ở phía dưới thác Khôn, cá thường di cư xuôi dòng xuống hạ lưu khi mùa mưa bắt đầu và ngược dòng lên thượng lưu vào đầu mùa khô. Chúng thường di chuyển giữa sông Mekong và các dòng nhánh [6] và di cư vào vùng ngập [68]. Trứng hoặc cá con được ghi nhận xuất hiện trong thời gian dài (tháng 3 – 9) và chủ yếu là vào tháng 5 – 7, do đó, cá vồ đém được cho

là đẻ trứng vào đầu mùa mưa [68] và vào tháng 8 [74] (**Hình 1.6C**). Thời gian trôi nổi của trứng và ấu trùng cá vồ đém được ghi nhận là khoảng 33 – 35 ngày [68].

Thông tin chính xác về địa điểm sinh sản của cá vồ đém hiện vẫn còn nhiều tranh cãi. Theo Rainboth (1996), chúng có thể di cư vào các vùng ngập lũ để sinh sản [68], trong khi Bardach (1959) cho rằng quá trình đẻ trứng diễn ra ở đoạn sông gần Stung Treng và ấu trùng mất khoảng 6 – 8 ngày để di chuyển đến sông Bassac (Campuchia) [75]. Poulsen *et al.* (2004) cho rằng cá vồ đém di cư qua thác Khôn khi mùa khô bắt đầu nhằm mục đích sinh sản và bãi đẻ có thể nằm ở khu vực ngay phía trên thác Khôn. Sau khi sinh sản, ấu trùng di chuyển vào các vùng ngập và sinh sống ở đó trong suốt mùa mưa. Khi mùa mưa kết thúc và mực nước bắt đầu rút, cả cá con và cá trưởng thành quay trở lại sông. Một số cá thể lớn, đặc biệt là ở khu vực thượng lưu, có thể ở lại sông ngay cả trong mùa mưa. Cá thường tìm đến các vực sâu để trú ẩn trong mùa khô. Khi đạt đến tuổi thành thực, chúng rời khỏi nơi trú ẩn mùa khô và bắt đầu cuộc di cư sinh sản ngược dòng vào thời điểm bắt đầu mùa mưa [74].



Hình 1.6. Hình thái ngoài (A), bản đồ phân bố (chấm màu tím, B) [76] và hướng di cư theo mùa (C) [6] của cá vồ đém *Pangasius larnaudii*

(1) đốm đen lớn ở gốc vây ngực

1.3. Tổng quan về phương pháp nghiên cứu

1.3.1. Kỹ thuật giải trình tự DNA dựa trên vị trí cắt hạn chế

Công nghệ giải trình tự thế hệ mới (*Next Generation Sequencing* – NGS) đã phát triển nhanh chóng và trở thành công cụ quan trọng trong di truyền học hiện đại nhờ hiệu suất cao, độ chính xác lớn và chi phí ngày càng giảm. So với phương pháp Sanger truyền thống chỉ cho phép giải trình tự từng đoạn gen riêng lẻ (thường nhỏ hơn 1,500 bp), NGS cho phép giải trình tự hàng triệu đoạn DNA song song trong cùng một phản ứng, từ đó tạo ra một khối lượng dữ liệu khổng lồ phục vụ nghiên cứu đa dạng sinh học, phát sinh loài và di truyền quần thể [23].

Trong số các phương pháp dựa trên công nghệ giải trình tự NGS, kỹ thuật giải trình tự DNA dựa trên vị trí cắt giới hạn (*Restriction site-Associated DNA Sequencing* – RAD-seq), được giới thiệu lần đầu bởi Baird *et al.* (2008) [23], đã mở ra một hướng tiếp cận hiệu quả trong việc thu thập hàng nghìn chỉ thị đa hình nucleotide đơn (SNPs) phân bố rộng khắp hệ gen mà không cần giải trình tự toàn bộ bộ gen. Về nguyên lý, RAD-seq sử dụng các enzyme cắt hạn chế để phân cắt hệ gen thành các đoạn DNA có kích thước xác định, sau đó gắn các đoạn adapter và tiến hành giải trình tự trên nền tảng NGS, giúp giảm độ phức tạp của dữ liệu và tăng khả năng phát hiện biến dị di truyền trên toàn hệ gen [23, 77]. Kỹ thuật này đặc biệt hữu ích cho các loài chưa có hệ gen tham chiếu vì không yêu cầu thông tin ban đầu về trình tự, đồng thời thích hợp với các nghiên cứu có số lượng mẫu lớn và phạm vi phân bố rộng [78, 79].

Cho đến nay, nhiều biến thể của RAD-seq đã được phát triển nhằm nâng cao độ chính xác, giảm chi phí và tối ưu hóa quy trình tạo thư viện gen, tiêu biểu như double digest RAD-ddRAD (sử dụng 2 enzyme cắt hạn chế) [80], 2b-RAD (sử dụng enzyme cắt giới hạn thuộc nhóm IIB) [81] và EzRAD (sử dụng một hay nhiều enzyme cắt giới hạn) [82]. Các phương pháp này khác nhau chủ yếu ở số lượng enzyme sử dụng, cách chọn lọc kích thước đoạn DNA, và yêu cầu về thiết kế adapter, qua đó cho phép điều chỉnh số lượng locus thu được và độ sâu giải trình tự tùy theo mục tiêu nghiên cứu [79]. Đáng chú ý, kỹ thuật EzRAD, được mô tả bởi Toonen *et al.* (2013), đã đơn

giảm hóa đáng kể quy trình RAD-seq bằng cách cho phép sử dụng linh hoạt các enzyme cắt hạn chế kết hợp với việc sử dụng bộ kit thương mại Illumina TruSeq (Mỹ) trong bước chuẩn bị thư viện gen. Phương pháp này không yêu cầu thiết kế adapter đặc thù cho từng enzyme, cho phép thử nghiệm linh hoạt nhiều loại hoặc tổ hợp enzyme cắt giới hạn khác nhau mà không cần đầu tư vật liệu mới. EzRAD phù hợp với các loài không mô hình (*non-model organisms*) vì có thể tạo thư viện gen từ lượng DNA đầu vào nhỏ, không cần hệ gen tham chiếu, và vẫn cho phép phát hiện hàng nghìn SNPs có độ bao phủ cao, kể cả khi chỉ một phần nhỏ dữ liệu đạt tiêu chuẩn chất lượng, đây là ưu điểm vượt trội so với các kỹ thuật khác như GBS [83], ddRAD hay 2bRAD, vốn đòi hỏi thiết kế adapter chuyên biệt và quy trình phức tạp hơn.

Nhờ các ưu điểm này, RAD-seq nói chung và EzRAD nói riêng đã trở thành công cụ mạnh mẽ trong lĩnh vực nghiên cứu di truyền quần thể, bao gồm: phân tích cấu trúc quần thể, đánh giá dòng gen, xác định các locus chịu chọn lọc tự nhiên để hiểu rõ cơ chế thích nghi sinh thái, cũng như định lượng tác động của các yếu tố con người lên đa dạng di truyền [79].

1.3.2. Các công cụ lắp ráp *de novo* hệ gen sử dụng dữ liệu RAD-seq

Đối với các loài chưa có hệ gen tham chiếu, quá trình lắp ráp *de novo* hệ gen (*de novo assembly*) đóng vai trò then chốt trong phân tích RAD-seq [84]. Trong trường hợp này, “lắp ráp *de novo* hệ gen” không nhằm xây dựng một hệ gen hoàn chỉnh, mà chỉ tạo ra tập hợp trình tự đại diện cho một phần nhỏ của hệ gen - chủ yếu bao gồm các đoạn liền kề vị trí nhận biết của enzyme cắt hạn chế được thu nhận bởi phương pháp RAD [85]. Về mặt phân tích, tập hợp trình tự này được sử dụng như một hệ gen tham chiếu rút gọn (*reduced reference genome*) để căn chỉnh các đoạn đọc và xác định biến thể di truyền (SNPs) giữa các cá thể cho phân tích di truyền quần thể.

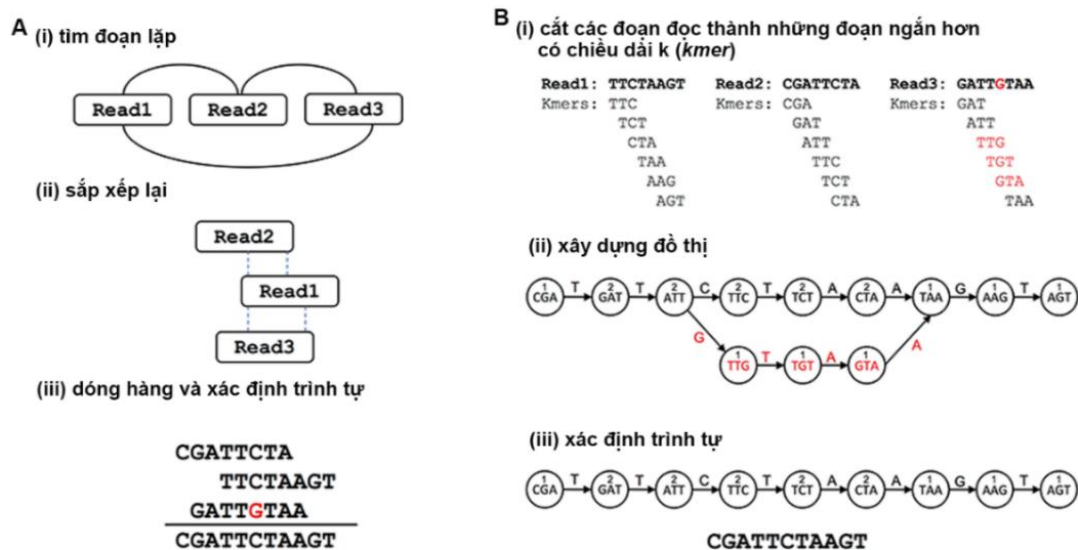
Quá trình lắp ráp *de novo* hệ gen từ các đoạn đọc ngắn bao gồm ba giai đoạn chính: (i) kết nối các đoạn đọc để tạo ra các đoạn trình tự liên tục (*contig*), (2) sắp xếp và liên kết các contig thành các đoạn khung (*scaffold*), và (3) sử dụng thông tin từ các đoạn đọc để lấp đầy các khoảng trống (*gap*) trong scaffold (*gap filling*) [86]. Các bộ

công cụ lắp ráp *de novo* được phát triển dựa trên hai thuật toán chủ yếu là đồ thị (*graph-based*) và tham lam (*greedy-based*).

❖ **Thuật toán đồ thị** biểu diễn các đoạn đọc dưới dạng một tập hợp các đỉnh và vùng chồng lặp (*Overlap*) giữa các đoạn đọc là các cạnh có hướng kết nối các đỉnh để tạo thành một đồ thị hoàn chỉnh thông qua đường Hamiltonian (đi qua các đỉnh chính xác 1 lần) hoặc đường Eulerian (đi qua tất cả các cạnh chính xác 1 lần). Thuật toán đồ thị gồm 2 dạng là Overlap – Layout – Consensus (OLC) và *de Bruijn* graph (DBG).

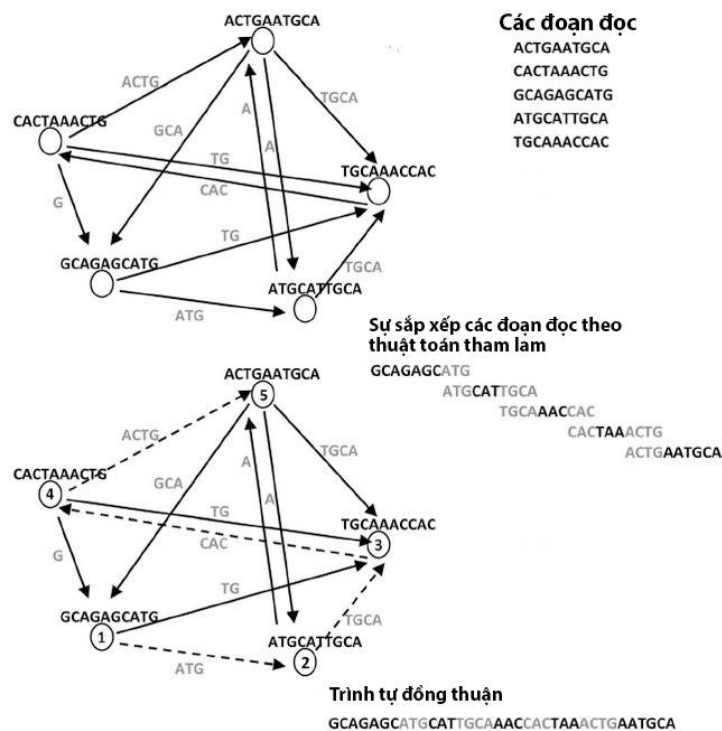
+ **Thuật toán OLC** được thực hiện qua 3 bước: (i) so sánh từng cặp đoạn đọc để tìm ra các đoạn lặp và xây dựng đồ thị với các đoạn đọc là nút và các đoạn lặp là cạnh; (ii) đơn giản hóa đồ thị và xác định các contig dọc theo đường Hamiltonian; (iii) sắp xếp các contig để tạo ra trình tự đồng thuận (*consensus sequences*) (**Hình 1.7A**). Một số phần mềm ứng dụng thuật toán OLC là CAP3 [87], Newbler [88].

+ **Thuật toán DBG** gồm 3 bước: (i) các đoạn đọc được chia thành các đoạn ngắn hơn có chiều dài k , được gọi là *kmer*; (ii) xây dựng một đồ thị *de Bruijn* với các *kmer* là các nút và các đoạn lặp có kích thước $k-1$ là các cạnh; (iii) xác định trình tự đồng thuận dựa trên đường Eulerian (**Hình 1.7B**). Thuật toán DBG được ứng dụng trong các phần mềm như Velvet [89], SOAPdenovo2 [90] và Stack [91].



Hình 1.7. Mô phỏng các bước thực hiện của thuật toán dựa trên đồ thị trong lắp ráp *de novo* hệ gen. (A) OLC và (B) BDG [92]

❖ **Thuật toán tham lam** giải quyết vấn đề bằng cách liên tục tìm kiếm và đưa ra lựa chọn tối ưu tại chỗ ở mỗi giai đoạn, với kỳ vọng đạt được kết quả tốt nhất trên toàn bộ quá trình [93]. Thuật toán tham lam được thực hiện bằng cách kết hợp các đoạn đọc thành những đoạn contig dựa vào số điểm vùng lặp cao nhất và quá trình này được lặp lại cho đến khi chỉ còn lại một trình tự duy nhất hoặc không thể tiếp tục thực hiện (**Hình 1.8**) [86]. Thuật toán này được ứng dụng trong các phần mềm lắp ráp như SSAKE [94], SHARCGS [95], Rainbow [96] và VSEARCH [97].

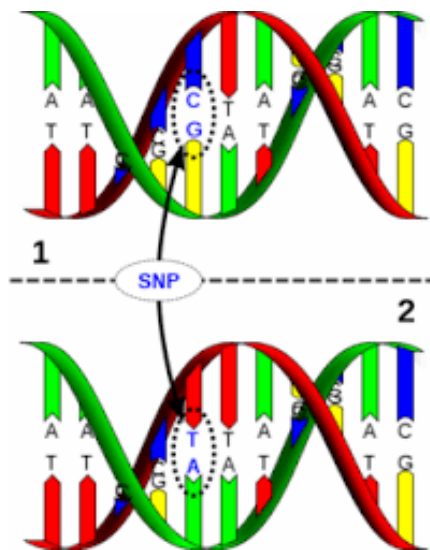


Hình 1.8. Mô phỏng các bước thực hiện của thuật toán tham lam trong lắp ráp *de novo* hệ gen [86]

Thuật toán đồ thị có ưu điểm là khả năng xử lý dữ liệu phức tạp và độ chính xác cao, nhưng lại yêu cầu tài nguyên tính toán lớn và có độ phức tạp thuật toán cao. Ngược lại, thuật toán tham lam có tốc độ tính toán nhanh và yêu cầu tài nguyên tính toán thấp hơn, song độ chính xác thường thấp hơn và gặp khó khăn trong việc xử lý các vùng lặp lại. Việc lựa chọn thuật toán phù hợp phụ thuộc vào nhiều yếu tố, bao gồm kích thước và độ phức tạp của hệ gen, tài nguyên tính toán có sẵn và yêu cầu về độ chính xác của kết quả lắp ráp [86, 92].

1.3.3. Chỉ thị đa hình nucleotide đơn (SNPs)

Chỉ thị phân tử DNA là các đoạn trình tự đặc trưng trong hệ gen thể hiện sự đa hình di truyền thông qua các biến đổi như thay thế, chèn hoặc mất đoạn nucleotide giữa các cá thể trong cùng loài hoặc giữa các loài khác nhau [98]. Trong lịch sử nghiên cứu di truyền, nhiều chỉ thị phân tử đã được phát triển và ứng dụng rộng rãi như đa hình độ dài đoạn cắt giới hạn (*Restriction fragment length polymorphism* - RFLP), đa hình độ dài đoạn cắt khuếch đại (*Amplified Fragment Length Polymorphism* - AFLP), đa hình nhân bản DNA ngẫu nhiên (*Random amplification of polymorphic DNA* - RAPD) và microsatellite [98]. Tuy nhiên, với sự bùng nổ của các công nghệ giải trình tự thế hệ mới, chỉ thị đa hình nucleotide đơn (SNP) đã trở thành công cụ mạnh mẽ và tối ưu nhất [94]. SNP được định nghĩa là sự thay thế của một nucleotide đơn (A, T, G hoặc C) tại một vị trí cụ thể trong hệ gen, tạo nên dạng biến dị di truyền phổ biến nhất hiện nay (**Hình 1.9**) [99, 100].



Hình 1.9. Hình minh họa về chỉ thị đa hình nucleotide đơn (SNP)

So với các chỉ thị truyền thống, SNP sở hữu những ưu thế vượt trội về độ phân giải và quy mô dữ liệu (**Bảng 1.1**). Với mật độ phân bố dày đặc, trung bình cứ 300–1.000 bp lại có một vị trí đa hình (cả vùng mã hóa và không mã hóa), SNP cho phép bao phủ hệ gen với độ phân giải cao hơn gấp nhiều lần so với chỉ thị DNA ti thể (kích thước khoảng 16.000-18.000 bp) hay microsatellite (thường phân bố thưa thớt từ

5.000–50.000 bp) [101, 102]. Với tỉ lệ đột biến thấp và đặc tính hai alen (*biallelic*), SNP mang lại tính ổn định cao, dễ dàng chuẩn hóa và so sánh dữ liệu giữa các nghiên cứu khác nhau. Đặc biệt, thông qua các kỹ thuật như RAD-seq, GBS và RNA-seq, việc phát hiện đồng thời hàng nghìn đến hàng trăm nghìn locus SNP trở nên khả thi với chi phí thấp và độ chính xác cao, ngay cả đối với những đối tượng nghiên cứu chưa có thông tin về hệ gen tham chiếu [78, 79].

Bên cạnh đó, SNPs còn cho phép phân tách giữa các quá trình tiến hóa ngẫu nhiên và chọn lọc tự nhiên bằng cách phân loại thành locus trung tính (*neutral loci*) và locus thích nghi (*adaptive loci*). Trong khi các locus trung tính giúp làm sáng tỏ dòng gen và trôi dạt di truyền, các locus thích nghi cung cấp bằng chứng trực tiếp về áp lực chọn lọc và khả năng ứng phó của quần thể trước biến đổi môi trường [103]. Chính vì vậy, SNPs hiện đang được ứng dụng rộng rãi trong di truyền học hiện đại nhằm lập bản đồ gen, xác định nguồn gốc cá thể và xây dựng các chiến lược bảo tồn nguồn gen bền vững trước những thách thức của biến đổi khí hậu [104].

Bảng 1.1. So sánh ưu và nhược điểm của các loại chỉ thị phân tử phổ biến

Loại chỉ thị	Ưu điểm	Nhược điểm
DNA ti thể	Dễ khuếch đại, di truyền dòng mẹ và không tái tổ hợp.	Chỉ phản ánh lịch sử dòng mẹ và lượng thông tin hạn chế.
Microsatellite	Tính đa hình cao (đa alen), nhạy với các biến dị di truyền gần đây.	Dễ sai số do alen giả, khó chuẩn hóa giữa các phòng thí nghiệm, số lượng locus hạn chế, chi phí phát triển đoạn mồi ban đầu cao.
SNPs	Mật độ dày đặc, độ ổn định cao, phân tích quy mô lớn nhờ NGS	Mỗi locus đơn lẻ ít thông tin, yêu cầu hệ thống xử lý tin sinh học mạnh.

1.3.4. Phương pháp đánh giá đa dạng di truyền, khác biệt di truyền và cấu trúc quần thể

Đa dạng di truyền là một trong ba cấp độ của đa dạng sinh học, thể hiện sự phong phú và khác biệt về vật chất di truyền giữa các cá thể, quần thể hoặc vùng địa lý trong cùng một loài. Đây là nền tảng quyết định khả năng tiến hóa và thích nghi của sinh vật

trước sự biến đổi môi trường và áp lực chọn lọc tự nhiên [105]. Mức độ đa dạng di truyền cao giúp quần thể duy trì tiềm năng thích nghi dài hạn, trong khi sự suy giảm đa dạng - thường do giao phối cận huyết hoặc thu hẹp kích thước quần thể hiệu quả (*Effective population size-Ne*) - có thể dẫn đến suy thoái di truyền, làm giảm sức sống, khả năng sinh sản và tăng nguy cơ tuyệt chủng. Theo lý thuyết di truyền quần thể, kích thước hiệu quả tối thiểu cần duy trì ở mức trên 5.000 để đảm bảo cân bằng giữa đột biến và trôi dạt di truyền, giúp bảo tồn tiềm năng thích nghi lâu dài [106].

Để đánh giá đa dạng di truyền, các thông số phổ biến bao gồm số lượng alen và alen hiệu quả trên mỗi locus, tỉ lệ dị hợp tử quan sát (H_o) và mong đợi (H_e) và hệ số cận huyết (F_{IS}) [106]. Các thông tin về ý nghĩa sinh học và công thức tính của mỗi thông số đánh giá mức độ đa dạng di truyền được trình bày ở **Phụ lục 1**. Hiện nay, cùng với sự phổ biến của các chỉ thị phân tử như SNPs, hàng loạt công cụ tin sinh học được phát triển để phân tích dữ liệu di truyền quần thể, như FSTAT [107], GENALEX [108], GENODIVE [109], PLINK [110] và GENEPOP [111].

Biến dị di truyền phản ánh sự khác biệt giữa các kiểu gen trong cùng một quần thể, bắt nguồn từ đột biến, trôi dạt di truyền, dòng gen và chọn lọc tự nhiên [112]. Trong đó, trôi dạt di truyền (*genetic drift*) là sự thay đổi ngẫu nhiên tần số alen qua các thế hệ, đặc biệt là ở các quần thể có kích thước nhỏ, dẫn đến suy giảm biến dị trong quần thể và tăng khác biệt giữa các quần thể [113]. Ngược lại, dòng gen có tác dụng làm giảm sự phân hóa bằng cách trao đổi vật chất di truyền giữa các quần thể [114]. Chọn lọc tự nhiên có thể làm giảm biến dị (qua chọn lọc loại trừ) hoặc duy trì biến dị (qua chọn lọc cân bằng hoặc phân hóa) tùy theo điều kiện môi trường và lợi thế thích nghi của alen [115]. Ngoài ra, các sự kiện lịch sử như hiệu ứng nút cổ chai (*bottleneck*) và hiệu ứng người sáng lập (*founder effect*) làm giảm mạnh tính đa hình di truyền và có thể gây mất mát lâu dài về đa dạng di truyền [113].

Phân tích cấu trúc quần thể là bước quan trọng nhằm xác định mức độ kết nối và phân tách di truyền giữa các nhóm cá thể. Các phương pháp truyền thống như đánh dấu, phân tích hình thái học hay xương đá tai cung cấp thông tin sinh thái cơ bản,

nhưng chỉ thị phân tử mang lại độ chính xác cao hơn, cho phép xác định cấu trúc quần thể tiềm ẩn và các dòng gen giữa các vùng địa lý. Các mô hình thống kê hiện đại như STRUCTURE [116], DAPC [117] thường được sử dụng để mô tả mức độ phân hóa di truyền (F_{ST} , AMOVA) và đánh giá ảnh hưởng của các rào cản địa lý hoặc môi trường đến sự phân bố alen [118].

Trong các hệ thống sông lớn như Mekong, đặc điểm thủy văn và các rào cản vật lý (đập, hồ chứa, phân mảnh dòng chảy) có thể tạo ra sự phân tách di truyền giữa các quần thể cá, từ những quần thể đồng nhất (*panmictic*) - nơi dòng gen mạnh và giao phối ngẫu nhiên - đến những quần thể cô lập với sự khác biệt di truyền rõ rệt [119]. Do đó, đánh giá đa dạng và cấu trúc di truyền của các loài cá sông có ý nghĩa đặc biệt trong quản lý nghề cá, bảo tồn nguồn gen và phục hồi quần đàn trong bối cảnh tác động của con người và biến đổi khí hậu ngày càng gia tăng.

1.3.5. Các phương pháp nghiên cứu sự di cư của cá

1.3.5.1. Các phương pháp phụ thuộc nghề cá

Phương pháp này chủ yếu dựa trên kiến thức sinh thái địa phương (*local ecological knowledge*) và giám sát các hoạt động khai thác thủy sản. Thông qua việc ghi nhận thời điểm và địa điểm ngư dân bắt được cá mang trứng hoặc cá con, các nhà nghiên cứu có thể đưa ra giả thuyết về hướng di cư và vị trí bãi đẻ của các loài cá. Mặc dù với chi phí thấp và dễ triển khai trên phạm vi rộng, phương pháp này tồn tại tính chủ quan cao và dữ liệu có thể bị sai lệch do khả năng định danh loài chưa chính xác hoặc báo cáo không đầy đủ từ phía ngư dân [7, 31, 120].

1.3.5.2. Các phương pháp không phụ thuộc nghề cá

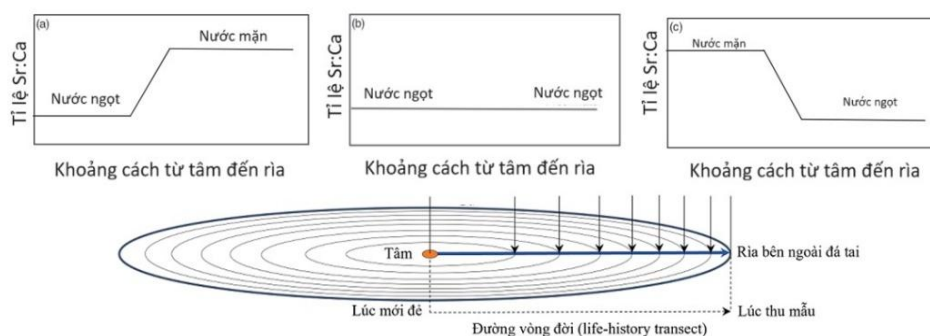
a. Nghiên cứu đặc điểm sinh học sinh sản và nuôi nhốt trong phòng thí nghiệm

Nghiên cứu sinh học sinh sản tập trung phân tích mô học tuyến sinh dục để xác định các giai đoạn phát triển của trứng và tinh trùng, từ đó dự báo thời điểm cá bắt đầu di cư [121]. Các thí nghiệm trong phòng thí nghiệm cho phép xác định các ngưỡng phản ứng sinh lý trước các yếu tố kích hoạt di cư như tốc độ dòng chảy, nhiệt độ và

cường độ ánh sáng dưới điều kiện kiểm soát [122]. Tuy nhiên, môi trường nuôi nhốt thường khó tái lập hoàn toàn các hành vi sinh thái phức tạp của cá trong tự nhiên [123].

b. Phân tích thành phần hóa học của đá tai

Phân tích thành phần hóa học đá tai là công cụ phản ánh chi tiết lịch sử cư trú của cá thông qua việc hấp thụ các nguyên tố vi lượng vào cấu trúc canxi cacbonat (CaCO_3) theo thời gian [124]. Bằng cách định lượng tỉ lệ của các “chất chỉ thị” như Strontium (Sr) và Barium (Ba) so với Calcium (Ca), các nhà khoa học có thể xác định chính xác thời điểm cá di chuyển giữa các môi trường sống khác nhau. Sự thay đổi nồng độ Sr:Ca từ tâm đá tai (lúc cá mới đẻ) đến rìa ngoài (lúc thu mẫu) cho phép dự đoán mô hình di cư từ nơi sinh sản ở nước ngọt đến vùng sinh trưởng ở nước mặn hoặc ngược lại (**Hình 1.10**) [125, 126]. Dù có độ phân giải thời gian cao, quy trình này mang tính xâm lấn bắt buộc (phải gây chết cá) và đòi hỏi chi phí đầu tư thiết bị phân tích hiện đại.



Hình 1.10. Tỉ lệ Sr:Ca từ lúc cá mới đẻ (tâm đá tai) đến lúc thu mẫu (rìa bên ngoài)
(a) cá đẻ ở nước ngọt và sinh trưởng ở nước mặn, (b) cá chỉ sống ở nước ngọt,
(c) cá đẻ ở nước mặn và sinh trưởng ở nước ngọt [125]

c. Phương pháp gắn thẻ

Phương pháp gắn thẻ và bắt lại, từ các thẻ vật lý đơn giản đến các thiết bị theo dõi điện tử phức tạp (như thẻ PIT hoặc thẻ vệ tinh), được sử dụng để khảo sát trực tiếp tuyến đường và môi trường sống của cá trong thời gian thực [127]. Tại lưu vực sông Mekong, phương pháp này đã được triển khai hiệu quả tại Campuchia [128] và Thái Lan [129, 130]. Để giảm thiểu tính xâm lấn, hệ thống đếm cá điện tử và camera

dưới nước thường được triển khai tại các khu vực dòng chảy hẹp để giám sát mật độ và thời gian di cư mà không gây căng thẳng cho cá [131, 132]. Tuy nhiên, các kỹ thuật này bị giới hạn bởi tỉ lệ thu hồi thể thấp và sự phụ thuộc vào độ trong của nước.

d. Công cụ di truyền

Công cụ di truyền thể hiện tính ưu việt trong việc dự đoán mô hình di cư dựa trên nguyên lý phân tích dòng gen và sự biến động tần số alen giữa các quần thể mà không cần quan sát trực tiếp hay gây chết cá [133]. Thông qua các thuật toán tin sinh học, phương pháp này cho phép chuyển đổi dữ liệu DNA thành các dữ liệu về hướng di cư và sự kết nối quần thể. Chẳng hạn như công cụ divMigrate cho phép xác định hướng di cư ưu thế thông qua sự trao đổi di truyền không đối xứng, từ đó phân định rõ vai trò của từng khu vực là “quần thể nguồn” (nơi xuất phát/sinh sản) hay “quần thể nhận” (nơi cư trú/sinh trưởng) [134]; TreeMix sử dụng phương pháp hàm tiện tối đa (*maximum likelihood*) để lồng ghép các sự kiện di cư vào sơ đồ quan hệ giữa các quần thể, giúp giải thích sự pha trộn di truyền phức tạp giữa các hệ thống sông [135].

Mỗi phương pháp nghiên cứu di cư đều có ưu và nhược điểm riêng (**Bảng 1.2**). Do đó, việc áp dụng đồng thời các phương pháp là cần thiết để khắc phục các hạn chế đơn lẻ, đồng thời hỗ trợ kiểm chứng chéo dữ liệu. Sự kết hợp này mang lại cái nhìn toàn diện và chính xác về đặc điểm di cư phức tạp của cá trong hệ thống thủy văn quy mô lớn như sông Mekong [136].

Bảng 1.2. Ưu và nhược điểm của các phương pháp nghiên cứu di cư của cá

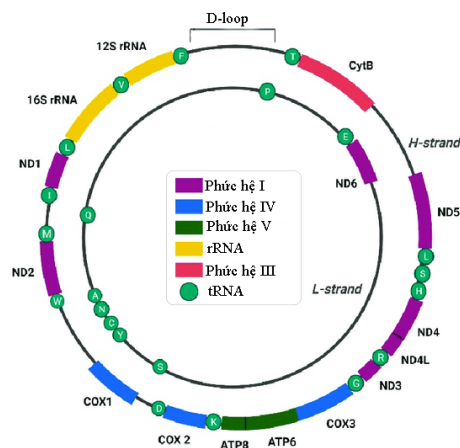
Phương pháp	Ưu điểm	Nhược điểm
Phụ thuộc nghề cá	Chi phí thấp, dễ thực hiện	Tính chủ quan và định danh loài không chính xác
Nghiên cứu sinh học sinh sản và nuôi nhốt	Làm rõ thời điểm và phản ứng sinh lý	Khó thiết lập hành vi tự nhiên trong nuôi nhốt.
Phân tích thành phần hóa học của đá tai	Ghi lại lịch sử cư trú suốt vòng đời	Xâm lấn (gây chết cá), đòi hỏi chuyên môn, chi phí thiết bị cao
Gắn thẻ và máy đếm	Dữ liệu thời gian thực, trực quan	Tỉ lệ thu hồi thể thấp, phụ thuộc độ trong của nước
Công cụ di truyền	Không xâm lấn, quy mô quần thể lớn.	Yêu cầu chuyên môn và kỹ thuật phân tích chuyên sâu

1.3.6. Lắp ráp, chú giải hệ gen ti thể và xác định mã vạch RAD

1.3.6.1. Đặc điểm và cấu trúc của hệ gen ti thể

Ti thể, thường được xem là “nhà máy năng lượng” của tế bào nhân thực, là các bào quan thiết yếu có mặt khắp nơi, chịu trách nhiệm chính trong việc sản xuất năng lượng dưới dạng ATP thông qua quá trình phosphoryl hóa oxy hóa [137]. Ti thể sở hữu hệ gen riêng biệt, gọi là DNA ti thể (*mitochondrion DNA* – mtDNA), có khả năng tự nhân lên độc lập. Ở động vật, mtDNA có cấu trúc vòng và kích thước khoảng 15 – 20 kilobase (kb) [137].

Hệ gen ti thể của động vật gồm 37 gen và không chứa intron. Cụ thể gồm 13 gen mã hóa protein (*protein coding genes*-PCGs), 2 gen mã hóa RNA ribosome (rRNA), 22 gen mã hóa RNA vận chuyển (*transfer RNA* – tRNA), và một vùng không mã hóa (*non-coding region* – NCR) [138]. Trong số 13 PCGs, có 7 gen mã hóa cho phức hệ I (*NADH-ubiquinone oxidoreductase*) là *ND1-ND6* và *ND4L*; 1 gen mã hóa cho phức hệ III, là cytochrome b (*Cyt b*); 3 gen mã hóa phức hệ IV, là *COX1-3*; và 2 gen mã hóa cho phức hệ V, là *ATP6* và *ATP8*. Các gen rRNA (*12S* và *16S*) và tRNA nằm xen kẽ giữa các gen PCGs. Vùng NCR hay *D-loop* đóng vai trò quan trọng vì chứa các điểm khởi đầu và kết thúc quá trình sao chép (Hình 1.11).



Hình 1.11. Minh họa cấu trúc của hệ gen ti thể ở động vật [138]

Màu sắc tương ứng với các phức hệ gen, gen rRNA và tRNA trong hệ gen ti thể. Các gen nằm ở vòng tròn trong và ngoài được mã hóa tương ứng trên chuỗi nhẹ và chuỗi nặng.

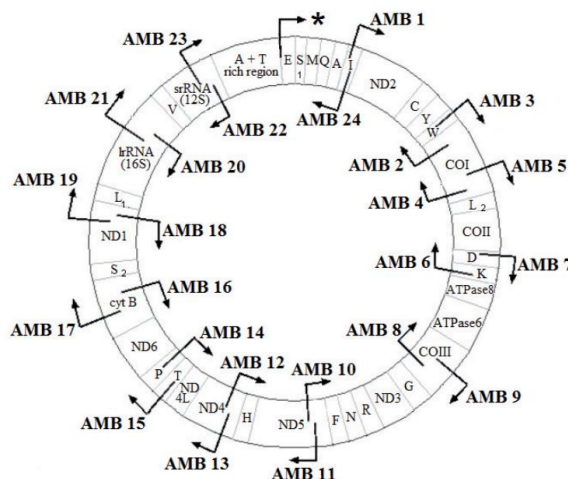
Gen ti thể có một số đặc điểm di truyền đặc trưng, bao gồm di truyền theo dòng mẹ, tốc độ đột biến cao, số lượng bản sao lớn và không xảy ra tái tổ hợp [139]. Nhờ những đặc điểm này, đặc biệt là cấu trúc vòng bền vững và tốc độ đột biến cao, mtDNA là công cụ phân tử mạnh mẽ và được ứng dụng rộng rãi trong việc định danh loài, phân tích mối quan hệ tiến hóa và đa dạng di truyền của quần thể [140].

1.3.6.2. Phương pháp lắp ráp hệ gen ti thể

Sự phát triển mạnh mẽ của các phương pháp giải trình tự hệ gen đã thúc đẩy mạnh mẽ việc công bố hệ gen ti thể hoàn chỉnh, đặc biệt ở các loài cá, với số lượng tăng từ 21 loài (năm 2000) lên hơn 4.200 loài (năm 2025, theo MitoFish <https://mitofish.aori.u-tokyo.ac.jp/>, cập nhật ngày 08/02/2025). Việc xác định trình tự và lắp ráp mtDNA được thực hiện thông qua bốn phương pháp sau:

(i) Khuếch đại PCR thông thường và giải trình tự Sanger

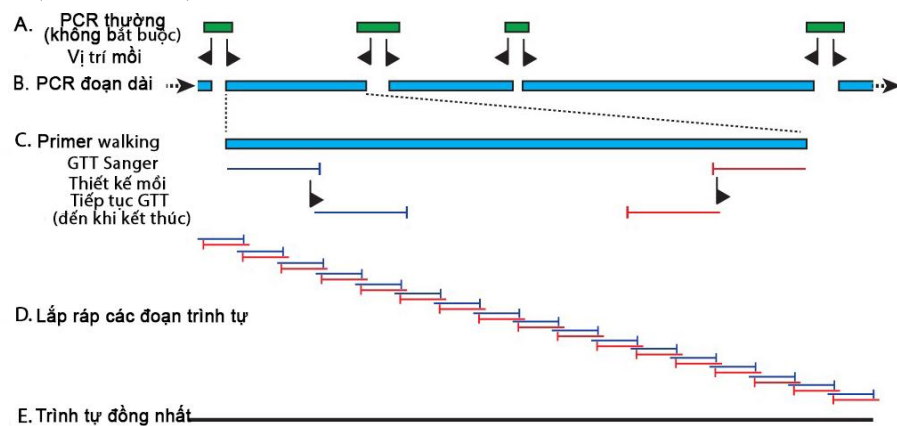
Phương pháp này sử dụng tập hợp các cặp mồi được thiết kế dựa trên trình tự mtDNA của loài có quan hệ gần gũi. Các mồi này khuếch đại các đoạn DNA ngắn với vùng chồng lấp (100 – 300 bp), sau đó các đoạn này được giải trình tự Sanger và lắp ráp để tạo thành hệ gen hoàn chỉnh. Vị trí của 24 đoạn mồi (AMB 1-24) được sử dụng trong khuếch đại và lắp ráp hệ gen ti thể được minh họa ở **Hình 1.12** [141].



Hình 1.12. Vị trí của 24 đoạn mồi (AMB 1-24) được sử dụng để khuếch đại và lắp ráp hệ gen ti thể [141]

(ii) Khuếch đại PCR chuỗi dài và giải trình tự Sanger bằng “primer walking”

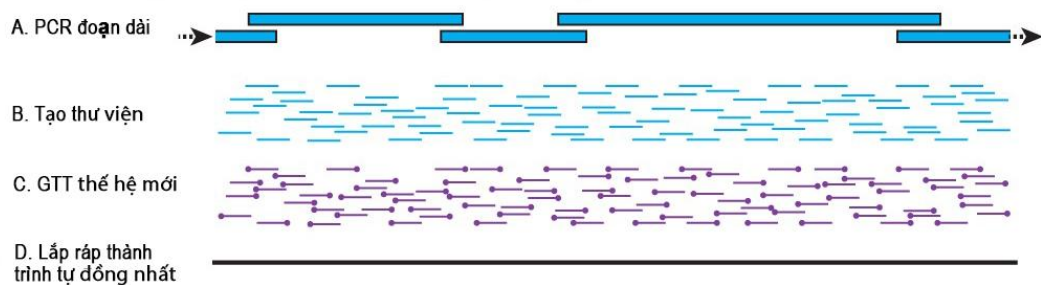
Kỹ thuật PCR chuỗi dài (*long-range PCR*) [142] được sử dụng để khuếch đại hai hay nhiều phân đoạn DNA lớn (khoảng 2 kb). Do giới hạn của Sanger, chiến lược “primer walking” được áp dụng để giải trình tự các đoạn dài này. Theo đó, mỗi mới cho lần giải trình tự tiếp theo được thiết kế dựa trên phần cuối của trình tự vừa thu được, lặp lại cho đến khi toàn bộ phân đoạn được giải trình tự và lắp ráp thành công [143]. Các đoạn trình tự được tiến hành lắp ráp để thu nhận một trình tự đồng nhất là hệ gen ti thể (**Hình 1.13**).



Hình 1.13. Quy trình khuếch đại, giải trình tự và lắp ráp hệ gen ti thể ứng dụng kỹ thuật PCR chuỗi dài và giải trình tự Sanger bằng “primer walking” [141]

(iii) Khuếch đại PCR chuỗi dài và giải trình tự thế hệ mới

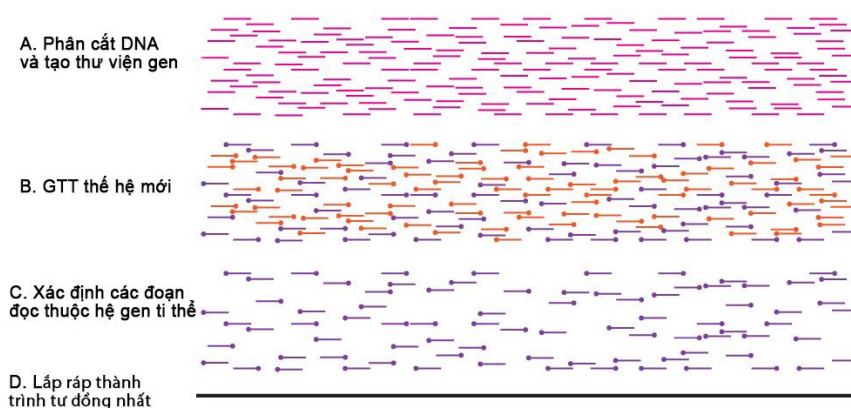
Tương tự, PCR chuỗi dài tạo ra các phân đoạn lớn (khoảng 2 kb), nhưng thay vì Sanger, các phân đoạn này được sử dụng để tạo thư viện và giải trình tự bằng các nền tảng NGS (như Roche 454 hoặc Illumina). Các đoạn đọc (*reads*) thu được sau đó được lắp ráp thành mtDNA hoàn chỉnh bằng các công cụ tin sinh học (**Hình 1.14**).



Hình 1.14. Quy trình khuếch đại, giải trình tự và lắp ráp hệ gen ti thể ứng dụng kỹ thuật PCR chuỗi dài và giải trình tự thế hệ mới [141]

(iv) Tạo thư viện gen và giải trình tự thế hệ mới

DNA hệ gen được phân cắt thành những phân đoạn ngắn bằng enzyme cắt giới hạn, sau đó được chuẩn bị thư viện gen và giải trình tự bằng các phương pháp đọc ngắn (như hệ thống Illumina) [144] hoặc đoạn đọc dài (như Nanopore) [145]. Các đoạn đọc dài/ngắn được xử lý và lắp ráp thành hệ gen ti thể sử dụng các công cụ tin sinh học [146] (**Hình 1.15**).



Hình 1.15. Quy trình tạo thư viện gen và lắp ráp hệ gen ti thể ứng dụng kỹ thuật giải trình tự thế hệ mới [141]

1.3.6.3. Xác định mã vạch RAD

Mã vạch DNA (*DNA barcoding*) là một kỹ thuật phân tử dùng một đoạn DNA tiêu chuẩn, thường là vùng gen cytochrome c oxidase subunit I (COI) thuộc gen ti thể ở động vật, để xác định loài hoặc nhận diện biến thể địa lý thông qua việc so sánh trình tự với các cơ sở dữ liệu như Trung tâm Thông tin Công nghệ Sinh học Quốc gia (*National Center for Biotechnology Information* - NCBI) và Hệ thống Dữ liệu Mã vạch sự sống (*The Barcode of Life Data System* - BOLD) [147]. Mặc dù mã vạch DNA đã được chứng minh là công cụ hiệu quả cho định danh loài và nghiên cứu tiến hóa [140], việc chỉ sử dụng một hoặc một vài đoạn gen ti thể có hạn chế trong việc phản ánh đầy đủ mối quan hệ phát sinh loài và lịch sử tiến hóa của sinh vật [148].

Với sự phát triển của công nghệ giải trình tự NGS, đặc biệt là dữ liệu RAD-seq, đã mở ra hướng nghiên cứu di truyền quần thể ở cấp độ hệ gen. Dữ liệu RAD-seq không chỉ giúp phát hiện và sàng lọc hàng nghìn chỉ thị phân tử (như SNPs) trên toàn

bộ hệ gen mà còn có thể được ứng dụng để lắp ráp hệ gen ti thể [149–151], cung cấp thông tin giá trị cho việc xác định loài, nghiên cứu mối quan hệ phát sinh [149, 151] và phân tích cấu trúc di truyền quần thể [150].

Mã vạch RAD (*RADbarcode*) là một quy trình mới sử dụng dữ liệu RAD-seq để xác định tập hợp các trình tự mã vạch thuộc hệ gen ti thể (*Aligned mitogenome segments*), cho phép độ phân giải cao hơn trong việc phân tích các biến thể địa lý và cấu trúc quần thể [152]. Quy trình này gồm:

(1) Sắp xếp vào hệ gen và tạo trình tự đồng thuận

Các đoạn đọc thuộc gen ti thể được trích xuất từ dữ liệu RAD-seq, tiến hành sắp xếp (*mapping*) chúng với hệ gen ti thể tham chiếu và bước đầu tạo thành các trình tự hệ gen ti thể dưới định dạng fasta của từng cá thể. Các trình tự này chứa những vị trí chưa xác định và được gán bởi nucleotide N. Việc sắp xếp các đoạn đọc với hệ gen tham chiếu, các vị trí đa hình nucleotide đơn và chèn/xóa được xác định và kết hợp thành vị trí đa alen (*multiallelic*). Sau đó, các trình tự hệ gen ti thể đồng thuận (*consensus genome sequences*) của từng cá thể được tạo thành bằng cách kết hợp trình tự hệ gen ban đầu và những vị trí đa alen.

(2) Dóng hàng với hệ gen tham chiếu và thu nhận mã vạch RAD

Các trình tự hệ gen ti thể đồng thuận được dóng hàng (*align*) với các hệ gen ti thể tham chiếu sẵn có. Những vị trí nucleotide chèn, xóa hoặc khoảng trống (*gap*) và các cá thể có tỉ lệ dóng hàng thấp được loại bỏ. Cuối cùng, tập dữ liệu mã vạch RAD được thu nhận và là đầu vào cho các nghiên cứu khảo sát mối quan hệ phát sinh loài và cấu trúc di truyền quần thể [152].

1.4. Tình hình nghiên cứu liên quan đến các nội dung của luận án

1.4.1. Lắp ráp *de novo* hệ gen và xác định chỉ thị phân tử SNPs của các loài cá phân bố ở lưu vực sông Mekong

Tại lưu vực sông Mekong, nơi thông tin hệ gen của các loài sinh vật (kể cả cá) còn hạn chế [153], việc sử dụng dữ liệu từ kỹ thuật RAD-seq để lắp ráp *de novo* các hệ gen rút gọn phục vụ nghiên cứu di truyền quần thể được thực hiện trong hơn một

thập kỷ qua. Kỹ thuật này giảm độ phức tạp này chỉ giải trình tự các phân đoạn DNA tại các vị trí cắt giới hạn, do đó chỉ lắp ráp được khoảng 10% kích thước hệ gen thực tế [154]. Ứng dụng kỹ thuật này, hệ gen cá xác sọc *Pangasius macronema* lắp ráp được chỉ đạt 28,96 Mb [155] và cá phèn vàng *Polynemus melanochir* là 20,4 Mb [16]. Bên cạnh đó, sử dụng phương pháp giải trình tự toàn bộ hệ gen (*whole-genome sequencing* – WGS), Kim et al. (2018) công bố bản phác thảo đầu tiên về hệ gen của cá tra *Pn. hypophthalmus* (Siluriformes: Pangasiidae) từ một cá thể cá tra được nuôi ở ĐBSCL. Từ các đoạn đọc được giải trình tự trên hệ thống Illumina (MiSeq và HiSeq 2500), hệ gen cá tra được lắp ráp bằng phần mềm PLATANUS sử dụng thuật toán *De Bruijn*. Kích thước hệ gen cá tra là 715 Mbp, với tỉ lệ GC là 38,3%. Bản lắp ráp hệ gen bao gồm 568 scaffolds (*tập hợp các contigs*), trong đó 563 scaffolds có độ dài trên 1000 bp. Scaffold dài nhất đạt 37,5 Mbp và giá trị N50 là 14,29 Mbp [183].

Việc sử dụng các kỹ thuật RAD-seq và GBS đã trở thành công cụ hiệu quả, xác định thành công một số lượng lớn các chỉ thị SNPs cho các loài cá ở lưu vực sông Mekong. Cụ thể, các nghiên cứu đã phát hiện 1.735 SNPs từ 8 quần thể cá phèn vàng *P. melanochir* [16]; 731 SNPs từ 7 quần thể cá xác sọc *P. macronema* [155]; và 638 SNPs từ 7 quần thể cá sấu *Boesemania microlepis* [156] sử dụng kỹ thuật EzRAD; 756 SNPs trên 6 quần thể cá lăng *Hemibagrus spilopterus* (Siluriformes: Bagridae) [15]; 1.791 SNPs từ 5 quần thể cá linh *Henicorhynchus lobatus* (Cypriniformes: Cyprinidae) [17]; 3.768 SNPs từ 2 quần thể cá xác *Helicophagus leptorhynchus* (Siluriformes: Pangasiidae) [17] bằng kỹ thuật ddRAD; và 7.263 SNPs từ 7 quần thể cá tra tự nhiên và nuôi [18]. Việc phát hiện thành công hàng ngàn chỉ thị SNPs này đã vượt trội so với các chỉ thị truyền thống, cho phép các nhà nghiên cứu thực hiện các phân tích chuyên sâu về di truyền quần thể, từ đó cung cấp dữ liệu khoa học vững chắc về quản lý nguồn lợi thủy sản và bảo tồn đa dạng di truyền cho các loài cá phân bố ở khu vực này.

1.4.2. Nghiên cứu về đa dạng di truyền quần thể và mô hình di cư của các loài cá ở lưu vực hạ lưu sông Mekong

1.4.2.1. Di truyền quần thể nhóm cá không di cư

Ở nhóm cá không di cư, các nghiên cứu đều ghi nhận sự đa dạng di truyền cao và cấu trúc quần thể có sự phân tách tương ứng với các khu vực địa lý. Sử dụng chỉ thị ND2 mtDNA, Takagi *et al.* (2011) ghi nhận sự khác biệt di truyền rất lớn giữa các quần thể cá rô đồng (*Anabas testudineus*) tại Lào (2 quần thể) và Campuchia (6 quần thể) với giá trị F_{ST} dao động từ 0,52 đến 0,98 ($P < 0,05$) [20]. Đối với cá chạch lấu (*Mastacembelus favus*), các chỉ thị *cytb* mtDNA và microsatellite đã giúp xác định hai nhóm quần thể chính: nhóm hạ lưu (từ Kratié - Campuchia đến ĐBSCL - Việt Nam) và nhóm trung - thượng lưu (từ Lào đến Thái Lan). Đáng chú ý, quần thể Pakse ở trên khu vực thác Khôn tại đây không chỉ thể hiện sự đa dạng di truyền cao nhất (đa dạng haplotype 0,9105) mà còn có hiện tượng trộn lẫn thông tin di truyền giữa các nhóm và hệ số cận huyết (F_{IS}) cao [157, 158]. Ngoài yếu tố địa lý, môi trường sống cũng là tác nhân gây phân hóa; điển hình là nghiên cứu trên cá lóc (*Channa striata*) sử dụng *cytb* và D-loop mtDNA đã ghi nhận sự phân hóa đặc trưng giữa các quần thể từ hồ, vùng ngập lũ đến môi trường nuôi trồng [159].

Trong phạm vi nhỏ tại khu vực ĐBSCL, đặc điểm chung của nhóm cá không di cư là sự đa dạng di truyền cao nhưng mức độ khác biệt giữa các địa phương thấp do tính kết nối thủy văn. Các nghiên cứu sử dụng chỉ thị RAPD và ISSR trên cá rô đồng [160], cá hường (*Helostoma temminckii*) [161], cá rô biển (*Pristolepis fasciata*) [162] và cá dầy *Channa lucius* [163] đều cho thấy tỉ lệ gen đa hình dao động từ 55% đến 90% và chỉ số Shannon trung bình khoảng 0,25 – 0,39. Sự khác biệt di truyền giữa các quần thể tại Cà Mau, Đồng Tháp, Hậu Giang hay Long An là không đáng kể (ví dụ ở cá rô biển $G_{ST} = 0,051$ [162]; cá dầy $G_{ST} = 0,09$ [163]), trong khi hệ số trao đổi gene lại ở mức cao (N_m từ 5,06 đến 9,3). Gần đây, việc ứng dụng chỉ thị SNPs trên cá phèn vàng (*Polynemus melanochir*) đã ghi nhận sự kết nối cao giữa sông Tiền (5 quần thể) và sông Hậu (3 quần thể) nhưng đồng thời cảnh báo về rủi ro bảo tồn khi

kích thước quần thể hiệu quả (N_e) chỉ đạt 290 (dưới ngưỡng an toàn 500) và hệ số cận huyết tăng cao [16].

Đối với cá chạch lá tre (*M. siamensis*), nghiên cứu duy nhất về cấu trúc quần thể loài này cho đến nay được thực hiện tại khu vực Lào và Campuchia, cho thấy sự phân hóa di truyền rõ rệt theo địa lý. Sử dụng chỉ thị ND2 mtDNA, Takagi *et al.* (2006) ghi nhận giá trị F_{ST} giữa các quần thể xuyên biên giới là rất cao ($F_{ST} = 0,93 - 0,97$; $P < 0,05$), trong khi sự sai khác nội bộ trong mỗi quốc gia thấp hơn đáng kể (Lào: 0,56 và Campuchia: 0,35). Kết quả phân tích mạng lưới haplotype cũng xác nhận sự phân tách thành hai nhóm riêng biệt tương ứng với hai khu vực địa lý nêu trên. Đây là minh chứng điển hình về sự cách ly di truyền mạnh mẽ của nhóm cá không di cư trong lưu vực sông Mekong [164].

1.4.2.2. Di truyền quần thể nhóm cá di cư chặng ngắn

Đối với nhóm cá di cư chặng ngắn, các nghiên cứu về đa dạng di truyền và cấu trúc quần thể hiện nay vẫn còn khá hạn chế và chủ yếu tập trung vào việc làm rõ tác động của các rào cản địa lý và nhân tạo. Điển hình là nghiên cứu của Ackiss *et al.* (2019) trên loài cá lăng (*Hemibagrus spilopterus*) sử dụng kỹ thuật ddRAD-seq để phát hiện các chỉ thị SNPs. Kết quả không chỉ ghi nhận sự phân tách quần thể rõ rệt giữa khu vực trên thác Khôn (Pakse, Lào) và dưới thác Khôn (Kratie, Stung Treng, Campuchia), mà còn ở các nhánh sông thuộc hệ thống 3S (Sesan, Serepok và Sekong). Đáng chú ý, nghiên cứu này đã phát hiện hệ số cận huyết cao và kích thước quần thể hiệu quả thấp tại quần thể sông Serepok (Việt Nam), đưa ra giả thuyết rằng sự gia tăng của các đập thủy điện đã tạo nên sự phân tách của quần thể [15].

Đối với cá ét mọi (*L. chrysophekadion*), việc xác định cấu trúc quần thể và tập tính di cư dựa trên dữ liệu di truyền vẫn còn nhiều tranh luận. Mặc dù loài này thường được xếp vào nhóm cá xám (di cư chặng ngắn), nhưng nghiên cứu của Mashyaka và Duong (2021) sử dụng chỉ thị ISSR lại ghi nhận mức đa dạng di truyền cao với tỉ lệ loci đa hình đạt 81,7% – 85,9% và chỉ số Shannon từ 0,415 – 0,451 trên 4 quần thể tại Lào (Paksan) và Việt Nam (An Giang, Cần Thơ, Đồng Tháp). Đặc biệt, khoảng

cách di truyền giữa quần thể Lào và Việt Nam rất thấp (chỉ dao động từ 0,016 – 0,022), mặc dù khoảng cách địa lý giữa hai khu vực này là 1.200 km. Kết quả này gợi mở giả thuyết loài cá này có khả năng di cư đường dài tương tự nhóm cá trắng [70].

1.4.2.3. Di truyền quần thể nhóm cá di cư chặng dài

Đối với nhóm cá di cư chặng dài, các nghiên cứu về đa dạng di truyền và cấu trúc quần thể tập trung chủ yếu vào việc kiểm chứng vai trò của thác Khôn – rào cản tự nhiên lớn nhất trên dòng chính sông Mekong tại biên giới Lào và Campuchia.

Ở khu vực hạ lưu thác Khôn, hầu hết các nghiên cứu trên các loài cá da trơn thuộc họ cá tra (*Pangasiidae*) đều ghi nhận sự thiếu hụt cấu trúc quần thể rõ rệt. Cụ thể, So *et al.* (2006) sử dụng chỉ thị RFLP đã ghi nhận sự tương đồng di truyền giữa 06 quần thể cá tra *Pangasianodon hypophthalmus* tại Campuchia với giá trị F_{ST} thấp (0,022; $P = 0,139$) [10]. Kết quả tương tự cũng được tìm thấy khi mở rộng phạm vi khảo sát lên 10 quần thể tại Việt Nam và Campuchia bằng chỉ thị microsatellite [13]. Hiện tượng đồng nhất di truyền này cũng được báo cáo trên nhiều loài cá trắng có giá trị kinh tế khác thông qua nhiều loại chỉ thị phân tử khác nhau. Các nghiên cứu trên cá basa *P. bocourti* bằng chỉ thị RFLP [13], cá tra dầu *Pn. gigas* dựa trên sự kết hợp giữa microsatellite và mtDNA [11] đều cho thấy sự thiếu hụt cấu trúc quần thể rõ rệt giữa Thái Lan và Campuchia. Tương tự tại khu vực ĐBSCL, nghiên cứu trên cá bông lau *P. krempfi* [165] và cá sát sọc *Pangasius macronema* [166] sử dụng chỉ thị ISSR cũng khẳng định sự tương đồng về cấu trúc di truyền giữa các quần thể ở ĐBSCL.

Tuy nhiên, khi phạm vi nghiên cứu được mở rộng qua thác Khôn, cấu trúc quần thể bắt đầu có sự phân hóa rõ rệt. Ứng dụng kỹ thuật GBS để thu nhận bộ chỉ thị SNPs, Vu *et al.* (2020) ghi nhận sự phân tách quần thể cá tra tự nhiên thành hai nhóm riêng biệt: nhóm phía trên thác Khôn (Thái Lan) và nhóm phía dưới thác (Campuchia và Việt Nam) [18]. Sự khác biệt này cũng được ghi nhận trên các loài cá có kích thước nhỏ như cá linh (*Henicorhynchus* spp.). Đối với cá linh *H. siamensis*, các chỉ thị mtDNA đã xác nhận sự phân tách giữa quần thể thượng lưu và hạ lưu thác Khôn [9, 12], trong khi đó chỉ thị microsatellite không ghi nhận sự khác biệt này [14].

Tương tự, đối với cá linh *H. lobatus*, trong khi các chỉ thị truyền thống ghi nhận sự đồng nhất, thì kỹ thuật ddRAD-seq với độ phân giải cao đã phát hiện sự khác biệt đáng kể giữa quần thể Pakse (Lào) so với các quần thể tại Campuchia và Việt Nam [17]. Ngoài ra, nghiên cứu trên cá trèn răng *Belodontichthys truncatus* còn chỉ ra rằng bên cạnh dòng chính, các quần thể ở dòng nhánh như sông Chi (Thái Lan) cũng có sự khác biệt di truyền do tác động của trôi dạt di truyền [167].

Tóm lại, các dẫn liệu khoa học hiện nay cho thấy cấu trúc di truyền của cá sông Mekong bị chi phối phức tạp bởi sự tương tác giữa tập tính di cư và các rào cản địa lý. Các kết quả nghiên cứu thể hiện được mô hình phân hóa gen đặc trưng, phản ánh sự tương phản giữa tính đồng nhất di truyền trên toàn lưu vực của nhóm cá trắng với sự phân tách quần thể theo sinh cảnh của nhóm cá đen và cá xám. Việc ứng dụng hệ thống chỉ thị đa dạng, từ các chỉ thị truyền thống (DNA ti thể, RFLP, ISSR) đến các chỉ thị hiện đại (microsatellite, SNPs), đã cung cấp các dữ liệu quan trọng về mức độ đa dạng di truyền và cấu trúc quần thể cá tại lưu vực này.

Tuy nhiên, các nghiên cứu này vẫn còn hạn chế về độ phân giải của chỉ thị và tính hệ thống trong các đánh giá tổng thể. Về mặt kỹ thuật, các chỉ thị truyền thống thường có độ lặp lại thấp hoặc chỉ cung cấp thông tin về các vùng gen bảo thủ, dẫn đến việc không phát hiện được những sự phân hóa di truyền mới hình thành hoặc các cấu trúc quần thể ở những loài có dòng gen mạnh. Điều này dẫn đến tình trạng mâu thuẫn dữ liệu giữa các nghiên cứu khi sử dụng các hệ thống chỉ thị khác nhau trên cùng một đối tượng, gây khó khăn cho việc xác định ranh giới giữa các tập tính di cư. Về tính hệ thống, phần lớn các nghiên cứu hiện nay được thực hiện đơn lẻ trên từng loài hoặc từng khu vực địa lý nhất định, thiếu các nghiên cứu đối chiếu đồng thời giữa các loài và các nhóm sinh thái. Điều này nêu bật lên yêu cầu cần thiết về việc thực hiện một nghiên cứu được triển khai đồng thời trên 03 loài cá điển hình đại diện cho 03 nhóm tập tính di cư và sử dụng thống nhất chỉ thị SNPs phân bố trên toàn hệ gen. Cách tiếp cận này giúp đồng bộ dữ liệu, xác định cấu trúc di truyền và mức độ

kết nối quần thể của các nhóm cá trước những biến động môi trường và rào cản địa lý tại lưu vực sông Mekong.

1.4.2.4. Nghiên cứu về mô hình di cư

Nghiên cứu về mô hình di cư của các loài cá tại LMB đã được thực hiện thông qua các phương pháp khác nhau. Dựa trên khảo sát kiến thức sinh thái địa phương, các dữ liệu về thành phần loài, tính mùa vụ và tuyến đường di cư tại khu vực Lào, Campuchia [6, 72] và hệ thống thác Khôn [26, 67] đã được ghi nhận. Nhằm cung cấp những bằng chứng thực nghiệm, phương pháp gắn thẻ đã được triển khai để theo dõi quá trình di chuyển của từng cá thể. Điển hình, Hogan *et al.* (2006) ghi nhận cá vồ đém (*P. larnaudii*) có thể di cư từ sông Tonlé Sap, vượt qua Phnom Pênh và được đánh bắt trở lại tại Đồng Tháp (Việt Nam) với quãng đường hơn 100 km [128]. Những tiến bộ gần đây trong kỹ thuật giám sát hiện đại như hệ thống thẻ tích hợp thụ động và mạng lưới đo từ xa bằng sóng âm cung cấp những bằng chứng mới về hành vi di cư của cá. Robinson *et al.* (2024) xác định khả năng di cư ngược dòng của cá mè vinh *Hypsibarbus malcomi* (Cypriniformes: Cyprinidae) từ vùng hạ lưu lên thượng lưu đập thủy điện Xayaburi (Lào), đạt khoảng cách 354 km với tốc độ trung bình 0,8 – 1,01 km/ngày [168]. Đáng chú ý, Eschenroeder *et al.* (2025) phát hiện tuyến đường di cư mới của cá vồ đém (*P. larnaudii*) giữa sông Mekong và sông Sekong (Lào), đồng thời ghi nhận cá lăng nha (*Hemibagrus wyckioides*) có khả năng di cư chằng dãi trái ngược với các giả định về di cư chằng ngắn như trước đây [169].

Bên cạnh các phương pháp theo dõi vật lý, việc phân tích thành phần hóa học trong đá tai thông qua tỉ lệ các nguyên tố vi lượng đã trở thành công cụ hữu hiệu để phân định các chiến lược di cư ở cấp độ cá thể. Dựa trên tỉ lệ Sr:Ca và Ba:Ca, các nghiên cứu đã xác định tập tính di cư ngược dòng từ vùng nước mặn - lợ vào sâu trong nước ngọt để sinh sản (*anadromous*) của cá bông lau (*P. krempfi*) và cá tra bần (*P. mekongensis*) với quãng đường lần lượt là 1.400 km và 700 km [136, 170]. Ngược lại, các loài như cá basa (*P. bocourti*), cá hú (*P. conchophilus*) và cá vồ đém (*P. larnaudii*) được khẳng định thuộc nhóm di cư trong sông (*potamodromous*) dựa trên

nồng độ Sr ổn định trong suốt vòng đời [171, 172]. Bên cạnh đó, các nghiên cứu còn phát hiện sự đa dạng chiến lược di cư ngay trong cùng một loài. Cụ thể, có ít nhất ba chiến lược di cư khác nhau được ghi nhận ở các loài thuộc họ cá tra Pangasiidae và lên đến bốn chiến lược ở các loài khác (như cá chình hoa *Anguilla marmorata*, cá ngát *Plotosus canius*, cá phèn vàng *Polynemus melanochir*) [136, 172].

Mặc dù các phương pháp trên cung cấp dữ liệu chi tiết về tuyến đường di cư ở cấp độ cá thể, việc dự đoán mô hình di cư ở quy mô quần thể chỉ được thực hiện khi ứng dụng các công cụ di truyền. Đáng chú ý, Nguyen *et al.* (2020) khảo sát mô hình di cư giữa các quần đàn cá tra tự nhiên (*P. hypophthalmus*) ở Thái Lan, Campuchia và Việt Nam dựa trên dữ liệu di truyền (công cụ divMigrate). Kết quả đã phát hiện sự cô lập di truyền rõ rệt của quần đàn phía trên thác Khôn (Thái Lan) so với sự kết nối liên tục ở hạ lưu giữa Campuchia và Việt Nam, chứng minh thác Khôn không chỉ là rào cản địa lý mà còn là ranh giới di truyền quan trọng đối với các loài di cư tại khu vực này [18].

1.4.3. Các nghiên cứu về lắp ráp hệ gen ti thể và xác định mã vạch RAD của các loài cá thuộc giống *Labeo*

Giống *Labeo* (họ cá chép Cyprinidae) có 105 loài được ghi nhận, phân bố rộng rãi ở châu Phi, khu vực Nam Á và Đông Nam Á [173]. Tính đến thời điểm hiện tại (truy cập ngày 20/10/2025), có 54 trình tự hệ gen ti thể của 21 loài thuộc giống *Labeo* được đăng ký mã số trên Ngân hàng Gen. Các hệ gen ti thể này được lắp ráp từ dữ liệu giải trình tự Sanger và giải trình tự thế hệ mới, bao gồm cả đoạn đọc ngắn và đoạn đọc dài [174–176]. Đối với cá ét mọi (*L. chrysophekadion*), một trình tự hệ gen ti thể (mã số AP011199) được lắp ráp từ dữ liệu giải trình tự Sanger [177]. Nghiên cứu này tập trung khảo sát mối quan hệ phát sinh loài ở cấp độ tông Labeonini (Cypriniformes: Cyprinidae), do đó, các thông tin chi tiết của hệ gen ti thể cá ét mọi không được nhóm tác giả trình bày [177].

Đến nay, chưa có nghiên cứu nào sử dụng dữ liệu RAD-seq để lắp ráp hệ gen ti thể và xác định mã vạch RAD ở giống *Labeo* cũng như ở các loài cá sông Mekong, nhằm khảo sát đa dạng di truyền và cấu trúc quần thể.

1.5. Cơ sở khoa học của luận án

Cơ sở khoa học của luận án được xây dựng dựa trên sự kết hợp giữa sinh thái học cảnh quan thủy văn và hệ gen học quần thể so sánh (*Comparative population genomics*), nhằm phân tích mối quan hệ giữa đặc điểm sinh học của các loài cá và cấu trúc di truyền quần thể của chúng trong bối cảnh biến động môi trường tại sông Mekong.

Thứ nhất, luận án dựa trên nền tảng về tính liên tục của hệ thống thủy văn và vai trò của các hành lang di cư trong việc duy trì dòng gen. Lưu vực hạ lưu sông Mekong không chỉ là một điểm nóng về đa dạng sinh học với hơn 1.300 loài cá [178] mà còn là một hệ thống sinh thái động, nơi các chu kỳ nhịp lũ tự nhiên đóng vai trò là tín hiệu sinh học kích hoạt các quá trình di cư [7]. Trong điều kiện tự nhiên, dòng gen giữa các quần thể được duy trì thông qua các tuyến di cư dọc theo dòng chính và các tuyến di cư ngang kết nối với các vùng ngập lũ, bao gồm Biển Hồ và hệ thống sông 3S [6, 7, 31]. Tuy nhiên, sự gia tăng các rào cản nhân tạo và tác động của BĐKH có thể làm gián đoạn các hành lang này, từ đó dẫn đến sự phân hóa cấu trúc di truyền giữa các quần thể và suy giảm tính thích nghi của các loài bản địa [179, 180].

Thứ hai, luận án dựa trên lý thuyết địa sinh học so sánh và hệ gen học quần thể. Trong một hệ sinh thái đa loài, các cách tiếp cận quản lý hoặc phân tích đơn lẻ từng loài thường khó phản ánh đầy đủ các tác động mang tính hệ thống [31]. Cách tiếp cận này cho phép so sánh đồng thời các chỉ số đa dạng di truyền và cấu trúc quần thể của những loài có chiến lược sống khác nhau dưới cùng một áp lực môi trường [181]. Việc lựa chọn ba loài cá đại diện cho ba nhóm chính: cá chạch lá tre (*M. siamensis* – cá đen, không di cư), cá ét mọi (*L. chrysophekadion* – cá xám, di cư chặng ngắn và tùy nghi) và cá vô đém (*P. larnaudii* – cá trắng, di cư chặng dài) cho phép kiểm chứng giả thuyết về mối liên hệ giữa tập tính di cư và cấu trúc di truyền quần thể. Điều này

giúp làm rõ liệu các rào cản vật lý gây ra sự phân hóa đồng nhất hay có sự khác biệt tùy thuộc vào đặc điểm sinh học của từng loài.

Thứ ba, luận án dựa trên ưu thế về độ phân giải của dữ liệu giải trình tự thế hệ mới và kỹ thuật EzRAD. Đối với các loài cá Mekong chưa có hệ gen tham chiếu, việc sử dụng các chỉ thị truyền thống thường phản ánh chưa đầy đủ cấu trúc di truyền quần thể do kích thước mẫu hạn chế hoặc độ phân giải thấp [13, 17]. Kỹ thuật EzRAD cho phép thu nhận hàng ngàn chỉ thị SNPs phân bố ngẫu nhiên trên toàn hệ gen, giải quyết hạn chế về mặt thông tin di truyền [182]. Điều này không chỉ giúp đánh giá cấu trúc quần thể mà còn cho phép ước tính các thông số quan trọng như kích thước quần thể hiệu quả, hệ số cận huyết và các loci dưới tác động của chọn lọc – những thông tin mà dữ liệu DNA ti thể hay microsatellite chưa thể phản ánh toàn diện [103].

Cuối cùng, sự bổ trợ giữa hệ gen nhân và hệ gen ti thể thông qua việc lắp ráp *de novo* hệ gen rút gọn để xác định các chỉ thị SNPs kết hợp với chú giải hệ gen ti thể và mã vạch RAD tạo nên một hệ thống kiểm chứng chéo. Sự kết hợp này không chỉ tăng độ tin cậy cho các phân tích cấu trúc di truyền quần thể mà còn cung cấp bộ dữ liệu nền tảng phục vụ việc dự báo các kịch bản biến đổi di truyền, tạo cơ sở khoa học cho việc phân định các vùng bảo tồn và hành lang đa dạng sinh học tại sông Mekong.

CHƯƠNG 2:

ĐỐI TƯỢNG, PHẠM VI VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Sơ đồ khối mô tả các bước thực hiện chính theo ba nội dung nghiên cứu của luận án được thể hiện ở **Hình 2.1**.



Hình 2.1. Sơ đồ khối thực hiện ba nội dung nghiên cứu của luận án

2.1. Đối tượng, phạm vi nghiên cứu và phương pháp thu mẫu

Mẫu của 03 loài (cá chạch lá tre *M. siamensis*, cá ét mọi *L. chrysophekadion* và cá vô đém *P. larnaudii*) được thu thập ngẫu nhiên từ các ngư dân đánh bắt, chợ địa phương hoặc cảng cá tại lưu vực sông Mekong từ 2017 – 2021 (**Phụ lục 2**), gồm Myanmar (Thượng lưu Mekong - UMB) và Lào, Thái Lan, Campuchia, Việt Nam (Hạ lưu Mekong - LMB) (**Hình 2.2**).

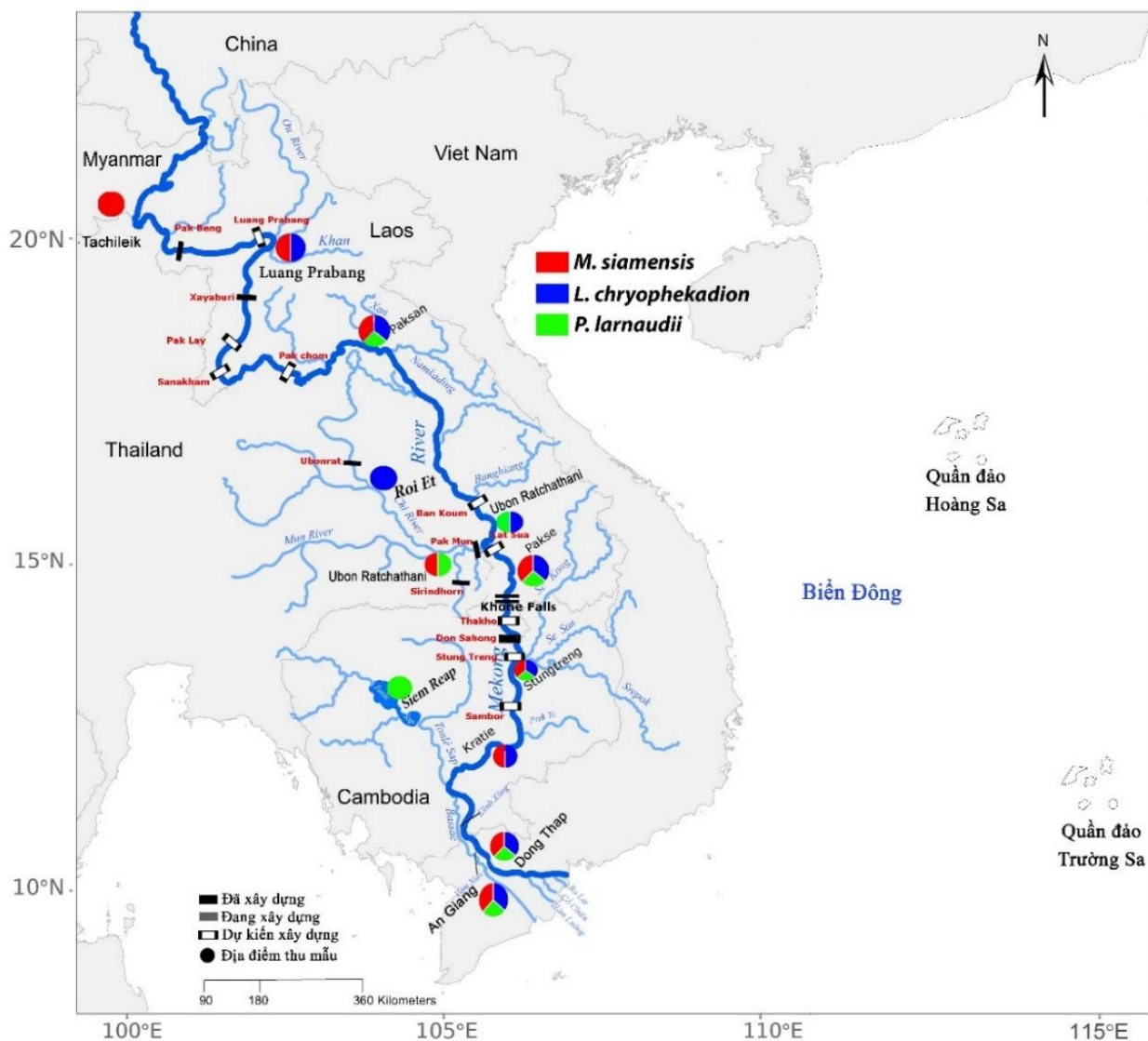
Chiến lược lựa chọn quần thể được thực hiện dựa trên bốn tiêu chí sau: *i) Đại diện quốc gia* thuộc lưu vực sông Mekong, lựa chọn 1–2 quần thể/ quốc gia; *ii) Đại diện thủy vực*, bao gồm các quần thể phân bố ở cả dòng chính và dòng nhánh, phản ánh sự khác biệt về chế độ dòng chảy, cấu trúc sinh cảnh và tính kết nối di cư của cá; *iii) Đại diện vùng địa lý – sinh thái*, với các quần thể trải dài trên năm vùng sinh thái đặc trưng của sông Mekong [27]: Thượng Lan Thương (khu vực có đặc trưng là sông núi); Hạ Lan Thương (sông suối đồi và đồng bằng ngập nước cục bộ); cao nguyên Korat (khu vực có mạng lưới sông nhánh phong phú như Mun, Chi); Kratie – Stung Treng (vùng hợp lưu giữa dòng chính Mekong và hệ thống 3S, nơi chuyển tiếp giữa vùng sông chảy xiết và đồng bằng ngập nước); và châu thổ Mekong (đồng bằng ngập nước rộng lớn và chế độ triều); và *iv) Ranh giới sinh thái – địa lý* quan trọng, chiến lược thu mẫu bao gồm các quần thể ở phía trên và phía dưới thác Khôn, được xem là ranh giới sinh học tự nhiên phân tách các khu hệ cá của vùng trung và hạ Mekong. Thông tin về vị trí địa lý và vùng sinh thái của các địa điểm thu mẫu được trình bày ở **Hình 2.2** và **Bảng 2.1**.

Nhằm giảm thiểu tối đa ảnh hưởng của sự biến động di truyền theo thời gian đến kết quả phân tích quần thể và dự đoán mô hình di cư, nghiên cứu áp dụng phương pháp thu thập ngẫu nhiên các mẫu cá thể trưởng thành, được đánh giá là có vốn gen ổn định [183]. Các mẫu cá thu thập được định loại ngay tại thực địa dựa vào các đặc điểm hình thái đặc trưng và các tài liệu của Rainboth (1996) [68] và Trần Đắc Định và cộng sự (2013) [52]. Cụ thể, cá chạch lá tre được nhận diện qua dãy 4-6 đốm tròn đen dọc theo vây lưng và đuôi; cá ét mọi có thân màu đen thẫm, có nhiều chấm đen và đỏ; và cá vồ đêm có 1 đốm đen lớn ở gốc vây ngực (các mô tả chi tiết được trình bày ở mục 1.2) (**Phụ lục 3**). Bên cạnh đó, để đảm bảo tính xác thực về nguồn gốc và địa điểm thu mẫu, ngư dân địa phương đã được phỏng vấn trực tiếp với sự hỗ trợ chuyên môn của các đối tác, gồm Uy Sophorn-Viện Nghiên cứu và Phát triển Nghề cá Nội đồng (Campuchia); Latsamy Phounvisouk-Trung tâm nghiên cứu nguồn lợi thủy sản (Lào);

Chaiwut Grudpan-Đại học Ubon Rachathani (Thái Lan) và Mie Mie Kyaw- Đại học Mandalay (Myanmar).

Tổng cộng 272 mẫu cá chạch lá tre (9 quần thể), 263 mẫu cá ét mọi (10 quần thể), 193 mẫu cá vồ đém (8 quần thể) được thu thập. Tại Tacheilek (Myanmar) thuộc lưu vực thượng lưu sông Mekong (UMB), 01 quần thể cá chạch lá tre được thu thập. Mặc dù cá ét mọi và cá vồ đém cũng được ghi nhận phân bố ở khu vực này, nhưng số lượng cá thể thu thập được (8 cá thể cá ét mọi và 1 cá thể cá vồ đém) không đủ để phân tích di truyền quần thể (**Bảng 2.1**).

Mẫu cơ của từng cá thể được thu thập và lưu giữ trong ống 2 mL có chứa cồn 95%, tiến hành mã số hóa và bảo quản ở -20°C để tiến hành các nghiên cứu về di truyền tại phòng thí nghiệm Sinh học phân tử, Trường Đại học Nha Trang.



Hình 2.2. Bản đồ các địa điểm thu mẫu các loài cá trong nghiên cứu hiện tại.
 (Màu sắc tương ứng với các loài cá; chữ màu đỏ thể hiện các đập thủy điện đã, đang và dự kiến xây dựng), và “=” là vị trí của thác Khôn (Khôn Falls).

Bảng 2.1. Thông tin vị trí địa lý, vùng sinh thái và số lượng cá thể của các loài cá thu thập tại các địa điểm nghiên cứu ở lưu vực sông Mekong

Quốc gia	Vị trí trên sông Mekong		Vùng sinh thái	Địa điểm thu mẫu	Kí hiệu quần thể	Tọa độ địa lý		Cá chạch lá tre	Cá ét mọi	Cá vồ dẽm
Myamnar	Thượng lưu - UMB	Dòng nhánh	Thượng Lan Thương	Tacheilek	TK	20°21'08.4"N	100°05'05.9"E	32	8*	1*
Lào	Thượng lưu - LMB	Dòng nhánh - Sông Khan	Hạ Lan Thương	Luang Prabang	LP	19°88'77.2"N	102°12'94.2"E	32	22	0
		Dòng chính		Paksan	PA	18°21'11.31"N	103°57'21.7"E	34	32	21
		Dòng chính		Pakse	PE	15°07'30.0"N	105°48'47.8"E	27	32	0
Thái Lan	Trung lưu - LMB	Dòng chính	Cao nguyên Korat	Ubon Ratchathani	UB-MK	15°25'70.8"N	104°76'15.7"E	0	28	29
		Dòng nhánh - Sông Mun		Ubon Ratchathani	UB-MR	12°15'54.22"N	109°6'15.5"E	32	0	24
		Dòng nhánh - Sông Chi		Roi Et	RE	15°23'51.47"N	105°21'19.6"E	0	30	0
Campuchia	Hạ lưu - LMB	Hợp lưu giữa dòng chính-3S	Kratie - Stung Treng	Stung Treng	ST	13°31'46.7"N	105°57'06.6"E	32	28	32
		Dòng chính		Kratie	KT	12°49'31"N	106°01'71.5"E	28	27	0
		Dòng nhánh		Siêm Riệp	SR	13°0'N	104°3'E	0	0	30
Việt Nam		Dòng chính	Châu thổ Mekong	Đồng Tháp	DT	10°57'04.1"N	105°38'11"E	25	32	32
		Dòng chính		An Giang	AG	11°07'37.7"N	105°10'50.8"E	30	24	24
	Tổng cộng							272	263	193

*Ghi chú: *số lượng mẫu không đủ quần thể nên bị loại khỏi phân tích*

2.2. Lắp ráp *de novo* hệ gen của các loài cá và xác định chỉ thị SNPs

2.2.1. Xây dựng thư viện DNA các loài cá sử dụng kỹ thuật EzRAD và giải trình tự

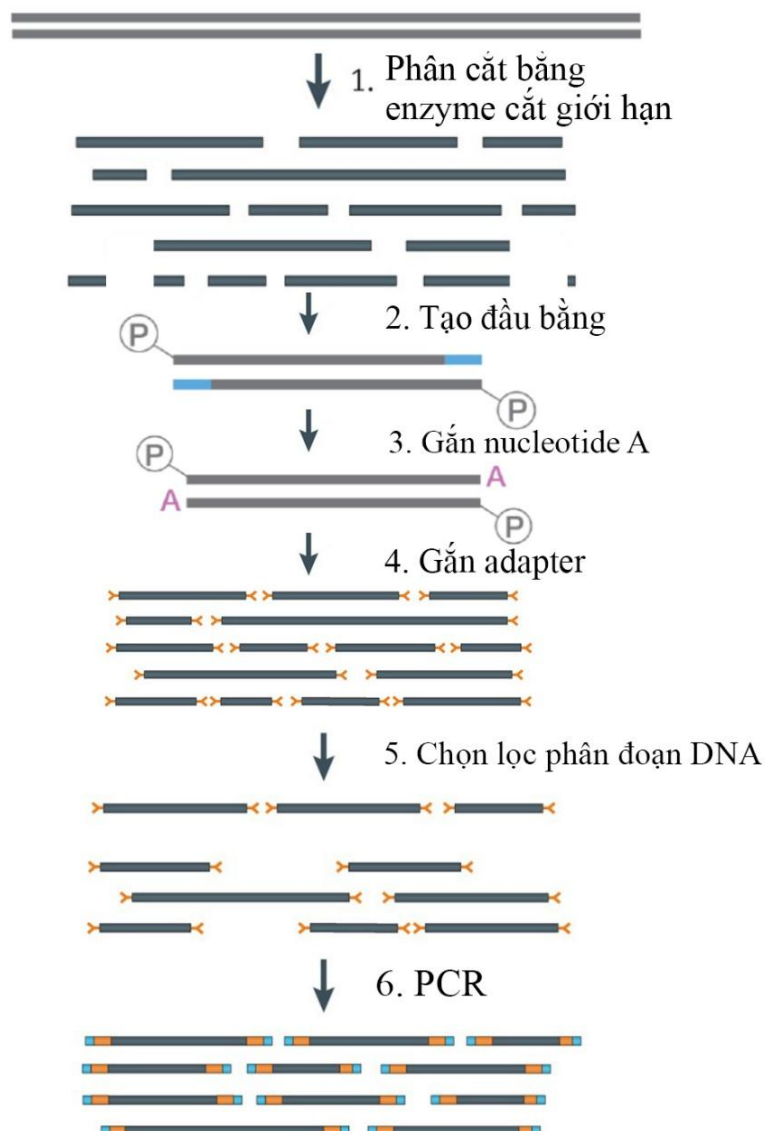
2.2.1.1. Tách chiết và xác định nồng độ DNA

DNA được tách chiết từ 30 mg mẫu cơ đã bảo quản trong cồn của mỗi cá thể cá bằng bộ kit Wizard® Genomic DNA Purification (Promega, Mỹ) theo hướng dẫn của nhà sản xuất, có điều chỉnh. Cụ thể, trong bước thu hồi DNA, nghiên cứu tiến hành rửa giải ba lần riêng biệt với 100 μ L đệm AE thay vì sử dụng 250 μ L Nuclease-Free Water, mỗi lần ủ ở các khoảng thời gian khác nhau (5 phút, 10 phút và 15 phút, chi tiết ở **Phụ lục 4**). DNA tách chiết được điện di trên gel agarose 1,5% để kiểm tra độ tinh sạch. Những mẫu DNA tốt (băng sáng rõ và không bị đứt gãy) được xác định nồng độ sử dụng bộ kit Qubit dsDNA High Sensitivity (Thermo Fisher Scientific) và thực hiện trên máy đo huỳnh quang Qubit® 2.0 Fluorometer (Invitrogen, Mỹ) theo hướng dẫn của nhà sản xuất (**Phụ lục 4**). Mẫu DNA đạt nồng độ tối thiểu 3 ng/ μ L được chọn để tiếp tục cho các bước thí nghiệm tiếp theo.

DNA được tiến hành tinh sạch bằng AMPure XP Beads (Beckman Coulter) theo tỉ lệ 1:2. Mẫu được lắc trong 15 phút ở 200 vòng/phút trên máy lắc GFL (mã máy 3005, Đức) để DNA gắn kết với hạt bead. Sau khi đặt trên đế từ trong 5 phút, DNA được giữ lại trên hạt bead, cho phép loại bỏ dung dịch chứa tạp chất. Cuối cùng, rửa hạt bead bằng cồn 80% để thu được DNA tinh sạch.

2.2.1.2. Xây dựng thư viện DNA hệ gen

Thư viện DNA hệ gen của các loài cá được tạo ra thông qua kỹ thuật EzRAD sử dụng bộ kit TruSeq Nano HT Library Preparation (Illumina). Quy trình này dựa trên phương pháp được mô tả bởi Toonen *et al.* (2013) [82], với những điều chỉnh theo Dang *et al.* (2019) [16]. Các bước chính của quy trình được minh họa trong **Hình 2.3**.



Hình 2.3. Quy trình tạo thư viện hệ gen bằng kỹ thuật EzRAD ([82], có hiệu chỉnh)

a. Phân cắt DNA hệ gen bằng enzyme hạn chế

Mẫu DNA tinh sạch được phân mảnh bằng cách sử dụng đồng thời hai enzyme hạn chế, MboI và Sau3AI (NEB). Cả hai enzyme đều nhận biết và cắt tại trình tự 5'-GATC-3' và tạo ra đầu dính. Tuy nhiên, MboI bị ức chế bởi sự methyl hóa Adenine (nucleotide A), trong khi Sau3AI vẫn hoạt động hiệu quả, do đó, việc kết hợp cả hai enzyme giúp đảm bảo việc cắt thành công trên các mẫu DNA có mức độ methyl hóa không đồng nhất, từ đó giảm thiểu thiên vị và tăng tính đại diện của thư viện DNA. Phản ứng này được thực hiện trong tổng thể tích 25 μ L, bao gồm đệm 2,5 μ L đệm

SmartCut 10X (NEB), 0,5 μ L MboI (5 U/mL), 0,5 μ L Sau3AI (5 U/mL) và 21,5 μ L mẫu DNA tinh sạch. Phản ứng được tiến hành trên máy luân nhiệt C1000 Touch (Biorad) ở 37°C trong khoảng từ 3 đến 24 giờ. Sau đó, nhiệt độ được tăng lên 65°C và duy trì trong 20 phút nhằm bất hoạt enzyme và cuối cùng bảo quản mẫu ở 4°C.

b. Tạo đầu bằng (Blunt End)

Các đoạn DNA có đầu dính, sản phẩm của quá trình phân cắt bằng enzyme hạn chế, được biến đổi thành đầu bằng bằng cách bổ sung 13,3 μ L End Repair Mix (Illumina) và thực hiện phản ứng tại 30°C trong 30 phút. Quá trình này dựa trên cơ chế hoạt động của DNA Polymerase, enzyme tổng hợp DNA theo chiều 5' \rightarrow 3'.

c. Chọn lọc phân đoạn DNA có kích thước mục tiêu (Size selection)

Để chọn lọc các đoạn DNA có kích thước từ 350 đến 550 bp, quy trình hai bước được thực hiện sử dụng dung dịch Sample Purification Beads (SP Beads, Illumina) với các tỉ lệ khác nhau. Các hạt bead trong SP Beads được thiết kế để ưu tiên liên kết với các đoạn DNA lớn hơn, sau đó là các đoạn DNA nhỏ hơn theo thứ tự giảm dần kích thước.

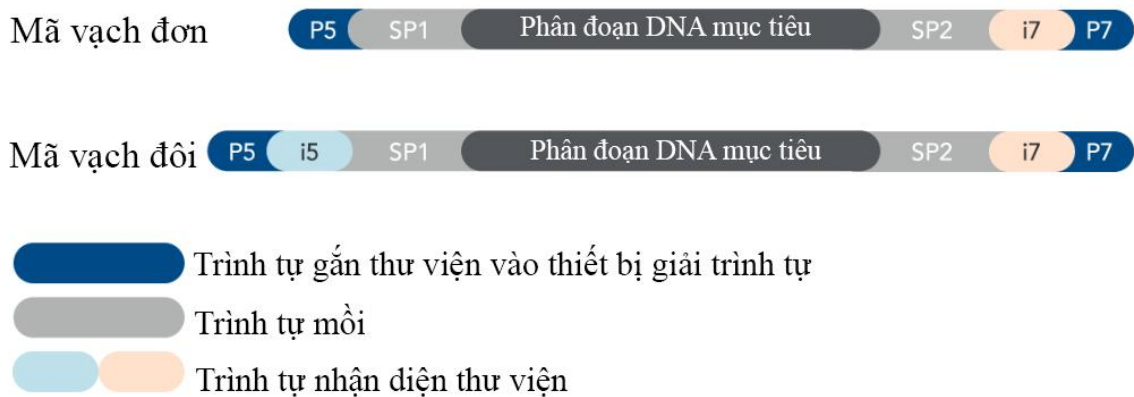
- Bước 1: dung dịch SP Beads pha loãng (tỉ lệ 1 H₂O:1,5 SP Beads) được thêm vào, cho phép các đoạn DNA lớn hơn 550 bp gắn kết với hạt bead. Các hạt bead này sau đó được loại bỏ, giữ lại dung dịch chứa các đoạn DNA mục tiêu.

- Bước 2: thêm SP Beads không pha loãng vào dung dịch ở bước 1. Các đoạn DNA từ 350 đến 550 bp sẽ liên kết với hạt bead, trong khi các đoạn nhỏ hơn bị loại bỏ trong quá trình rửa với cồn 80%. Cuối cùng, các đoạn DNA mục tiêu được thu hồi bằng dung dịch Resuspension buffer (Illumina).

d. Gắn nucleotide A (A-tailing) vào đầu 3' và Adapter

Để ngăn chặn các đoạn DNA tự liên kết và chuẩn bị cho việc gắn adapter, các nucleotide A (dung dịch A-tailing, Illumina) được thêm vào đầu 3' của chúng. Phản ứng gắn nucleotide A được thực hiện theo quy trình nhiệt như sau: 37°C trong 30 phút, 70°C trong 5 phút, 4°C trong 5 phút và mẫu được giữ ở 4°C.

Sau đó, adapter với nucleotide T bổ sung ở đầu 3' được gắn vào các đoạn DNA mục tiêu. Cấu trúc của Adapter gồm (1) trình tự gắn DNA lên bề mặt hệ thống máy giải trình tự (*Flow cell binding sites*); (2) trình tự môi giải trình tự (*Sequencing primer binding sites*); (3) trình tự nhận diện mẫu DNA (*Index sequences*), có thể là 1 (*single-index, mã vạch đơn*) hoặc 2 (*dual-index, mã vạch đôi*); và (4) Trình tự môi đọc chỉ số (*Index primer binding sites*) (**Hình 2.4**). Phản ứng gắn adapter được thực hiện ở 30°C trong 10 phút. Ngay khi kết thúc phản ứng, nhiệt độ mẫu đạt 4°C, dung dịch Stop Ligation (Illumina) được bổ sung vào phản ứng để dừng quá trình gắn adapter.



Hình 2.4. Cấu trúc của adapter

Sau quá trình gắn adapter, các đoạn DNA được tiếp tục làm sạch bằng dung dịch SP Bead (tỉ lệ 1:1). Tiếp theo rửa hạt bead bằng cồn 80% và thu hồi DNA bằng dung dịch Resuspension.

e. Khuếch đại thư viện DNA bằng PCR

Thư viện DNA được khuếch đại với tổng thể tích 15 μL , bao gồm 1,5 μL Illuminia PCR Primer Cocktail, 6 μL Illuminia Enhanced PCR Mix, 1,875 μL nước tinh khiết và 5,625 μL thư viện DNA. Phản ứng được thực hiện trên máy luân nhiệt với chu trình nhiệt độ: biến tính ban đầu tại 95°C trong 3 phút, tiếp theo là 8 chu kỳ của 98°C trong 20 giây, 60°C trong 15 giây và 72°C trong 30 giây và bước kéo dài cuối cùng ở 72°C trong 5 phút, giữ mẫu ở 4°C.

Để kiểm tra kết quả khuếch đại, các sản phẩm PCR được điện di trên gel agarose 1,5%. Gel được nhuộm Ethidium Bromide và quan sát dưới tia UV. Kỹ thuật EzRAD

tạo ra thư viện DNA với các phân đoạn DNA có kích thước từ 350 đến 550 bp và thư viện này được bảo quản ở -40°C.

2.2.1.3. Giải trình tự thế hệ mới

Thư viện DNA được tinh sạch bằng SP Beads theo tỉ lệ 1:1 và xác định nồng độ DNA bằng bộ đo huỳnh quang Qubit 2.0 (Introvigen) theo hướng dẫn của nhà sản xuất. Các thư viện DNA được giải trình tự hai chiều (pair-end) tại phòng Thí nghiệm Di truyền của Trường Đại học Texas A&M Corpus Christi (Hoa Kỳ) sử dụng hệ thống giải trình tự thế hệ mới HiSeq 4000 (Illumina).

Tất cả dữ liệu trong nghiên cứu được tiến hành đăng ký mã số trên Ngân hàng Gen thông qua cổng trực tuyến (Submission Portal), bao gồm tập hợp thông tin về mẫu (BioProject - <https://submit.ncbi.nlm.nih.gov/subs/bioproject/> và BioSample - <https://submit.ncbi.nlm.nih.gov/subs/biosample/>) và trình tự đoạn đọc thô (Sequence Read Archive, <https://submit.ncbi.nlm.nih.gov/subs/sra/>).

2.2.2. Lắp ráp de novo hệ gen của các loài cá và xác định chỉ thị SNPs

2.2.2.1. Đánh giá chất lượng trình tự đoạn đọc

Dữ liệu giải trình tự thư viện DNA của ba loài cá trong nghiên cứu là các đoạn đọc thô (định dạng file FASTQ) được minh họa ở **Hình 2.5**. Quy trình xử lý dữ liệu (**Hình 2.6**) được thực hiện trên hệ điều hành Ubuntu Server 21.10, các phần mềm và công cụ tin sinh học được cài đặt trên hệ thống máy chủ dung lượng cao có cấu hình như sau: 24 CPU E5-2620 v2 x 2 core x Intel(R) Xeon(R) 2.10 GHz, và dung lượng RAM là 78.6 GB.

```
@HWI-D00372:342:C7JLBANXX:4:1101:4553:2385 1:N:0:ATTCAGAATATAGCCT
GATCAACTGTGTGGCCCATCTGTCATCTAACAAGGTAACAATGTAACTCTCCACTCACAGTATC
+
/BBBBFFFFFFFFFFFF<FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF/BF
```

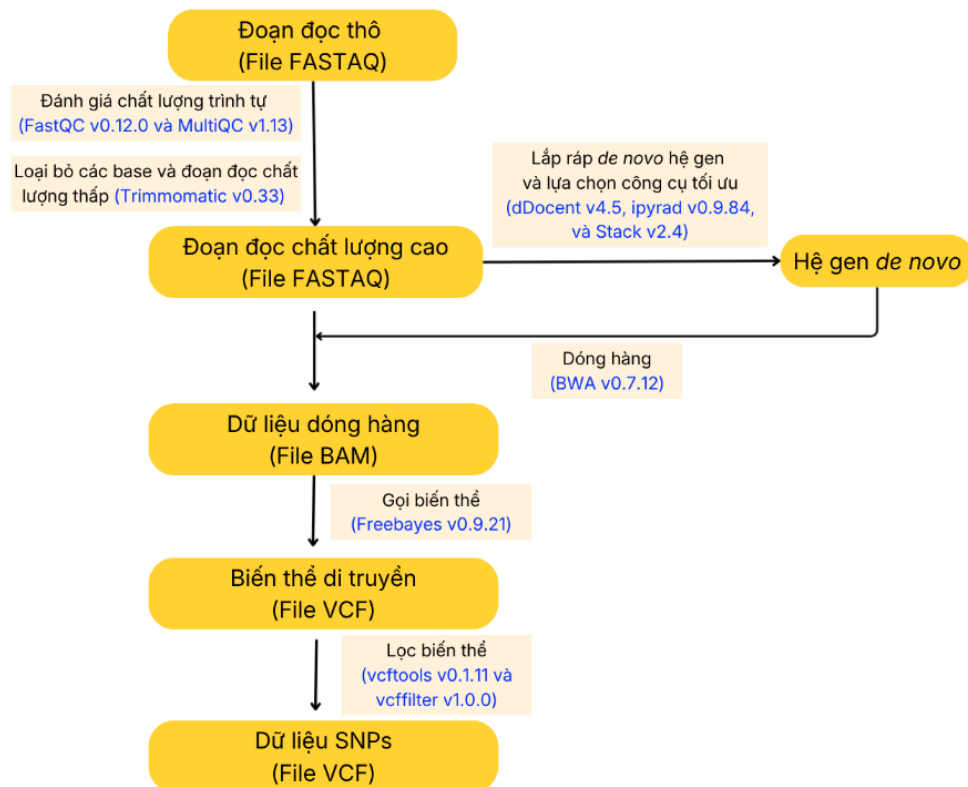
Tên đoạn đọc

Trình tự

Chất lượng nucleotide theo
thang Q score (Bộ mã ASCII)

Base: T
Q score: F = 37

Hình 2.5. Hình ảnh minh họa file trình tự với định dạng FASTQ



Hình 2.6. Sơ đồ khối quy trình xử lý số liệu

Các trình tự được đánh giá chất lượng thông qua các thông số gồm điểm chất lượng trên số lượng trình tự (*Per sequence quality scores*), điểm chất lượng trên từng vị trí nucleotide (*Per base sequence quality*), tỉ lệ trình tự chứa base N (*Per base N content*) và tỉ lệ trình tự chứa Adapter (*Adapter content*) sử dụng công cụ FastQC v0.12.0 [184] và MultiQC v1.13 [185]. Sau đó, nhằm thu nhận các đoạn đọc chất lượng cao, nghiên cứu sử dụng phần mềm Trimmomatic v0.33 [186] để loại bỏ các base và đoạn đọc chất lượng thấp như sau: (i) đối với base, các đoạn adapter và các base có điểm chất lượng Phred nhỏ hơn 30 (tương ứng xác suất nucleotide bị đọc sai bởi thiết bị giải trình tự là 1/1000, nghĩa là độ chính xác 99,9%; **Phụ lục 5**); (ii) đối với trình tự đoạn đọc: các trình tự có tỉ lệ N (nucleotide không xác định) trên 10% và chiều dài nhỏ hơn 50 bp. Sau đó, các đoạn đọc chất lượng cao được kết nối dựa trên đoạn chồng lấp (*overlap*) tối thiểu 40bp để tạo thành các contigs sử dụng công cụ zcat

v1.10 (<https://linux.die.net/man/1/zcat>). Các đoạn contigs có chiều dài tối thiểu 147bp được sử dụng làm dữ liệu đầu vào cho quá trình lắp ráp *de novo* hệ gen.

2.2.2.2. Khảo sát công cụ tối ưu lắp ráp *de novo* hệ gen sử dụng dữ liệu trình tự cá vồ đém

Nghiên cứu sử dụng tập hợp các contigs chất lượng cao của cá vồ đém làm dữ liệu đầu vào, sau đó tiến hành khảo sát và đánh giá ba bộ công cụ lắp ráp *de novo* hệ gen mã nguồn mở phổ biến: gồm dDocent v4.5 [187], ipyrad v0.9.84 [188] – đều sử dụng thuật toán tham lam (*greedy-base*) và Stack v2.4 [91] – sử dụng thuật toán đồ thị (*graph-base*). Quy trình lắp ráp *de novo* hệ gen cho từng công cụ được tiến hành theo hướng dẫn cụ thể từ các nghiên cứu trước (Bird, 2022 [189] đối với dDocent 4.5; Rochette *et al.*, 2019 [190] đối với Stacks 2.4; và Eaton and Overcast, 2020 [188] đối với ipyrad 0.9.84) với các thông số mặc định (**Phụ lục 6**). Tuy nhiên, để đảm bảo tính khách quan trong việc so sánh và đánh giá chính xác hiệu suất giữa các công cụ, nghiên cứu đã tiến hành cài đặt hai thông số liên quan đến độ bao phủ và số lượng contigs đại diện cho mỗi cá thể. Hai thông số này được cài đặt giống nhau ở cả ba công cụ (chi tiết tại **Bảng 2.2**) nhằm loại bỏ sự khác biệt do cài đặt tham số và tập trung so sánh kết quả lắp ráp *de novo* hệ gen.

Bảng 2.2. Các thông số được cài đặt để lắp ráp *de novo* hệ gen

Công cụ	dDocent v4.2		Stack v2.41		Ipyrad v0.93	
Thông số	Cut-off 1	Cut-off 2	stack depth (-m)	mismatch (-M)	mindepth	min_samples_locus
Giá trị	3	9	3	9	3	9
Giải thích thông số	độ bao phủ tối ưu (được tính cho mỗi cá thể)	số lượng contigs của mỗi cá thể	đoạn đọc tối thiểu được yêu cầu cho mỗi alen của từng cá thể	khoảng cách tối đa của 2 nucleotides giữa các alen trong một locus	độ bao phủ tối thiểu của từng base	số lượng cá thể tối thiểu trên mỗi locus

Để đánh giá kết quả lắp ráp *de novo* hệ gen, nghiên cứu sử dụng các thông số lắp ráp gồm số lượng các đoạn contig lắp ráp, chỉ số N50 (là kích thước của các contig đã được sắp xếp từ lớn đến nhỏ tại vị trí kích thước đạt 50% hệ gen, **Phụ lục 5**), kích thước hệ gen, số lượng, kích thước lớn nhất và kích thước trung bình của cụm (*cluster*). Các thông số này được xác định bằng công cụ awk v5.1.0 (<https://linux.die.net/man/1/awk>) và grep v3.4 (<https://www.gnu.org/software/grep/>). Đồng thời, tỉ lệ các đoạn đọc được đóng hàng chính xác vào hệ gen được xác định bằng tính năng ‘samtools-flagstat’ trong Samtools v1.9 [191]. Sau đó, công cụ lắp ráp *de novo* hệ gen tối ưu được lựa chọn và tiếp tục áp dụng để lắp ráp *de novo* hệ gen cho cá chạch lá tre và cá ét mọi.

2.2.2.3. Lắp ráp *de novo* hệ gen của cá chạch lá tre và cá ét mọi

Dựa vào quy trình tối ưu của loài cá vô đốm, hệ gen của cá chạch lá tre và cá ét mọi được tiến hành lắp ráp *de novo*. Sử dụng các đoạn contigs đã được kết nối của cá chạch lá tre và cá ét mọi (mục 2.2.2.1), nghiên cứu tiến hành xác định giá trị của thông số cut-off 1 và cut-off 2 theo quy trình dDocent nhằm thu nhận các contigs đặc trưng và tần suất xuất hiện trong toàn bộ dữ liệu. Các thông số về hệ gen gồm số lượng các đoạn contigs, kích thước hệ gen, thành phần các nucleotide (A, T, G, C và N) và tỉ lệ GC được xác định bằng công cụ awk v5.1.0 và grep v3.4. Để đánh giá chất lượng hệ gen sau quá trình lắp ráp *de novo*, nghiên cứu tiến hành xác định chỉ số N50 và tỉ lệ các đoạn đọc được đóng hàng chính xác vào hệ gen bằng công cụ awk v5.1.0 và grep v3.4.

2.2.3. Xác định chỉ thị SNPs của các loài cá

Để xác định chỉ thị SNPs, các đoạn đọc chất lượng cao từ mỗi cá thể cá được đóng hàng với hệ gen *de novo* sử dụng thuật toán MEM (*Maximum Entropy Method*) của công cụ BWA v0.7.12 (*Burrows-Wheeler Aligner*) [192], tạo ra các tệp dữ liệu đóng hàng ở định dạng BAM. Dựa trên các tệp đóng hàng, các biến thể di truyền (như SNPs, vị trí chèn/xóa - INDELs, nhiều cặp nucleotide nằm liền kề nhau - Multiple Nucleotide Polymorphisms (MNPs) và các biến thể phức hợp khác) được gọi đồng thời bằng công cụ Freebayes [193] và tập hợp thành dữ liệu biến thể thô (bao gồm cả SNP/INDEL thô).

Để đảm bảo thu được tập hợp các chỉ thị SNP có độ tin cậy cho phân tích tiếp theo, các biến thể di truyền thô được tiến hành sàng lọc nhằm loại bỏ các biến thể không mong muốn (như INDELs) và các vị trí có độ tin cậy thấp sử dụng công cụ vcftools v0.1.11 [194] và vcfilter v1.0.0 [193]. Cụ thể là, nghiên cứu chỉ thu nhận các SNPs có 2 alen và loại bỏ các vị trí chèn/xóa. Về chất lượng, các SNPs có chất lượng trung bình từ 30 trở lên (tương đương với độ chính xác gọi biến thể là 99,9%), độ sâu bao phủ trung bình tại mỗi vị trí đảm bảo từ 10 trở lên và chỉ giữ lại các SNP dị hợp tử có cân bằng alen nằm trong khoảng 0,25-0,75 nhằm loại bỏ các biến thể có độ thiên lệch giải trình tự cao. Về tần suất xuất hiện, các biến thể hiếm được loại bỏ với ngưỡng tần suất alen thay thế nhỏ hơn 0,05 và số alen thay thế nhỏ hơn 3 trong toàn bộ quần thể. Về dữ liệu thiếu, loại bỏ các SNPs có tỉ lệ thiếu trên 80% cá thể và các cá thể có tỉ lệ phát hiện SNPs nhỏ hơn 80%. Cuối cùng, loại bỏ các SNPs không tuân thủ cân bằng Hardy-Weinberg với $p < 0,001$, đồng thời chỉ giữ lại 1 SNP trên mỗi contig để đảm bảo tính độc lập của các chỉ thị trong phân tích liên kết (chi tiết tại **Bảng 2.3**).

Sau khi sàng lọc chất lượng, tập hợp dữ liệu SNPs có độ tin cậy cao được đưa vào phân tích để phân biệt các SNPs trung tính (*neutral loci*) với các SNPs ngoại vi (*outlier loci*) – những vị trí chịu tác động của chọn lọc tự nhiên. Việc phát hiện các loci ngoại vi chịu nhiều yếu tố gây nhiễu như tỉ lệ dương tính giả cao, lỗi kiểu gen, cấu trúc quần thể phức tạp và độ nhạy khác nhau của các thuật toán [195]. Do đó, để tăng tính tin cậy và kiểm tra chéo kết quả, nghiên cứu này sử dụng đồng thời hai phương pháp được đề xuất bởi Narum và Hess (2011) [195]: Lositan Selection Workbench v1.0 (sử dụng thuật toán FDIST2) [196] và BayeScan v2.1 (sử dụng thuật toán hồi quy Bayes) [197]. Sự kết hợp hai phương pháp nhằm khắc phục các hạn chế riêng lẻ của từng thuật toán, từ đó cung cấp các locus chịu chọn lọc đáng tin cậy hơn. Dữ liệu đầu vào được chuẩn bị ở định dạng genpop (“.gen”).

Bảng 2.3. Các thông số và giá trị lọc được sử dụng trong nghiên cứu hiện tại

STT	Các thông số lọc	Giá trị sàng lọc	Ý nghĩa
1	Bi-allelic		Chỉ giữ lại các SNPs có 2 alen
2	Loại bỏ các vị trí chèn/xóa		Loại các vị trí SNPs chèn và xóa
3	Chất lượng trung bình (<i>mean quality score - Q</i>)	≥ 30	Loại những SNPs có điểm chất lượng Phred nhỏ hơn 30, tương ứng với độ chính xác là 99,9%
4	Độ sâu bao phủ trung bình tối thiểu (<i>minimum mean DP</i>)	≤ 10	Độ bao phủ tối thiểu của các đoạn đọc trên mỗi locus của tất cả các cá thể
5	Tỉ lệ phát hiện SNPs trên tất cả các cá thể	80%	Dữ liệu thiếu (<i>missing data</i>): những vị trí SNPs không xác định được kiểu gen của tất cả các cá thể trong quần thể nghiên cứu. Loại bỏ các SNPs có tỉ lệ “ <i>missing data</i> ” $\geq 80\%$, thường loại bỏ số lượng lớn SNPs
6	Cân bằng alen (<i>allele balance - AB</i>)	$< 0,25$ và $> 0,75$	Tỉ lệ giữa số lần đọc của hai alen tại một SNP, và chỉ áp dụng cho SNP dị hợp tử.
7	Số alen thay thế (<i>minor allele count - MAC</i>)	< 3	Số lần xuất hiện của một alen khác biệt với alen tham chiếu (<i>reference allele</i>) trong một tập hợp dữ liệu di truyền
8	Tần số alen thay thế (<i>minor Allele Frequency - mAF</i>)	0,05	Tỉ lệ xuất hiện của alen thay thế (<i>alternative allele</i>) so với tổng số tất cả các alen tại một vị trí cụ thể trong quần thể
9	Tỉ lệ phát hiện SNPs trên từng cá thể	$> 80\%$	Cá thể thiếu (<i>Missing individuals</i>) xảy ra khi một số cá thể trong quần thể nghiên cứu không có đủ dữ liệu SNPs để phân tích (Loại các cá thể có tỉ lệ SNPs $\leq 80\%$)
10	Cân bằng Hardy-Weinberg	$p < 0,001$	Lọc các SNPs mà giá trị p nhỏ hơn 0,1% trong quần thể
11	1 SNP/contig		Chỉ giữ lại một SNP trên mỗi contig

Đối với công cụ Lositan, mỗi locus được tính toán giá trị khác biệt di truyền (F_{ST}) và dị hợp tử mong đợi (H_e), sau đó được hiển thị trên biểu đồ phân tán F_{ST} so với H_e . Phương pháp này giả định rằng hầu hết các locus trong hệ gen tiến hóa trung tính; do đó, các locus có giá trị F_{ST} hoặc H_e nằm ngoài vùng tin cậy của phân phối kỳ vọng được xem là các locus chịu chọn lọc. Phân tích được thực hiện với các thông số gồm 100.000 lần lặp mô phỏng, khoảng tin cậy 95% và giá trị kiểm soát tỉ lệ phát hiện sai (*False Discovery Rate* – FDR) < 0,001. Các locus ngoại vi được xác định là những locus có giá trị F_{ST} cao hơn vùng tin cậy (chọn lọc tích cực – *positive selection*) hoặc có giá trị H_e cao hơn vùng tin cậy (chọn lọc cân bằng – *balancing selection*).

Trong khi đó, công cụ BayeScan v2.1 phát hiện các locus ngoại vi dựa trên sự khác biệt về tần số alen giữa các quần thể. Phân tích được thực hiện với các thông số mặc định, bao gồm 20 lần chạy thử, mỗi lần 5.000 vòng lặp, tiếp theo là 50.000 vòng lặp “burn-in”. Giá trị tỉ lệ cược tiên nghiệm cho mô hình trung tính được đặt là 10. Đối với mỗi locus, giá trị “log10(PO)” được tính toán, locus được xem là ngoại vi khi $\log_{10}(PO) > 0$ và $FDR < 0,001$.

Để đảm bảo tính chính xác khi ước tính các thông số đa dạng di truyền quần thể (chủ yếu dựa trên các quá trình tiến hóa trung tính), các loci ngoại vi được xác định bởi cả hai phương pháp (LOSITAN và BayeScan) được loại bỏ khỏi tập dữ liệu để thu được các SNPs trung tính. Đồng thời, tập hợp các loci trùng lặp được chia sẻ giữa BayeScan và LOSITAN cũng được thu nhận làm tập dữ liệu thứ hai phục vụ cho các phân tích chuyên sâu về thích nghi.

Cuối cùng, hai tập dữ liệu SNPs trung tính và loci đáp ứng (định dạng “.vcf”) được sử dụng để khảo sát đa dạng di truyền và cấu trúc quần thể của các loài cá ở hạ lưu sông Mekong. Phần mềm PGDSpider v2.1.1.3 [198] được sử dụng để chuyển đổi định dạng từ “.vcf” sang “.gen”, “.arp” và “.strc”, nhằm tạo ra dữ liệu đầu vào phù hợp cho các phân tích đa dạng di truyền và cấu trúc quần thể.

2.3. Khảo sát sự đa dạng di truyền, cấu trúc quần thể và dự đoán mô hình di cư của ba loài cá điển hình ở lưu vực hạ lưu sông Mekong

2.3.1. Khảo sát sự đa dạng di truyền của các quần thể cá

Sự đa dạng di truyền của các quần thể cá được đánh giá sử dụng tập dữ liệu SNPs trung tính ở định dạng file genpop (.gen). Các thông số di truyền được tính toán gồm dị hợp tử quan sát (*observed heterozygosity* - H_o) và dị hợp tử mong đợi (*expected heterozygosity* - H_e). Trong đó, giá trị H_o phản ánh tần số thực tế của các cá thể dị hợp tử trong quần thể, trong khi H_e được ước tính dựa trên giả định cân bằng Hardy-Weinberg (HWE), thể hiện mức độ đa dạng di truyền tiềm năng của quần thể. Cả hai thông số này đều dao động trong khoảng 0 đến 1, tuy nhiên, do SNP là chỉ thị có hai alen, giá trị H_e tối đa về mặt lý thuyết chỉ đạt 0,5 khi hai alen có tần số bằng nhau [199]. Các giá trị H_o và H_e được thực hiện tính riêng cho từng quần thể thu mẫu và cho toàn bộ tập hợp cá thể thuộc lưu vực sông Mekong sử dụng phần mềm GenAlEx v6.5.2 [108].

2.3.2. Xây dựng cấu trúc di truyền quần thể các loài cá ở hạ lưu sông Mekong

2.3.2.1. Xác định sự khác biệt di truyền của các quần thể cá

Sử dụng bộ dữ liệu SNPs trung tính và loci đáp ứng của từng loài cá, sự khác biệt di truyền chung (G_{ST}) [200], khác biệt di truyền theo cặp (F_{ST}) [201] giữa các quần thể và giá trị độ tin cậy được tính toán bằng phần mềm Alrequin v3.5 [202] với dữ liệu đầu vào ở định dạng file “.arp” và tần số hoán vị được cài đặt là 9.999. Các giá trị G_{ST} và F_{ST} đều nằm trong khoảng từ 0 (không có sự khác biệt di truyền) đến 1 (khác biệt di truyền hoàn toàn). Đối với F_{ST} , giá trị $< 0,05$ được xem là mức sai khác di truyền nhỏ; $0,05 < F_{ST} < 0,15$ biểu thị sai khác di truyền trung bình và $> 0,15$ phản ánh sai khác di truyền rõ rệt giữa các quần thể [201].

Để kiểm soát tỉ lệ phát hiện sai (FDR), các giá trị p thu được từ phân tích F_{ST} được hiệu chỉnh đa kiểm định bằng phương pháp Benjamini và Hochberg (1995) [203] (công thức tính p hiệu chỉnh được thể hiện ở **Phụ lục 5**). Sự khác biệt di truyền

theo cặp (F_{ST}) chỉ được xem là có ý nghĩa thống kê khi giá trị p nhỏ hơn p hiệu chỉnh ($p_{adjusted}$). Quá trình hiệu chỉnh giá trị độ tin cậy này được thực hiện bằng công cụ trực tuyến theo đường link <https://tools.carbocation.com/>. Cuối cùng, biểu đồ heatmap được xây dựng bằng công cụ seaborn v0.13.2 [204] để trực quan hóa mức độ khác biệt theo cặp giữa các quần thể (F_{ST}).

2.3.2.2. Xây dựng cấu trúc di truyền quần thể của các loài cá

Nghiên cứu cấu trúc di truyền quần thể của các loài cá được tiến hành bằng cách phân tích dữ liệu SNPs trung tính và loci đáp ứng, sử dụng cả phân tích nhóm và phân tích thành phần chính (*Principal Component Analysis* - PCA).

Phân tích nhóm, sử dụng phần mềm STRUCTURE v2.3.4 [116], cho phép xác định cấu trúc quần thể dựa trên sự phân nhóm các cá thể theo thông tin di truyền. Các nhóm được hình thành đại diện cho sự chiếm ưu thế của một hoặc nhiều quần thể và có thể được coi là đặc điểm di truyền đặc trưng của từng quần thể. Phương pháp này áp dụng thuật toán suy luận Bayesian để phân bổ các cá thể vào các nhóm thông qua quy trình Markov chain Monte Carlo (MCMC). Giả định ban đầu là các nhóm được chọn ngẫu nhiên, các cá thể được gán vào một nhóm dựa trên kiểu gen của nhiều locus và tần số alen ước tính của mỗi nhóm. Dữ liệu đầu vào cho phân tích là định dạng file “.strc” (chứa thông tin kiểu gen của các cá thể tại các locus) và mô hình hỗn hợp (*admixture*) với tần số alen tương quan (*correlated allele frequencies*) được sử dụng. Mô hình hỗn hợp cho phép các cá thể có thể có tổ tiên từ nhiều quần thể. Số nhóm K được giả định từ 1 đến 10 (với 3 lần lặp cho mỗi giá trị K), sử dụng thông số Length of Burn-in Period là 5.000 và 50.000 lần lặp. Giá trị K tối ưu được xác định dựa trên giá trị trung bình của log-likelihood ($\ln P(D)$) thông qua công cụ trực tuyến Structure Harvester v0.6.94 [205].

Phân tích thành phần chính (PCA) [206] được sử dụng để giảm chiều dữ liệu, giữ lại các thành phần quan trọng mà vẫn duy trì các đặc điểm cơ bản của dữ liệu. Phân tích được thực hiện bằng gói adegenet v1.3.1 [207] trên phần mềm RStudio, với dữ liệu đầu vào là file genpop (.gen) và kí hiệu của các quần thể (popmap). Các thành

phần chính (PC) được giữ lại dựa trên giá trị Eigenvalues, với các PC đầu tiên thường có giá trị lớn hơn, thể hiện các hướng biến thiên lớn nhất trong dữ liệu.

2.3.2.3. Khảo sát sự liên kết quần thể theo khoảng cách địa lý

Để kiểm tra sự liên kết của các quần thể cá theo khoảng cách địa lý (*Isolation-by-distance*), nghiên cứu tiến hành chuyển đổi giá trị khác biệt di truyền (F_{ST}) ở mục 2.4.2 sang $F_{ST}/(1-F_{ST})$ và khoảng cách địa lý giữa 2 quần thể được ước lượng dựa trên chiều dài của sông bằng Google Map (**Phụ lục 7**). Mối tương quan này được đánh giá qua 2 phương pháp:

(1) kiểm định Mantel (*Mantel test*) được thực hiện bằng tính năng ‘vegan’ trên phần mềm RStudio, với tần số hoán vị là 9.999 và độ lặp lại là 1.000 lần [208]. Trong đó, mức độ giải thích của khoảng cách được thể hiện qua hệ số R^2 hiệu chỉnh (*adjusted R^2*) và ý nghĩa của mô hình được xác định dựa trên giá trị P.

(2) Phân tích bản đồ vector riêng Moran dựa trên khoảng cách (dbMEM) được sử dụng nhằm khắc phục hạn chế của kiểm định Mantel trong việc phát hiện cấu trúc không gian đa biến [209]. Phương pháp này sử dụng các hàm riêng (*eigenfunctions*) để phân tách cấu trúc không gian của tập dữ liệu thành nhiều quy mô khác nhau thông qua tính năng “adespatial” trên phần mềm RStudio với tần số hoán vị là 9.999 [210]. Các vector riêng (trục MEM) được sắp xếp theo thứ tự giảm dần của mức độ tương quan không gian, trong đó, các trục đầu tiên đại diện cho xu hướng biến đổi ở quy mô rộng và các trục sau phản ánh quy mô hẹp hơn [209].

2.3.2.4. Phân tích phương sai phân tử (AMOVA) của các quần thể cá

Mức độ khác biệt bên trong, giữa các quần thể và giữa các phân nhóm của ba loài cá điển hình được phân tích bằng phương sai phân tử (*Analysis of molecular variance* - AMOVA) dựa trên phần mềm Alrequin v3.5 sử dụng bộ dữ liệu SNPs trung tính và loci đáp ứng (định dạng “.arp”). Đồng thời, các giá trị thống kê F (F_{IS} , F_{ST} , F_{IT} - so sánh sự khác biệt bên trong, giữa các quần thể) và (F_{SC} , F_{CT} - so sánh sự khác biệt giữa các quần thể khi phân nhóm) cũng được ghi nhận. Phân tích AMOVA được thực hiện theo hai cách tiếp cận: (i) Dựa trên các quần thể thu mẫu theo vị trí

địa lý; và (ii) Dựa trên kết quả phân nhóm di truyền được xác định bởi phân tích STRUCTURE và PCA.

2.3.2.5. Xác định hệ số cận huyết và ước lượng kích thước quần thể hiệu quả

Hệ số cận huyết (F_{IS}) và kích thước quần thể hiệu quả (N_e) được ước tính cho từng nhóm quần thể đã được xác định từ kết quả phân tích cấu trúc di truyền (STRUCTURE) và phân tích thành phần chính (PCA). Tập dữ liệu SNPs trung tính (định dạng “.gen”) được sử dụng làm dữ liệu đầu vào cho cả hai phân tích.

Hệ số F_{IS} được sử dụng để đánh giá mức độ cận huyết trong quần thể bằng cách định lượng sự sai lệch giữa tần suất dị hợp tử quan sát (H_o) và tần suất dị hợp tử mong đợi (H_e) so với trạng thái cân bằng Hardy-Weinberg. Hệ số F_{IS} được tính toán bằng phần mềm GenoDive v2.0b27 [109]. Giá trị F_{IS} dao động từ -1 đến 1, với $F_{IS} > 0$ biểu thị sự thiếu hụt dị hợp tử ($H_o < H_e$), thường là dấu hiệu của cận huyết hoặc hiệu ứng Wahlund do sự pha trộn giữa các quần thể phụ. Ngược lại, $F_{IS} < 0$ cho thấy sự dư thừa dị hợp tử ($H_o > H_e$), có thể phát sinh từ cơ chế tránh giao phối cận huyết, chọn lọc ưu thế dị hợp tử hoặc lai xa giữa các quần thể có kiểu gen khác biệt [211].

Kích thước quần thể hiệu quả - thông số thể hiện số lượng cá thể tham gia sinh sản lý tưởng - được ước tính bằng phần mềm NeEstimator v2.1 [212] dựa trên phương pháp liên kết mất cân bằng (*Linkage Disequilibrium*). Việc ước tính được thực hiện dưới giả định mô hình giao phối ngẫu nhiên (*random mating model*) và khoảng tin cậy 95% được tính toán để xác định độ tin cậy của giá trị N_e . Giá trị N_e được đánh giá dựa trên nguyên tắc 50/500 [213]. Trong đó, ngưỡng ngắn hạn ($N_e \geq 50$) là kích thước quần thể hiệu quả tối thiểu cần thiết để tránh suy thoái cận huyết, ngưỡng dài hạn ($N_e \geq 500$) là mức tối thiểu cần thiết để duy trì đa dạng di truyền và khả năng thích nghi lâu dài.

2.3.3. Dự đoán mô hình di cư của các loài cá ở hạ lưu sông Mekong

Nhằm kiểm chứng giả thuyết về cấu trúc quần đàn theo các hệ thống di cư trong lưu vực đã được ghi nhận [31], nghiên cứu tiến hành phân tích và dự đoán mô hình di cư của 2 loài cá có tập tính di cư khác nhau (cá ét mọi di cư chằng ngấn, tùy nghi

và cá vồ đém di cư chằng dài). Dữ liệu thông tin di truyền (chủ yếu là tần số alen) được phân tích dựa trên 2 phương pháp, mô hình di cư hiện tại (*divMigrate*) [134] và lịch sử phân tách và pha trộn quần thể (*Treemix*) [135]. Riêng đối với cá chạch lá tre, do tập tính ít vận động và chưa có dữ liệu về khả năng di cư [59], do đó, việc dự đoán mô hình di cư của loài cá này là không phù hợp.

2.3.3.1. Khảo sát mô hình di cư hiện tại (*divMigrate*)

Mô hình di cư hiện tại của cá ét mọi (9 quần thể) và cá vồ đém (7 quần thể) ở LMB được xác định dựa trên tính năng ‘*divMigrate*’ của gói dữ liệu ‘*diveRsity*’ [134] và thực hiện trên phần mềm RStudio. Dữ liệu đầu vào là tập dữ liệu SNPs trung tính ở định dạng file genpop (“.gen”). Đầu tiên, mô hình di cư đối xứng được giả định để phù hợp với các quá trình di cư xảy ra ở tất cả các quần thể. Tỷ lệ di cư tương đối (giá trị từ 0 đến 1) được ước tính dựa trên số lượng cá thể di cư hiệu quả trên mỗi thể hệ (*effective number of migrants - N_m*). Sau đó, hướng và tỷ lệ di cư tương đối được tính toán với khoảng tin cậy 95% và độ lặp lại 10.000 lần, và từ đó suy ra quần thể nguồn (*source*) và nhận (*sink*). Mô hình di cư tối ưu được lựa chọn dựa trên giá trị độ tương đồng cosine (*cosine similarity*) cao nhất, đảm bảo tính phản ánh xác thực nhất đối với dữ liệu thực tế. Cuối cùng, mạng lưới di cư được trực quan bằng gói dữ liệu “*qgraph*” [214] trên RStudio.

2.3.3.2. Dự đoán lịch sử phân tách và pha trộn quần thể

Lịch sử phân tách và pha trộn giữa các quần thể cá tại LMB được thực hiện bằng phần mềm *TreeMix* v1.13 [135]. Đầu tiên, một cây quần thể không gốc (*unrooted population tree*) được xây dựng dựa trên tần số alen của tập dữ liệu SNPs trung tính bằng thuật toán khả năng tối đa (ML). Sau đó, quy trình hồi quy từng bước (*stepwise likelihood*) được thực hiện để khảo sát các sự kiện di cư (*migration events, m*) có thể xảy ra, với $m = 0-10$. Tại mỗi giá trị m , một ma trận phương sai phần dư (*residual covariance matrix*) được tính toán, kết hợp với kỹ thuật lấy mẫu lại (*jackknife*) để đánh giá mức ý nghĩa thống kê của các dòng gen di cư. Để xác định số lượng sự kiện di cư tối ưu, phương pháp Evanno (Δm) được áp dụng thông qua công

cụ OptM v0.1.3 [215] trên nền tảng RStudio. Theo đó, giá trị Δm cao nhất được sử dụng làm tiêu chí để xác định mô hình tối ưu phản ánh lịch sử tiến hóa của các quần thể nghiên cứu.

2.4. Lắp ráp và chú giải hệ gen ti thể, xác định mã vạch RAD và khảo sát cấu trúc di truyền quần thể của cá ét mọi *Labeo chrysophekadion*

2.4.1. Lắp ráp và chú giải hệ gen ti thể cá ét mọi

Hệ gen ti thể của cá ét mọi được lắp ráp và chú giải theo quy trình MitoZ v3.4 [216]. Đầu tiên, các đoạn đọc chất lượng cao thu nhận từ mục 2.2.2.1 được kết nối bằng công cụ Megahit v1.2.9 sử dụng thuật toán *de Bruijn graph* [217]. Trong đó, tính năng ‘Quick mode’ với giá trị *kmer* mặc định là 71 được lựa chọn để thực hiện việc kết nối các đoạn đọc (trong đó *kmer* là chiều dài của đoạn trình tự). Sau đó, tập hợp các contigs và scaffolds thu được được sàng lọc để trích xuất các đoạn thuộc hệ gen ti thể bằng cách so sánh với cơ sở dữ liệu HMMER v3.1b2 (<http://hmmer.org/>; [218]) sử dụng mô hình Markov ẩn (Hidden Markov Model - HMM). Các gen mã hóa protein, RNA vận chuyển và RNA ribosome được chú giải lần lượt bằng các phần mềm GeneWise v2.2 [219], MiTFi v1.0 [220] và infernal v1.1.1 [221]. Các gen sau khi chú giải được sắp xếp theo trật tự bằng công cụ BWA v0.7.17 [192] và Samtools v1.15.1 [191]. Cấu trúc dạng vòng của hệ gen ti thể được trực quan hóa bằng phần mềm Circos [222].

Thành phần nucleotide được xác định bằng phần mềm MEGA X [223]. Hệ gen ti thể trong nghiên cứu hiện tại được so sánh độ tương đồng với các trình tự đã công bố trên ngân hàng Gen bằng chương trình BLAST (<https://www.ncbi.nlm.nih.gov/>) và tiến hành đăng ký mã số thông qua công cụ Bankit (<https://submit.ncbi.nlm.nih.gov/about/bankit/>).

2.4.2. Xác định mã vạch RAD thuộc hệ gen ti thể và khảo sát cấu trúc di truyền quần thể của cá ét mọi ở hạ lưu sông Mekong

2.4.2.1. Xác định mã vạch RAD thuộc hệ gen ti thể

Trình tự mã vạch RAD của các quần thể cá ét mọi *L. chrysophekadion* được xác định theo quy trình radBARCODER (<https://github.com/cbirdlab/radBARCODER>, [152]). Cụ thể:

(i) Các đoạn đọc chất lượng cao được sắp xếp (*mapping*) với hệ gen ti thể của cá ét mọi từ nghiên cứu hiện tại bằng thuật toán MEM [224] của công cụ BWA v0.7.17 [192]. Tính năng ‘stats’ của Samtools [191] được sử dụng để loại bỏ những đoạn đọc không đúng vào được hệ gen ti thể;

(ii) xác định độ bao phủ (*coverage*) tại mỗi vị trí base sử dụng tính năng ‘genomecov’ của bedtools, sau đó những vị trí không có độ bao phủ được xác định và xuất thành một tập dữ liệu dưới định dạng bed (.bed);

(iii) Trình tự định dạng fasta của từng cá thể được tạo thành dựa vào kết quả của việc sắp xếp hệ gen ti thể tham chiếu và những vị trí không xác định từ tập dữ liệu bed được lấp vào trình tự là nucleotide N bằng tính năng ‘maskfasta’ của bedtools;

(iv) xác định các vị trí đa hình nucleotide đơn (thay đổi một nucleotide) và chèn/xóa (*insertion and deletion-indel*) và kết hợp chúng thành vị trí đa alen (*multiallelic*) sử dụng lần lượt tính năng ‘call’ và ‘norm’ của bcftools;

(v) Trình tự đồng thuận (*consensus genome sequences*) ở định dạng file fasta của từng cá thể được tạo bằng cách kết hợp file trình tự fasta ở bước (iii) và file vị trí đa alen ở bước (iv) sử dụng tính năng ‘consensus’ của bcftools;

(vi) Các trình tự đồng thuận được đúng hàng (*alignment*) với hệ gen ti thể của cá ét mọi ở nghiên cứu hiện tại và nghiên cứu trước (mã số AP011199, [177]) sử dụng gói dữ liệu pagan2 [225]. Các vị trí nucleotide chèn, xóa hoặc khoảng trống, và các cá thể có tỉ lệ đúng hàng < 50% được tiến hành loại bỏ bằng ‘fltrGENOSITES.R’;

(vii) Tập dữ liệu mã vạch RAD dưới định dạng fasta được thu nhận và chuyển đổi sang định dạng Nexus bằng Seaview [226] để tiến hành các phân tích về đa dạng

di truyền và mạng lưới haplotype. Trình tự mã vạch RAD được xác định vị trí trên hệ gen ti thể bằng tính năng BLAST của ngân hàng Gen và hiển thị kết quả đóng hàng đa trình tự (*Multiple Sequence Alignment Viewer*).

2.4.2.2. Khảo sát đa dạng di truyền và mạng lưới haplotype của các quần thể cá ét mọi ở hạ lưu sông Mekong

Sử dụng dữ liệu mã vạch RAD thuộc hệ gen ti thể, nghiên cứu tiến hành khảo sát sự đa dạng di truyền của các quần thể cá ét mọi ở LMB. Các thông số đa dạng di truyền bao gồm tổng số haplotype (*number of haplotype* - N_h), đa dạng haplotype (*haplotype diversity* - H_d), đa dạng nucleotide (*nucleotide diversity* - π) và số lượng của vị trí đa hình (*number of polymorphic site* - S) được ước tính bằng phần mềm DnaSP v5 [227]. Trong đó, số lượng haplotype và vị trí đa hình thể hiện tổng số các biến thể di truyền và các điểm đột biến tồn tại trong quần thể; đa dạng haplotype và đa dạng nucleotide phản ánh tần suất xuất hiện và mức độ khác biệt trình tự giữa các cá thể. Các công thức dùng để tính toán các thông số được thể hiện ở **Phụ lục 1**. Mức độ đa dạng di truyền được đánh giá dựa trên tiêu chuẩn của Grant và Bowen (1998) [228] theo các ngưỡng: thấp ($\pi < 0,001$; $H_d < 0,5$), trung bình ($0,001 \leq \pi < 0,01$; $0,5 \leq H_d < 0,8$) và cao ($\pi \geq 0,01$; $H_d \geq 0,8$).

Đồng thời, mức độ khác biệt di truyền (F_{ST}) được xác định bằng phần mềm Arlequin v3.5 [202] nhằm đánh giá mức độ cách ly địa lý và dòng gen giữa các cặp quần thể. Cuối cùng, mạng lưới haplotype được xây dựng theo thuật toán TCS (Templeton, Crandall and Sing) trên phần mềm PopART [229].

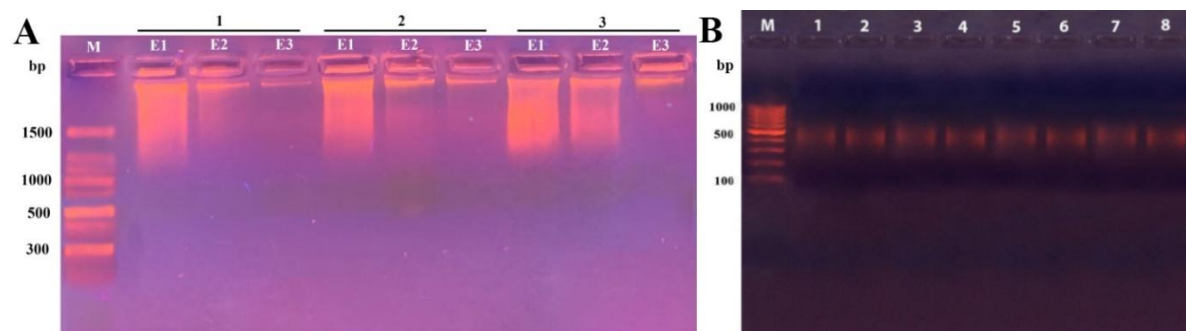
CHƯƠNG 3:

KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Lắp ráp *de novo* hệ gen của ba loài cá điển hình ở lưu vực hạ lưu sông Mekong và xác định các chỉ thị phân tử SNPs

3.1.1. Tạo thư viện hệ gen của ba loài cá và giải trình tự thế hệ mới

Tất cả các mẫu cá thu thập đều được sử dụng để tách chiết DNA tổng số. **Hình 3.1A** minh họa kết quả tách chiết của 3 mẫu thuộc ba loài cá được tiến hành rửa giải 3 lần (E1, E2 và E3) như phương pháp đã được trình bày ở mục 2.2.1.1. Ở cả 3 mẫu, DNA tổng số của lần rửa giải thứ 3 (E3) thể hiện vạch sáng rõ nét và không xuất hiện vết sáng kéo dài (smear) so với các lần còn lại, do đó, các mẫu này được lựa chọn để tiến hành xác định nồng độ DNA. Nghiên cứu ghi nhận tất cả các mẫu đều đạt nồng độ (từ 3 ng/μl trở lên) để thực hiện tạo thư viện gen bằng kỹ thuật EzRAD (**Phụ lục 8**). Thư viện hệ gen của các loài cá thu nhận có kích thước trong khoảng 350 – 550 bp (**Hình 3.1B**).



Hình 3.1. Hình ảnh minh họa kết quả điện di DNA tổng số (A) và thư viện hệ gen (B) của một số mẫu cá trên gel agarose 1,5%

Giếng M: thang DNA chuẩn 100bp (Thermo Scientific), Hình A. Các giếng từ E1-E3: DNA tổng số sau 3 lần rửa giải của các mẫu 1, 2 và 3; Hình B. Các giếng 1-8: Thư viện hệ gen của 8 mẫu cá

Tổng cộng **272** thư viện DNA cá chạch lá tre, **255** thư viện DNA cá ét mọi và **192** thư viện cá vồ đém được xác định nồng độ và giải trình tự sử dụng hệ thống Hiseq 4000 (Illumina) (**Phụ lục 8**). Kết quả giải trình tự thu được **36.447.827**,

1.062.049.264 và 313.556.27 đoạn đọc thô (chiều dài mỗi đoạn đọc là 150 – 151 bp) tương ứng đối với cá chạch lá tre, cá ét mọi và cá vồ đém.

Dữ liệu đoạn đọc và các thông tin sinh học được đăng ký trên Ngân hàng Gen với mã số với BioSample SAMN44070737 - SAMN44071008 và BioProject PRJNA1169209 ở cá chạch lá tre, BioSample SAMN40348039 - SAMN40348293, BioProject PRJNA1085793 ở cá ét mọi và BioSample SAMN47596690 - SAMN47596881, BioProject: PRJNA1242590 ở cá vồ đém.

Sau khi loại bỏ các nucleotide có chất lượng thấp, các nucleotide N, các adapter và đoạn đọc có chiều dài nhỏ hơn 50 bp, số lượng đoạn đọc chất lượng cao được thu nhận lần lượt là **28.733.914** (78,8%) ở cá chạch lá tre, **1.026.721.048** (96,6%) ở cá ét mọi và **259.912.528** (82,9%) ở cá vồ đém. Tất cả các đoạn đọc chất lượng cao được sử dụng để lắp ráp *de novo* hệ gen của mỗi loài cá.

3.1.2. Lắp ráp *de novo* và đánh giá chất lượng hệ gen của ba loài cá

3.1.2.1. Khảo sát phương pháp lắp ráp *de novo* hệ gen cá vồ đém

Khảo sát phương pháp lắp ráp *de novo* hệ gen cá vồ đém bằng ba bộ công cụ dDocent v4.2, Stacks v2.41 và ipyrad v0.93, kết quả được thể hiện ở **Bảng 3.1**. Qua **Bảng 3.1**, công cụ ipyrad thể hiện kết quả lắp ráp được số lượng đoạn contigs và clusters nhiều nhất (1.198.265 contigs và 214.522 clusters) và thấp nhất ở công cụ dDocent (169.083 contigs và 100.397 clusters). Tuy nhiên, hệ gen được lắp ráp bởi công cụ dDocent có chiều dài lớn nhất với 34.390.648 bp, trong khi đó, hệ gen có kích thước nhỏ hơn 31.299.528 bp và 23.726.444 bp lần lượt được lắp ráp bởi công cụ ipyrad và Stack. Bên cạnh đó, chiều dài contig N50 và kích thước cluster lớn nhất thể hiện kết quả tốt nhất ở dDocent so với hai công cụ còn lại.

Để đánh giá độ chính xác của việc lắp ráp hệ gen, tỉ lệ các đoạn contigs được đóng vào hệ gen được xác định. Kết quả ghi nhận tỉ lệ này ở công cụ ipyrad đạt cao nhất (76,58%), tiếp theo là dDocent (71,39%), và Stack (68,57%). Tuy nhiên, nếu xét về tỉ lệ các đoạn contigs đóng chính xác vào hệ gen thì công cụ dDocent đạt tỉ lệ cao nhất (84,07%) (**Bảng 3.1**).

Bảng 3.1. Các thông số lắp ráp và sắp xếp hệ gen cá vồ đém từ ba bộ công cụ

Thông số \ Công cụ	dDocent v4.2	Stack v2.41	Ipyrad v0.93
Thông số lắp ráp			
Số lượng contig	169.083	215.039	1.198.265
Số lượng clusters	100.397	161.310	214.522
Chiều dài hệ gen (bp)	35.729.239	23.726.444	31.299.528
Chiều dài contig N50 (bp)	344	147	147
Kích thước cluster lớn nhất (bp)	527	177	269
Kích thước cluster trung bình (bp)	148,15	148,25	149,23
Thông số các đoạn contig sắp xếp vào hệ gen			
Tỉ lệ mapping (%)	71,39	68,57	76,58
Tỉ lệ mapping chính xác (%)	84,07	83,80	76,89

In đậm thể hiện giá trị cao nhất của mỗi thông số

Dựa trên các tiêu chí so sánh, nghiên cứu hiện tại ghi nhận các công cụ sử dụng thuật toán tham lam (dDocent và ipyrad) đều cho kết quả lắp ráp ưu việt hơn so với công cụ sử dụng thuật toán đồ thị (Stack). Kết quả này cũng tương tự với nghiên cứu của Lacava *et al.* (2019) khi so sánh và đánh giá 6 công cụ lắp ráp *de novo* trên dữ liệu ddRAD giả lập của 2 loài đã biết thông tin hệ gen (thực vật *Arabidopsis thaliana* TAIR 10 và người *Homo sapiens* GRCh38). Qua kết quả đánh giá tỉ lệ lắp ráp so với hệ gen có sẵn, nghiên cứu ghi nhận công cụ CD-HIT – sử dụng thuật toán tham lam – thể hiện kết quả lắp ráp tốt nhất ở cả hai bộ dữ liệu [85].

Tuy nhiên, hiệu quả của các thuật toán thường không đồng nhất giữa các tập dữ liệu khác nhau. Li *et al.* (2018) lại cho rằng công cụ RADassembler – sử dụng thuật toán đồ thị vượt trội hơn so với dDocent hay Rainbow - thuật toán tham lam khi phân tích dữ liệu thực nghiệm từ cá lù đù vàng (*Larimichthys polyactis*), thể hiện qua các chỉ số N50 và chiều dài trung bình của contig, số lượng và tỉ lệ contig được sắp xếp chính xác vào hệ gen [230]. Sự khác biệt này cho thấy việc lựa chọn công cụ tối ưu không chỉ phụ thuộc vào bản chất thuật toán mà còn tùy vào mục đích lắp ráp, tập dữ liệu trình tự, tài nguyên sẵn có (dung lượng máy tính).

Nếu xét thêm về khía cạnh hiệu suất, nghiên cứu hiện tại nhận thấy sự khác biệt đáng kể: dDocent chỉ mất 2 ngày để hoàn thành quy trình, trong khi ipyrad đòi hỏi tới 7 ngày trên cùng một hệ thống. Ưu thế về tốc độ xử lý của dDocent cũng đã được xác nhận trong nghiên cứu của Li *et al.* (2018) [230]. Do đó, với sự cân bằng giữa độ chính xác của thuật toán tham lam và hiệu quả về mặt thời gian, dDocent đã được lựa chọn làm công cụ tối ưu và tiếp tục sử dụng để thực hiện lắp ráp *de novo* hệ gen cho cá chạch lá tre và cá ét mọi trong nghiên cứu này.

3.1.2.2. Lắp ráp *de novo* hệ gen của cá chạch lá tre và cá ét mọi

Từ các đoạn đọc chất lượng cao của cá chạch lá tre (**28.733.914** đoạn đọc) và cá ét mọi (**1.026.721.048** đoạn đọc), nghiên cứu tiến hành lựa chọn các đoạn đọc đặc trưng đại diện cho toàn bộ dữ liệu và tần suất xuất hiện của chúng nhằm loại bỏ những đoạn đọc chỉ chia sẻ bởi một vài cá thể hoặc đoạn đọc với tần suất xuất hiện thấp do lỗi trong quá trình giải trình tự thông qua giá trị cut-off 1 và 2 lần lượt là 3 và 8. Sau khi chọn lọc dữ liệu, **28.458.469** và **1.026.721.048** đoạn đọc đặc trưng được sử dụng để lắp ráp *de novo* hệ gen của cá chạch lá tre và cá ét mọi.

Sử dụng thuật toán tham lam của công cụ dDocent, kết quả nghiên cứu lắp ráp được **89.056** đoạn contigs (kích thước từ 154 – 470 bp) và **90.898** đoạn contigs (kích thước từ 154 – 470 bp) lần lượt ở cá chạch lá tre và cá ét mọi. Các thông số về hệ gen của cá chạch lá tre và cá ét mọi lắp ráp được trình bày ở **Bảng 3.3**. Hệ gen cá chạch lá tre và cá ét mọi được lắp ráp có chiều dài lần lượt là **30.372.557 bp** (tỉ lệ nucleotide GC là 36,1%) và **32.975.632 bp** (tỉ lệ nucleotide GC là 35,9%). Cả hai hệ gen này đều thể hiện tỉ lệ cao các đoạn đọc đóng hàng chính xác vào hệ gen, với 99,8% ở cá chạch lá tre và 99,9% ở cá ét mọi (**Bảng 3.2**).

Kích thước của các hệ gen lắp ráp *de novo* từ dữ liệu EzRAD của 03 loài cá lần lượt là 31 Mb ở cá chạch lá tre, 33 Mb ở cá ét mọi và 35,7 Mb ở cá vô đốm. Kết quả lắp ráp được tương đương 3-5% kích thước hệ gen của cùng loài hoặc loài cùng giống (như cá chạch bông *Macrognathus aculeatus* (638,6 Mb), cá ét mọi (1.076 Mb) và cá *Pangasius pangasius* (792 Mb)) được ước tính bằng hàm lượng DNA trong nhân của

tế bào đơn bội (*C-value*) và công bố trên cơ sở dữ liệu kích thước hệ gen động vật (*Animal Genome Size Database*, <https://www.genomesize.com/index.php>). Điều này cũng phù hợp với chiến lược của kỹ thuật RAD-seq, khoảng 0,1% - 10% kích thước hệ gen được lắp ráp [82], và cũng được ghi nhận ở các nghiên cứu trước đây [154, 231]. Kết quả này cho thấy lắp ráp *de novo* hệ gen không phải là mục đích của RAD-seq, mà thay vào đó, kỹ thuật này hiệu quả trong việc thu nhận hàng ngàn chỉ thị phân tử (như SNPs) phân bố trên toàn bộ hệ gen và được chia sẻ trên số lượng mẫu lớn để ứng dụng trong các nghiên cứu chuyên sâu về tiến hóa, mối quan hệ phát sinh loài và cấu trúc di truyền quần thể của các sinh vật [94].

Để đánh giá chất lượng hệ gen lắp ráp *de novo*, tỉ lệ các đoạn đọc được đóng hàng chính xác vào hệ gen được xác định và đạt 84,07 – 95,2% ở cả ba loài (**Bảng 3.1 và Bảng 3.2**), điều này chứng tỏ hệ gen có chất lượng lắp ráp tốt và đủ điều kiện để thực hiện bước tiếp theo.

Bảng 3.2. Các thông số về hệ gen của cá chạch lá tre và cá ét mọi

Các thông số	Cá chạch lá tre	Cá ét mọi
<i>Thông số lắp ráp</i>		
Chiều dài hệ gen lắp ráp (bp)	30.372.557	32.975.632
Thành phần các nucleotide (A, T, G, C, N)	7.867.091; 7.928.609; 5.469.674; 5.486.511; 3.620.672	8.526.176; 8.734.089; 5.997.475; 5.833.856; 3.884.036
Tổng số đoạn contigs	89.056	90.898
Chiều dài đoạn contig nhỏ nhất (bp)	154	154
Chiều dài đoạn contig lớn nhất (bp)	470	470
Tổng số đoạn contig của N50	44.528	45.449
Chiều dài đoạn contig của N50 (bp)	343	343
<i>Thông số các đoạn contig sắp xếp vào hệ gen</i>		
Tỉ lệ mapping (%)	99,1	99,4
Tỉ lệ mapping chính xác (%)	92,6	95,2

3.1.3. Xác định chỉ thị phân tử SNPs của ba loài cá điển hình ở lưu vực hạ lưu sông Mekong

3.1.3.1. Xác định và sàng lọc chỉ thị SNPs của ba loài cá

Sau khi lắp ráp *de novo* hệ gen của ba loài cá điển hình ở LMB, các chỉ thị SNPs được xác định dựa trên sự khác biệt nucleotide đơn giữa trình tự đoạn đọc của mỗi cá thể khi dóng hàng với hệ gen. Nghiên cứu phát hiện được số lượng SNPs thô lần lượt là **466.068** ở cá chạch lá tre, **103.408** đối với cá ét mọi và **59.769** đối với cá vô dẽm. Sau đó, dữ liệu SNPs thô được tiến hành sàng lọc qua các thông số để thu nhận các SNPs đặc trưng quần thể, kết quả về số lượng cá thể và SNPs qua mỗi bước sàng lọc của ba loài cá được thể hiện trên **Bảng 3.3**.

Bảng 3.3. Số lượng cá thể và SNPs qua các bước lọc của ba loài cá

Các thông số lọc	<i>M. siamensis</i>		<i>L. chrysophekadion</i>		<i>P. larnaudii</i>	
	Số lượng SNPs	Số lượng cá thể	Số lượng SNPs	Số lượng cá thể	Số lượng SNPs	Số lượng cá thể
Các loci có hai alen	441.572	272	103.408	256	59.769	192
Loại bỏ đoạn chèn/xóa	392.771	272	101.858	256	58.326	192
Chất lượng trung bình ($Q \geq 30$)	392.771	272	89.307	256	48.526	192
Độ sâu bao phủ tối thiểu ($\text{MeanDP} \leq 5$)	328.803	272	89.307	256	48.526	192
Tỉ lệ phát hiện SNPs trên tất cả các cá thể (80%)	45.618	272	60.380	256	33.622	192
Số lượng alen thay thế ($\text{MAC} < 3$)	39.463	272	9.029	256	9.022	192
Tần số alen thay thế ($\text{MAF} = 0,05$)	20.123	272	8.246	256	8.556	192
Tỉ lệ phát hiện SNPs trên từng cá thể (>80%)	9.200	239	2.058	256	8.556	192
Cân bằng HWE với độ tin cậy $p < 0,001$	9.043	239	1.774	232	1.842	160
1 SNP/ contig	4.237	239	825	232	1.270	160

Nhằm thu nhận các chỉ thị SNPs có ý nghĩa, nghiên cứu tiến hành sàng lọc qua nhiều thông số với các tiêu chí như đã trình bày ở **Bảng 2.3** trong phần phương pháp. Kết quả thu được ở **Bảng 3.3** cho thấy số lượng SNPs giảm dần qua mỗi chỉ số lọc. Cuối cùng, nghiên cứu tuyển chọn được **4.237 SNPs đặc trưng** của 239 cá thể (9 quần thể) cá chạch lá tre, **825 SNPs đặc trưng** của 232 cá thể (9 quần thể) cá ét mọi và **1.270 SNPs đặc trưng** của 160 cá thể (7 quần thể) cá vồ đém. Một số cá thể không đủ dữ liệu SNPs để phân tích (*Missing individuals*) và được loại bỏ, với số lượng cá thể của mỗi loài cá trước và sau khi lọc được trình bày ở **Phụ lục 9 và 10**.

Số lượng SNPs được xác định thể hiện sự dao động tùy thuộc vào đặc điểm sinh học của từng đối tượng, phạm vi khảo sát và quy trình lọc dữ liệu tin sinh học nghiêm ngặt. Cá chạch lá tre – loài cá không di cư, sống cố định, đạt mức đa hình cao nhất với 4.237 SNPs (từ 9 quần thể, 239 cá thể). Sự đa hình cao này cho thấy việc trao đổi gen thấp giữa các quần thể địa lý do ít di chuyển, cho phép mỗi quần thể cục bộ tích lũy các biến thể di truyền riêng biệt. Bên cạnh đó, việc thu thập mẫu cá chạch lá tre mở rộng thêm quần thể Tachilek (Myanmar) thuộc UMB, giúp thu thập thêm các SNPs đặc trưng ở khu vực này, góp phần làm tăng số lượng SNPs. Kết quả này tương đồng với các nghiên cứu sử dụng kỹ thuật ddRAD trên các loài cá không di cư như *Achondrostoma salmantinum* ở Tây Ban Nha (4.123 SNPs từ 8 quần thể, 74 cá thể) [232] và *Cichla temensis* ở Brazil (2.216 SNPs từ 6 quần thể, 68 cá thể) [233]. Ngược lại, cá vồ đém – loài di cư đường dài, ghi nhận số lượng SNPs trung bình (1.270 SNPs từ 7 quần thể, 160 cá thể), phản ánh dòng gen cao và liên tục giữa các quần thể đã khảo sát. Cơ chế di cư này có xu hướng làm đồng nhất di truyền giữa các quần thể địa lý, dẫn đến sự phân hóa thấp và số lượng SNPs đặc trưng cho từng vùng địa lý ít hơn. Sự khác biệt về số lượng SNPs được ghi nhận so với cá tra (*Pn. hypophthalmus*) (7.263 SNPs) – cũng là loài di cư đường dài – không phải do sự khác biệt về tập tính di cư mà là do yếu tố kỹ thuật. Cụ thể, số lượng SNPs cao ở cá tra là kết quả của việc sử dụng kỹ thuật GBS [18], có khả năng khảo sát số lượng vị trí gen lớn hơn nhiều so với kỹ thuật EzRAD được áp dụng cho cá vồ đém, đồng thời mẫu cá tra thu thập

bao gồm cả cá nuôi và tự nhiên trên quy mô rộng. Cuối cùng, cá ét mọi có số lượng SNPs thấp nhất (825 SNPs từ 9 quần thể, 232 cá thể). Điều này cũng tương tự nghiên cứu về cá lăng *H. spilopterus* (756 SNPs từ 6 quần thể) ở LMB và hệ thống 3S [15], phản ánh mức độ đa hình thấp có thể liên quan đến sự kiện thắt cổ chai di truyền hoặc tác động của sự phân mảnh môi trường sống.

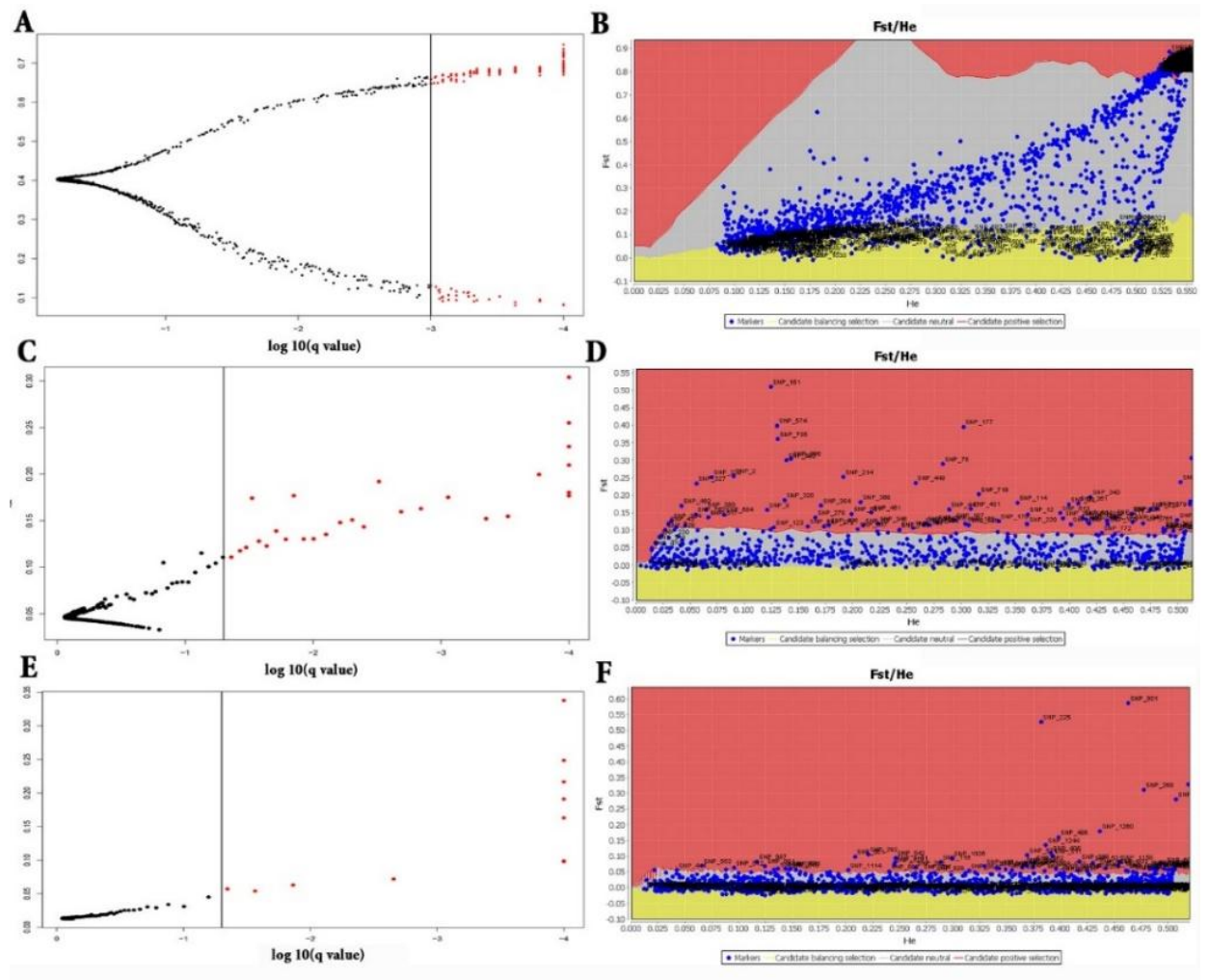
3.1.3.2. Kiểm định loci ngoại vi (*outlier loci*)

Kết quả kiểm định loci ngoại vi – những vị trí loci có giá trị cực khác biệt so với phần còn lại của bộ gen – của các quần thể ở ba loài cá từ hai phương pháp BayeScan và Lositan được thể hiện ở **Hình 3.2**.

Đối với các quần thể cá chạch lá tre, BayeScan xác định được 220 loci ngoại vi với tiêu chí khoảng tin cậy 95% và giá trị $FDR \leq 0,001$ (**Hình 3.2A**). Trong khi đó, LOSITAN phát hiện số lượng loci lớn hơn đáng kể là 470 loci (vùng màu đỏ ở **Hình 3.2B**). Tổng cộng, 501 loci được xác định là loci ngoại vi, với 189 loci trùng lặp được chia sẻ giữa hai phương pháp. Ở cá ét mọi, số lượng loci ngoại vi được phát hiện lần lượt là 28 ở BayeScan và 67 ở LOSITAN (**Hình 3.2C và D**), với 23 loci được chia sẻ giữa hai phương pháp. Tương tự, đối với cá vồ dêm, BayeScan chỉ xác định được 11 loci, trong khi LOSITAN phát hiện 94 loci (**Hình 3.2E và F**), và có 11 loci trùng lặp được ghi nhận. Tổng số loci ngoại vi được xác định cho cá ét mọi và cá vồ dêm lần lượt là 67 và 94.

Sự chênh lệch về số lượng loci ngoại vi được phát hiện bởi BayeScan và LOSITAN đã được ghi nhận trong nhiều nghiên cứu trước đây [156], kết quả này phản ánh sự khác nhau trong thuật toán thống kê và độ nhạy của từng phương pháp. BayeScan, dựa trên mô hình Bayes, thường có tỉ lệ dương tính giả thấp hơn và được xem là bảo thủ hơn, ít bị ảnh hưởng bởi lịch sử tiến hóa hay các quá trình lịch sử sinh học của quần thể (*demographic history*), do đó số lượng loci ngoại vi được phát hiện thường ít hơn nhưng có độ tin cậy cao hơn. Ngược lại, LOSITAN, dựa trên mô phỏng phân bố F_{ST} dưới giả định trung lập (*neutral model*), có độ nhạy cao hơn trong việc phát hiện các loci ngoại vi nhưng dễ dẫn đến tỉ lệ dương tính giả cao hơn, đặc biệt

trong các quần thể có cấu trúc phức tạp hoặc chịu tác động của trôi dạt di truyền và dòng gen không đồng nhất [195]. Do đó, việc kết hợp nhiều phương pháp phân tích và xem xét các loci được phát hiện trùng lặp giữa các mô hình thống kê là cần thiết nhằm xác định tập hợp loci chịu chọn lọc mạnh mẽ và đáng tin cậy nhất.



Hình 3.2. Đồ thị thể hiện kết quả kiểm định loci ngoại vi của cá chạch lá tre (A, B), ét mọi (C, D) và vồ dóm (E, F) theo BayeScan (A, C, E) và Lositan (B, D, F) (A, C, E) Các giá trị F_{ST} thể hiện các dấu chấm dựa trên \log_{10} của PO (posterior odds) cho tất cả các vị trí trong phương pháp BayeScan, trong đó các giá trị của \log_{10} với $FDR = 0,001$ là ngưỡng để xác định vị trí các loci đáp ứng; (B) Các dấu chấm trên đồ thị thể hiện giá trị theo sự phân bố của giá trị dị hợp tử mong đợi (H_e) và F_{ST} ở Lositan, với loci cân bằng (màu vàng), loci trung tính (màu xám) và loci thích ứng (màu đỏ).

So sánh giữa ba loài cá, nghiên cứu hiện tại ghi nhận số lượng loci ngoại vi cao nhất (501 loci) ở cá chạch lá tre. Điều này gợi ý rằng quần thể loài này có thể đang chịu áp lực chọn lọc cục bộ mạnh mẽ hơn, có thể liên quan đến sự phân mảnh môi trường sống hoặc sự khác biệt lớn về điều kiện sinh thái giữa các khu vực địa lý mà chúng sinh sống [104]. Hai loài cá còn lại, cá ét mọi và cá vồ đém, có số lượng loci ngoại vi tương đối thấp hơn, với 65 và 94 loci. Kết quả này tương đồng với các nghiên cứu trên các loài cá có sự trao đổi dòng gen cao do tập tính di cư làm giảm sự thích nghi cục bộ, dẫn đến chỉ một tỉ lệ nhỏ loci ngoại vi được phát hiện [156, 234].

Nghiên cứu tiến hành loại bỏ các loci ngoại vi và chỉ sử dụng dữ liệu các loci trung tính để ước tính các thông số đa dạng di truyền quần thể [104]. Nghiên cứu thu nhận tập dữ liệu SNPs trung tính lần lượt là **3.736 SNPs** (cá chạch lá tre), **760 SNPs** (cá ét mọi) và **1.176 SNPs** (cá vồ đém). Đồng thời, tập hợp các loci được chia sẻ giữa BayeScan và LOSITAN (gồm **189, 23 và 11 loci đáp ứng** lần lượt ở cá chạch lá tre, cá ét mọi và cá vồ đém) được sử dụng làm tập dữ liệu thứ hai trong nghiên cứu để phân tích về sự thích nghi của các quần thể cá. Thông tin về số lượng cá thể và SNPs của mỗi loài cá theo tiến trình xử lý dữ liệu được trình bày chi tiết ở **Phụ lục 10**.

3.2. Khảo sát sự đa dạng di truyền, cấu trúc quần thể và dự đoán mô hình di cư của các loài cá điển hình ở lưu vực hạ lưu sông Mekong

Nghiên cứu sử dụng dữ liệu SNPs trung tính và loci đáp ứng để phân tích sự khác biệt di truyền, cấu trúc quần thể và ước lượng phương sai phân tử (AMOVA) của các loài cá. Tuy nhiên, do không tìm thấy bằng chứng về sự thích nghi rõ ràng của quần thể với môi trường, các kết quả liên quan đến loci đáp ứng được coi là không cung cấp thêm thông tin có ý nghĩa. Do đó, phần kết quả sau đây chỉ trình bày phân tích dựa trên dữ liệu SNPs trung tính. Phân tích đầy đủ sử dụng dữ liệu loci đáp ứng được trình bày trong **Phụ lục 15**.

3.2.1. Sự đa dạng di truyền của các quần thể cá

Kết quả khảo sát đa dạng di truyền quần thể của ba loài cá điển hình ở lưu vực sông Mekong được trình bày ở **Bảng 3.4**. Nhìn chung, ở tất cả các quần thể cá, giá trị

dị hợp tử quan sát đều bằng hoặc nhỏ hơn dị hợp tử mong đợi, cho thấy có dấu hiệu của sự suy giảm đa dạng di truyền hoặc ảnh hưởng của cấu trúc quần thể. Mức độ đa dạng di truyền có sự phân hóa rõ rệt giữa các loài, phản ánh sự khác biệt về đặc điểm sinh học và tập tính di cư của mỗi loài. Trong đó, cá vồ đém (di cư chằng dài) thể hiện mức đa dạng cao nhất ($H_o/H_e = 0,220/0,266$), tiếp theo là cá ét mọi (di cư chằng ngắn và tùy nghi) với H_o/H_e là 0,185 và 0,233. Cá chạch lá tre (loài sống định cư) có mức đa dạng di truyền thấp nhất trong ba loài nghiên cứu ($H_o/H_e = 0,140/0,176$).

Đối với cá chạch lá tre, sự đa dạng di truyền có xu hướng tăng dần từ thượng lưu xuống hạ lưu. Các quần thể thuộc khu vực UMB (Tachileik) và các khu vực phía trên thác Khôn như thượng lưu-LMB, trung lưu-LMB, đặc biệt là các dòng nhánh thuộc cao nguyên Korat như sông Mun (UB-MR), ghi nhận mức đa dạng thấp nhất ($H_o = 0,106$; $H_e = 0,125$). Ngược lại, các quần thể phía dưới thác Khôn thuộc vùng sinh thái Kratié - Stung Treng và khu vực ĐBSCL thể hiện mức đa dạng cao hơn, với giá trị H_o đạt cao nhất tại Kratié (0,206) (**Bảng 3.4**).

Đối với cá ét mọi, mức đa dạng di truyền tập trung cao ở các khu vực trung lưu-LMB (Paksan, Pakse và Ubon Ratchathani) và hợp lưu giữa sông Mekong và 3S (Stung Treng). Các quần thể tại Paksan và Pakse thuộc trung lưu-LMB thể hiện mức đa dạng cao nhất (H_o dao động từ 0,217 đến 0,243). Tuy nhiên, mức đa dạng này có xu hướng giảm dần về phía hạ lưu, thấp nhất tại Đồng Tháp (0,139) (**Bảng 3.4**). Trong khi đó, các quần thể cá vồ đém thể hiện sự ổn định về mặt di truyền trên toàn lưu vực. Giá trị dị hợp tử quan sát cao nhất ở quần thể Ubon Ratchathani thuộc sông Mun (0,263) và thấp nhất ở Biển Hồ-Siêm Riệp (0,225), trong khi đó, dị hợp tử mong đợi cao nhất ở Paksan (0,298) và thấp nhất ở Biển Hồ-Siêm Riệp (0,261) (**Bảng 3.4**).

Bảng 3.4. Các thông số đa dạng di truyền ở các quần thể của ba loài cá điển hình ở lưu vực sông Mekong

Vị trí trên sông Mekong		Vùng sinh thái	Địa điểm thu mẫu	<i>Macrognathus siamensis</i>			<i>Labeo chrysophekadion</i>			<i>Pangasius larnaudii</i>		
				Nse	Ho	He	Nse	Ho	He	Nse	Ho	He
Thượng lưu-UMB	Dòng nhánh	Thượng Lan Thương	TK	32	0,140	0,140						
Thượng lưu-LMB - Trên thác Khôn	Dòng nhánh (Sông Khan)	Hạ Lan Thương	LP	20	0,128	0,141	20	0,193	0,208			
Trung lưu-LMB - Trên thác Khôn	Dòng chính		PA	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	
	Dòng nhánh (Sông Mun)	Cao nguyên Korat	UB-MR	31	0,106	0,125				15	0,268	0,266
	Dòng nhánh (Sông Chi)		RE				30	0,140	0,234			
Hạ lưu-LMB - dưới thác Khôn	Hợp lưu giữa dòng chính và 3S	Kratie - Stung Treng	ST	32	0,146	0,236	24	0,238	0,238	28	0,230	0,270
	Dòng chính		KT	27	0,206	0,239	25	0,143	0,237			
	Biển Hồ		SR							22	0,225	0,261
	Dòng chính	Châu thổ Mekong	AG	28	0,153	0,202	24	0,185	0,228	19	0,235	0,265
	Dòng chính		DT	21	0,158	0,210	27	0,139	0,222	30	0,223	0,269
Tổng cộng/Trung bình				239	0,140	0,176	232	0,185	0,233	160	0,220	0,266

Ghi chú: Nse: Số lượng mẫu phân tích, Ho: Dị hợp tử quan sát, He: Dị hợp tử mong đợi.

3.2.2. Cấu trúc di truyền quần thể của các loài cá ở LMB

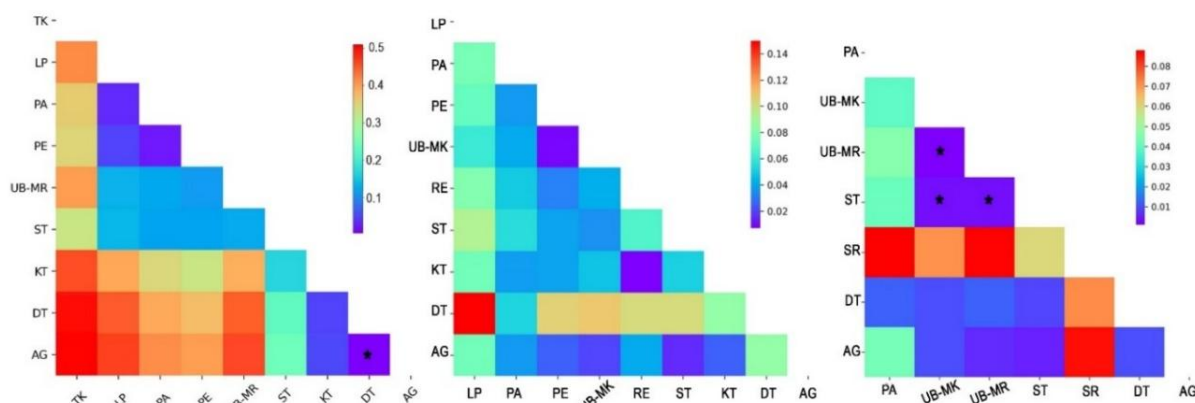
3.2.2.1. Sự khác biệt di truyền của các quần thể cá ở LMB

Sự khác biệt di truyền chung (G_{ST}) thể hiện sự phân hóa rõ rệt của ba loài cá nghiên cứu. Chỉ số G_{ST} được giá trị cao nhất ở loài cá chạch lá tre ($G_{ST}= 0,303$), tiếp theo là cá ét mọi ($G_{ST} = 0,038$) và thấp nhất ở cá vồ đém ($G_{ST} = 0,003$). Tất cả các giá trị G_{ST} đều có ý nghĩa thống kê ($P < 0,001$), cho thấy cấu trúc di truyền quần thể được hình thành rõ nét.

Đối với cá chạch lá tre, hầu hết các cặp quần thể đều thể hiện sự khác biệt di truyền có ý nghĩa thống kê ($P \leq 0,05$ đã được hiệu chỉnh FDR, **Phụ lục 13**), ngoại trừ cặp quần thể Đồng Tháp và An Giang. Trong đó, sự khác biệt rõ rệt được ghi nhận giữa quần thể Tachileik (TK) – dòng nhánh thuộc UMB so với tất cả các quần thể còn lại thuộc LMB, với giá trị F_{ST} dao động từ 0,331 (TK/ST) đến 0,509 (TK/AG). Tại lưu vực LMB, các quần thể hạ lưu-LMB ở phía dưới thác Khôn (KT, ST, DT và AG) thể hiện sự khác biệt di truyền cao hơn so với các quần thể thượng lưu-LMB và trung lưu-LMB ở phía trên thác (LP, PA, PE và UB-MR) (**Hình 3.3A; Phụ lục 11**).

Đối với cá ét mọi, giá trị F_{ST} dao động từ 0,007 (RE/KT) đến 0,15 (LP/DT) và tất cả các cặp so sánh đều có ý nghĩa thống kê ($P \leq 0,037$ đã được hiệu chỉnh FDR, **Phụ lục 13**). Trong đó, quần thể Đồng Tháp ($F_{ST} = 0,052 - 0,15$) và Luang Prabang ($F_{ST} = 0,06 - 0,15$) thể hiện sự khác biệt di truyền lớn so với các quần thể còn lại (**Hình 3.3B; Phụ lục 11**).

Cá vồ đém có mức độ khác biệt di truyền thấp nhất trong ba loài cá nghiên cứu ($F_{ST} = 0,001 - 0,088$), trong đó, 18/21 cặp quần thể được so sánh thể hiện sự khác biệt di truyền có ý nghĩa thống kê ($P < 0,05$ đã được hiệu chỉnh FDR, **Phụ lục 13**). Sự khác biệt di truyền lớn nhất được ghi nhận ở quần thể Siem Riệp (SR) với F_{ST} từ 0,06 – 0,088, tiếp theo là quần thể Paksan (PA) (F_{ST} từ 0,012 – 0,088). Đáng chú ý, có ba cặp quần thể (UB-MK/ST, UB-MK/UB-MR, UB-MR/ST) không cho thấy sự khác biệt di truyền có ý nghĩa (được đánh dấu * trên **Hình 3.3C; Phụ lục 11**), phản ánh sự trao đổi gen mạnh mẽ giữa các khu vực này.



Hình 3.3. Biểu đồ heatmap thể hiện giá trị khác biệt di truyền theo cặp quần thể (F_{ST}) của cá chạch lá tre (A), cá ét mọi (B) và cá vô đém (C). * chỉ ra sự khác biệt không có ý nghĩa thống kê sau khi hiệu chỉnh FDR

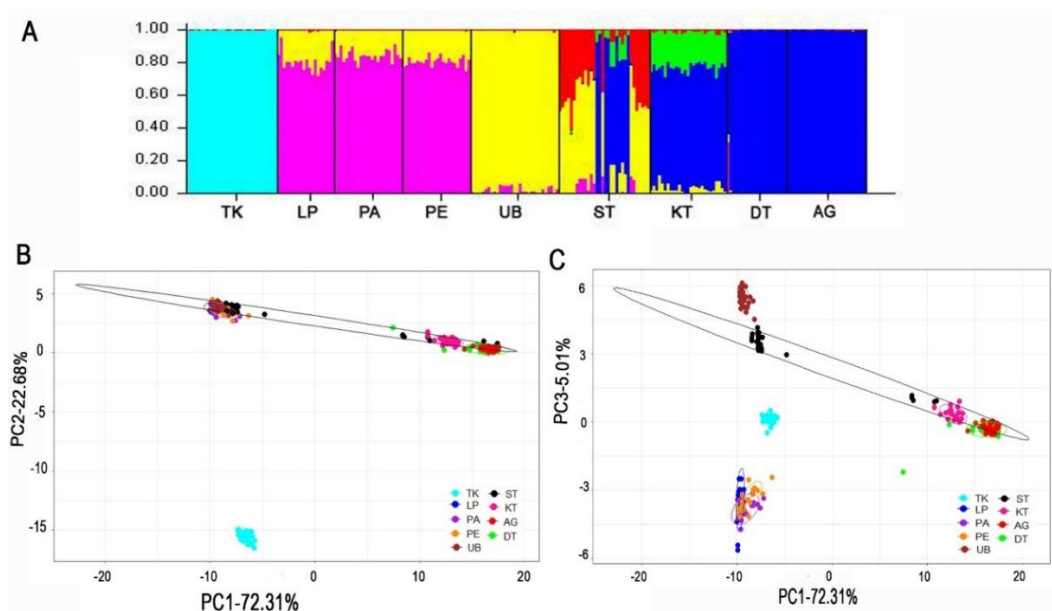
3.2.2.2. Cấu trúc quần thể của ba loài cá điển hình ở LMB

a. Cấu trúc quần thể của cá chạch lá tre

Phân tích STRUCTURE xác định số nhóm quần thể tối ưu (K) = 6 thông qua chỉ số likelihood cao nhất (**Phụ lục 12**). Tại $K = 6$, tỉ lệ xác suất các cá thể cá chạch lá tre thuộc từng quần thể thu mẫu được chỉ định thuộc mỗi nhóm tương ứng (**Hình 3.4A**). Quần thể Tachileik (Myanmar) thể hiện sự phân tách rõ rệt so với các quần thể còn lại, tạo thành một nhóm riêng biệt (màu xanh dương; 99,9%). Các cá thể thuộc các quần thể Luang Prabang, Paksan và Pakse (Lào), thuộc nhóm 2 (màu hồng; lần lượt chiếm 80,2%; 84,1%; và 80,6%). Mẫu thu thập từ quần thể Ubon Ratchathani-sông Mun (Thái Lan) thuộc nhóm 3 (màu vàng; 97,7%). Quần thể Stung Treng (nhóm 4) thể hiện sự trộn lẫn thông tin di truyền từ các nhóm (màu đỏ 26,8%; màu xanh đậm 28,1% và màu vàng là 39,5%; màu hồng và xanh lá chiếm tỉ lệ thấp (lần lượt là 2,6% và 3%)). Nhóm 5 gồm các cá thể từ Kratié - Campuchia và ĐBSCL - An Giang và Đồng Tháp (màu xanh đậm; lần lượt chiếm 73,4%; 97,9% và 99,8%), trong đó Kratié chia sẻ thông tin di truyền với Stung Treng (màu xanh lá 21,9%).

Đối với phân tích thành phần chính PCA, sự khác biệt di truyền tập trung ở trục thứ nhất (PC-1) với 72,31%, thứ hai (PC-2) và thứ ba (PC-3) chiếm lần lượt 22,68% và 5,01% (**Hình 3.4B&C**). Kết quả ghi nhận sự phân tách nhóm quần thể tương tự như phân tích STRUCTURE, ngoại trừ quần thể Ubon Ratchathani ở sông Mun (UB-

MR). Trong khi ở trục PC-1 và PC-2, quần thể UB-MR tạo thành nhóm với các quần thể LP, PA và PE (**Hình 3.4B**), thì ở trục PC-1 và PC-3, quần thể này thể hiện sự phân tách thành nhóm riêng biệt (**Hình 3.4C**).

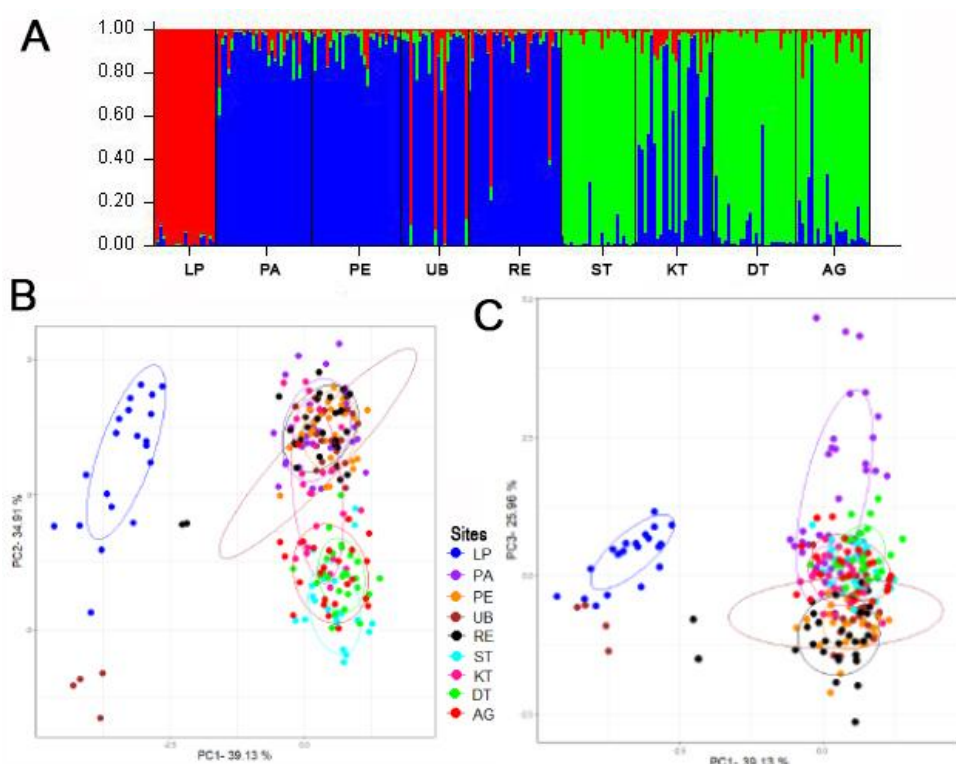


Hình 3.4. Cấu trúc quần thể cá chạch lá tre dựa trên phân tích nhóm Structure (A) và thành phần chính PCA (B, C) sử dụng 1.936 SNPs trung tính

b. Cấu trúc quần thể của cá ét mọi

Ba nhóm quần thể tối ưu được xác định dựa trên phân tích STRUCTURE của cá ét mọi ở LMB (**Phụ lục 12**), tương ứng với ba khu vực sinh thái. Các cá thể từ LP ở thượng lưu-LMB thuộc nhóm 1 (màu đỏ; 97,5%), các quần thể ở trung lưu-LMB (PA, PE, UB-MK và RE) thuộc nhóm 2 (màu xanh đậm; lần lượt chiếm 92,2%; 93,1%; 75,3% và 89,8%), và các quần thể ở hạ lưu-LMB (ST, KT, DT và AG) thuộc nhóm 3 (màu xanh lá; với tỉ lệ 94,8%; 45,3%; 91,8% và 84,2%). Bên cạnh đó, nghiên cứu ghi nhận 4 cá thể ở UB-MK và 2 cá thể ở RE chia sẻ thông tin di truyền với LP. Hơn nữa, sự trộn lẫn thông tin di truyền từ các quần thể ở trung lưu với Kratié cũng được ghi nhận, với tỉ lệ màu xanh đậm chiếm 49,6% ở nhóm này (**Hình 3.5A**).

Đối với phân tích PCA, quần thể Luang Prabang được phân tách thành một nhóm riêng biệt dựa trên hai thành phần chính PC-1/ PC-2 và PC-1/ PC-3. Các quần thể thuộc trung lưu-LMB và hạ lưu-LMB tạo thành 2 nhóm riêng biệt ở PC-1 và PC-2 (**Hình 3.5B**), trong khi đó, chúng lại tạo thành 1 nhóm ở PC-1 và PC-3 (**Hình 3.5C**).

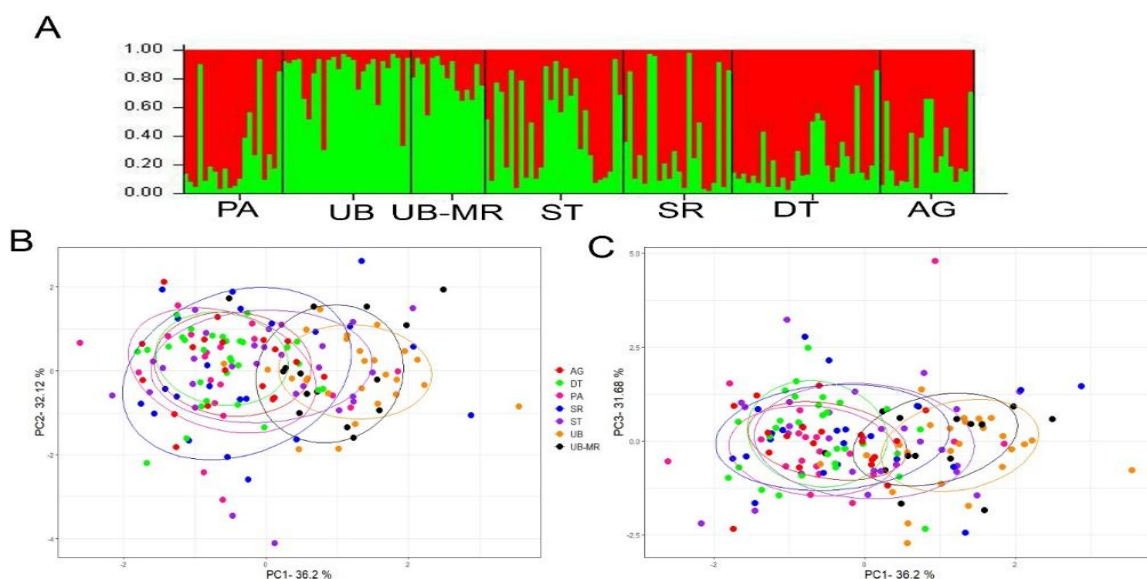


Hình 3.5. Cấu trúc quần thể cá ét mọi dựa trên phân tích nhóm Structure (A) và thành phần chính PCA (B, C) sử dụng 760 SNPs trung tính

c. Cấu trúc quần thể của cá vồ đém

Dựa trên phân tích STRUCTURE, hai nhóm quần thể cá vồ đém tương ứng với giá trị K tối ưu là 2 được xác định (**Phụ lục 12**). Nhóm 1 (màu đỏ) chiếm ưu thế ở các quần thể trung lưu-LMB (PA - 66,6%), và hạ lưu-LMB (SR - 60,1%; DT- 68,3%; AG - 69,3%). Nhóm 2 (màu xanh lá) chiếm ưu thế ở UB-MK (82%) và UB-MR (78,8%). Quần thể ST thể hiện sự trộn lẫn thông tin di truyền của 2 nhóm này (47,5% màu đỏ và 52,5% màu xanh lá) (**Hình 3.6A**).

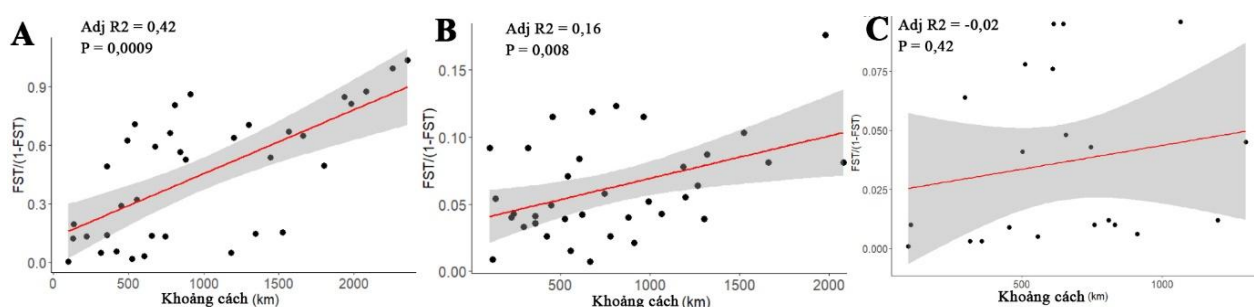
Đối với phân tích PCA, thành phần chính 1 (PC1), PC2 và PC3 giải thích lần lượt 36,2%; 32,12% và 31,68% sự khác biệt di truyền giữa các quần thể (**Hình 3.6B&C**). Kết quả cho thấy sự phân tách không rõ ràng, chứng tỏ có sự kết nối cao giữa các quần thể cá vồ đém.



Hình 3.6. Cấu trúc quần thể cá vồ đém dựa trên phân tích nhóm Structure (A) và thành phần chính PCA (B, C) sử dụng 1.176 SNPs trung tính

3.2.2.3. Sự liên kết quần thể của các loài cá theo khoảng cách địa lý (IBD)

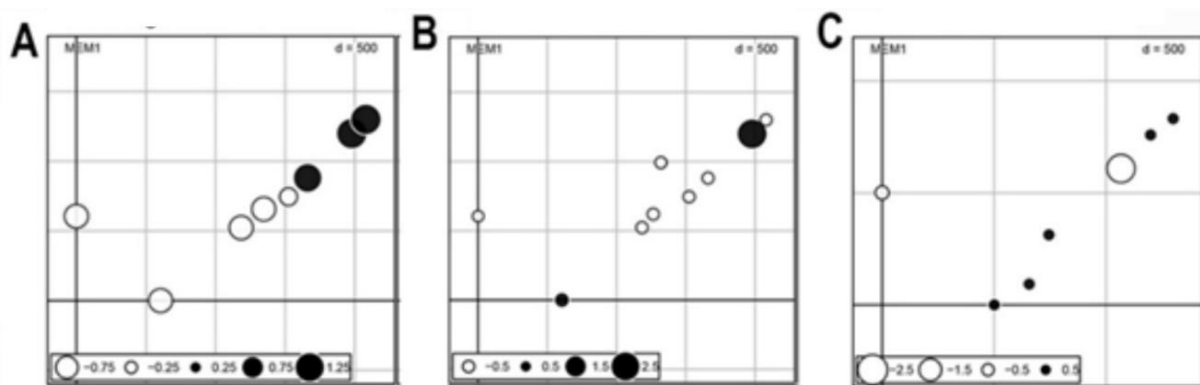
Kết quả kiểm định Mantel cho thấy mối tương quan thuận có ý nghĩa thống kê giữa khoảng cách địa lý và sự khác biệt di truyền ở quần thể cá chạch lá tre (adjusted $R^2 = 0,42$; $P = 0,0009$) và cá ét mọi (adjusted $R^2 = 0,16$; $P = 0,008$) (**Hình 3.7A&B**). Ngược lại, quần thể cá vồ đém không thể hiện mối tương quan này (adjusted $R^2 = -0,02$; $P = 0,42$; **Hình 3.7C**), trong đó giá trị R^2 hiệu chỉnh âm cho thấy khoảng cách địa lý không phải là yếu tố giải thích cho sự khác biệt di truyền của loài cá này.



Hình 3.7. Kết quả kiểm định Mantel về mối tương quan giữa khoảng cách địa lý và sự khác biệt di truyền của cá chạch lá tre (A), ét mọi (B) và vồ đém (C)

Đối với phân tích dbMEM, chỉ có trục MEM-1 thể hiện sự phân tách cấu trúc không gian của các quần thể ở cả 03 loài cá, với giá trị quan sát được ghi nhận là 0,7379; 0,1449 và 0,1776 lần lượt ở cá chạch lá tre, cá ét mọi và cá vồ đém (tất cả

đều có ý nghĩa thống kê $P < 0,01$) (**Hình 3.8**). Điều này cho thấy cấu trúc di truyền của các quần thể chỉ thể hiện sự phân tách ở quy mô rộng. Ngược lại, các trục MEM2-6 không có ý nghĩa ($P > 0,05$) (**Phụ lục 14**), chứng tỏ khoảng cách địa lý ở quy mô nhỏ không ảnh hưởng rõ rệt đến sự khác biệt di truyền của các loài này.



Hình 3.8. Các trục MEM-1 giữa sai khác di truyền và khoảng cách địa lý từ phân tích dbMEM của cá chạch lá tre (A), cá ét mọi (B) và cá vồ dằm (C). Các vòng tròn biểu thị vị trí và giá trị dbMEM của từng cặp phân tích theo F_{ST} và khoảng cách địa lý. Vòng tròn màu đen là giá trị dương và màu trắng là giá trị âm.

3.2.2.4. Phân tích phương sai phân tử (AMOVA) của các loài cá

Bảng 3.5 trình bày thông tin chi tiết về số lượng và nhóm quần thể được sử dụng trong phân tích AMOVA cho từng loài cá dựa trên 2 cách tiếp cận như đã trình bày trong phần phương pháp (**Mục 2.3.2.4**).

Bảng 3.5. Số nhóm quần thể sử dụng để phân tích phương sai phân tử AMOVA

Loài Nhóm	Cá chạch lá tre	Cá ét mọi	Cá vồ dằm
Vị trí thu mẫu	09 quần thể	09 quần thể	07 quần thể
Kết quả phân nhóm bằng STRUCTURE và PCA	04 nhóm quần thể: (i) UMB (TK) (ii) thượng và trung lưu-LMB (LP, PA, PE) (iii) trung lưu-LMB (UB-MR) (iv) hạ lưu-LMB (ST, KT, AG, DT)	03 nhóm quần thể: (i) thượng lưu-LMB (LP) (ii) trung lưu-LMB (PA, PE, UB-MK, RE) (iii) hạ lưu-LMB (ST, KT, AG, DT)	02 nhóm quần thể: (i) thượng và hạ lưu-LMB (PA, SR, KT, AG, DT) (ii) trung lưu-LMB (UB-MK, UB-MR)

Kết quả phân tích mức độ thay đổi phân tử cá chạch lá tre được trình bày ở **Bảng 3.6**. Dựa vào 9 quần thể thu mẫu, kết quả ghi nhận sự biến đổi di truyền xảy ra đáng kể ở cả ba cấp độ và đều có ý nghĩa thống kê ($P < 0,001$). Trong đó, biến động giữa các cá thể chiếm tỉ lệ cao nhất (40,4%), tiếp theo là sự khác biệt giữa các quần thể với 30,4% và giữa các cá thể trong quần thể chiếm 29,2%. Mặt khác, khi xem xét 4 nhóm quần thể, nghiên cứu phát hiện sự phân hóa di truyền đáng kể giữa các nhóm, chiếm 31,5% tổng biến dị ($P = 0,003 < 0,05$). Ở giả thuyết này, biến động giữa các cá thể vẫn chiếm tỉ lệ cao (47,4%). Các thành phần biến động còn lại bao gồm giữa các cá thể trong quần thể và giữa các quần thể trong cùng một nhóm lần lượt ghi nhận ở mức 11,0% và 10,1% với độ tin cậy cao ($P < 0,001$).

Bảng 3.6. Kết quả ước lượng mức độ thay đổi phân tử AMOVA của các quần thể cá chạch lá tre *Macrognathus siamensis*

Nguồn biến động	Bậc tự do	Tổng bình phương	Thành phần biến động	% biến động	Chỉ số cố định	Giá trị P
Theo 9 quần thể thu mẫu						
Giữa các quần thể	8	85.759,8	190,8 Va	30,4	$F_{ST} = 0,304$	$<0,001$
Giữa các cá thể trong quần thể	230	142.538,	183,1 Vb	29,2	$F_{IS} = 0,419$	$<0,001$
Giữa các cá thể	239	60.601	253,6 Vc	40,4	$F_{IT} = 0,596$	$<0,001$
Tổng cộng	477	288.899,2	627,5	100		
Theo 4 nhóm quần thể dựa trên phân tích STRUCTURE và PCA						
Giữa các nhóm	3	52.684,2	136,3 Va	31,5	$F_{CT} = 0,315$	0,003
Giữa các quần thể trong các nhóm	5	12.414,9	43,4 Vb	10,1	$F_{SC} = 0,147$	$<0,001$
Giữa các cá thể trong quần thể	230	69.064,5	47,6 Vc	11,0	$F_{IS} = 0,188$	$<0,001$
Giữa các cá thể	239	49.018,5	205,1 Vd	47,4	$F_{IT} = 0,525$	$<0,001$
Tổng cộng	477	183.182	432,4	100		

Kết quả phân tích AMOVA cá ét mọi được trình bày ở **Bảng 3.7**. Ở cả hai giả thuyết (9 quần thể thu mẫu và 3 nhóm từ phân tích STRUCTURE và PCA), nguồn biến dị di truyền chủ yếu đều từ sự khác biệt giữa các cá thể, với tỉ lệ chiếm 50% và 84,49% và có ý nghĩa thống kê ($F_{IT} = 0,158 - 0,496$; $P < 0,001$), trong khi khác biệt quần thể và khác biệt giữa các nhóm chiếm tỉ lệ thấp (6% và 2,37%; $P < 0,001$).

Bảng 3.7. Kết quả ước lượng mức độ thay đổi phân tử AMOVA của các quần thể cá ét mọi *Labeo chrysophekadion*

Nguồn biến động	Bậc tự do	Tổng bình phương	Thành phần biến động	% biến động	Chỉ số cố định	Giá trị P
Theo 9 quần thể thu mẫu						
Giữa các quần thể	8	4.629,9	7,720	6	$F_{ST} = 0,059$	<0,001
Giữa các cá thể trong quần thể	223	40.506,3	57,644	44	$F_{IS} = 0,465$	<0,001
Giữa các cá thể	232	15.394,5	66,356	50	$F_{IT} = 0,496$	<0,001
Tổng cộng	463	60.530,8	131,720	100		
Theo 3 nhóm quần thể dựa trên phân tích STRUCTURE và PCA						
Giữa các nhóm	2	125.132	0,271	2,37	$F_{CT}=0,024$	0,001
Giữa các quần thể trong các nhóm	6	169.493	0,307	2,69	$F_{SC}=0,028$	<0,001
Giữa các cá thể trong quần thể	223	2.686,8	1,225	10,74	$F_{IS}=0,113$	<0,001
Giữa các cá thể	232	2.227	9,599	84,19	$F_{IT}=0,158$	<0,001
Tổng cộng	463	5.208,4	11,401	100		

Kết quả phân tích AMOVA của 7 quần thể cá vồ đếm cho thấy mức độ thay đổi phân tử thấp giữa các quần thể (3%). Phần lớn biến dị di truyền được tìm thấy giữa các cá thể (57%), trong khi biến dị di truyền giữa các cá thể trong quần thể là 39%, và tất cả các cấp độ đều có ý nghĩa thống kê ($P < 0,001$). Khi xem xét 2 nhóm quần thể được phân định bởi STRUCTURE và PCA, phân tích AMOVA cho thấy sự khác biệt di truyền không đáng kể giữa 2 nhóm (0,4%) và giữa các quần thể trong các nhóm (-0,1%) và đều không có ý nghĩa thống kê ($P > 0,05$). Mức độ thay đổi phân tử

giữa các cá thể vẫn được ghi nhận rất cao (87,3%; $P < 0,001$), tương tự như kết quả phân tích dựa trên 7 quần thể thu mẫu (**Bảng 3.8**).

Bảng 3.8. Kết quả ước lượng mức độ thay đổi phân tử AMOVA của các quần thể cá vồ đém *Pangasius larnaudii*

Nguồn biến động	Bậc tự do	Tổng bình phương	Thành phần biến động	% biến động	Chỉ số cố định	Giá trị P
Theo 7 quần thể thu mẫu						
Giữa các quần thể	6	3.658,7	7,044	3	$F_{ST} = 0,033$	0,001
Giữa các cá thể trong quần thể	153	44.417,8	83,900	39	$F_{IS} = 0,406$	0,001
Giữa các cá thể	160	19.602	122,513	57	$F_{IT} = 0,426$	0,001
Tổng cộng	319	67.678,5	213,456	100		
Theo 2 nhóm quần thể dựa trên phân tích STRUCTURE và PCA						
Giữa các nhóm	1	56,559	0,139 Va	0,4	$F_{CT}=0,004$	0,09
Giữa các quần thể trong các nhóm	5	193,328	-0,034 Vb	-0,1	$F_{SC}=-0,001$	0,825
Giữa các cá thể trong quần thể	162	6.529,2	4,458 Vc	12,4	$F_{IS}=0,124$	<0,001
Giữa các cá thể	169	5.304,5	31,388 Vd	87,3	$F_{IT}=0,127$	<0,001
Tổng cộng	337	12.083,6	35,95099	100		

3.2.2.5. Hệ số cận huyết và kích thước quần thể hiệu quả

Kết quả về hệ số cận huyết và kích thước quần thể hiệu quả của ba loài cá được trình bày ở **Bảng 3.9**. Về hệ số cận huyết, tất cả các nhóm quần thể của ba loài cá trong nghiên cứu hiện tại đều cho thấy giá trị F_{IS} dương (dao động từ 0,17 đến 0,61), điều này chỉ ra một sự thiếu hụt dị hợp tử đáng kể so với trạng thái cân bằng HWE trong các nhóm quần thể. Cá chạch tre có mức độ cận huyết cao nhất, đặc biệt là nhóm Trung lưu-LMB (thuộc sông Mun) với giá trị F_{IS} cao nhất (0,61), phản ánh sự cô lập và phân mảnh mạnh mẽ của nhóm này. Ngược lại, cá vồ đém có F_{IS} thấp nhất (0,17 cho tất cả các quần thể ở LMB), cho thấy mức độ giao phối cận huyết tương

đôi ít hơn. Các phân nhóm quần thể cá ét mọi thể hiện hệ số cận huyết tương tự nhau, dao động từ 0,23 (thượng lưu-LMB), 0,25 (trung lưu-LMB) và 0,30 (hạ lưu-LMB).

Bảng 3.9. Hệ số cận huyết và kích thước quần thể hiệu quả của ba loài cá

Nhóm quần thể	Nse	F _{IS}	Ne (95% C _{IS})
<i>Cá chạch lá tre M. siamensis</i>			
UMB (Tachilek – Myanmar)	32	0,49	109,4 (106,2 – 112,8)
Trung lưu-LMB – Dòng chính và trên thác Khôn	68	0,58	292,1 (283,5 – 301,3)
Trung lưu-LMB – Dòng nhánh và trên thác Khôn	31	0,61	454,3 (399,2 – 526,7)
Hạ lưu-LMB – Dưới thác Khôn	108	0,40	128,7 (113,6 – 140,1)
<i>Cá ét mọi L. chrysophekadion</i>			
Thượng lưu-LMB (Luang Prabang)	20	0,23	193,5 (147,9 - 277,7)
Trung lưu-LMB – Trên thác Khôn	112	0,25	101,2 (98,7 - 103,9)
Hạ lưu-LMB – Dưới thác Khôn	100	0,30	442,6 (399,4 - 495,7)
<i>Cá vồ đém P. larnaudii</i>			
Thượng lưu-LMB và hạ lưu-LMB	126	0,23	347,2 (332,3 – 363,4)
Trung lưu-LMB – Trên thác Khôn	43	0,18	116,7 (111,8 – 120)
Tất cả các quần thể ở LMB	169	0,17	430,6 (414,1 – 448,3)

Ghi chú: Nse: Số lượng mẫu phân tích; F_{IS}: Hệ số cận huyết; Ne: kích thước quần thể hiệu quả

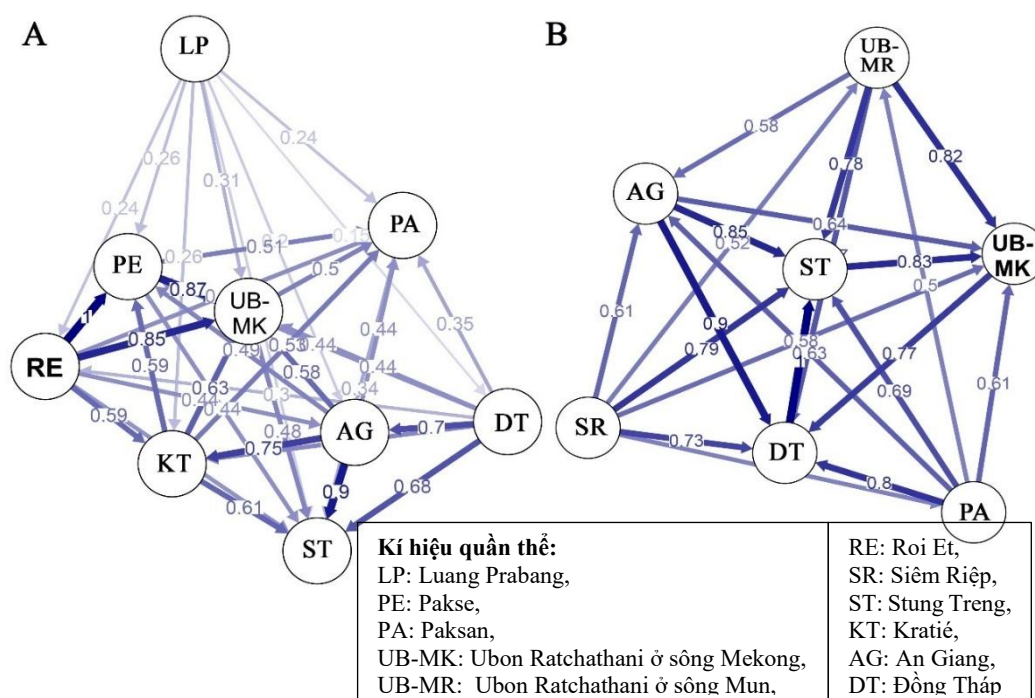
Kích thước quần thể hiệu quả (Ne) của các nhóm quần thể cá cho thấy sự phân hóa rõ rệt và hầu hết đều thấp hơn ngưỡng duy trì đa dạng di truyền dài hạn (Ne < 500) [213]. Ở cá chạch lá tre, giá trị Ne cao nhất (454,3) được ghi nhận ở nhóm quần thể tại Trung lưu-LMB (thuộc sông Mun) và thấp nhất (109,4) ở UMB (Tachilek). Hai nhóm còn lại là Trung lưu-LMB (dòng chính và trên thác Khôn) và hạ lưu-LMB (dưới thác Khôn) có Ne lần lượt là 292,1 và 128,7. Đối với cá ét mọi, giá trị Ne được ước tính lần lượt là 193,5 và 442,6 cho các nhóm thượng lưu-LMB (Luang Prabang) và hạ lưu-LMB (dưới thác Khôn). Riêng nhóm quần thể ở trung lưu-LMB (dòng chính và trên thác Khôn) chỉ ghi nhận giá trị Ne là 101,2; điều này cho thấy nhóm

quần thể đang có nguy cơ mất đa dạng di truyền do trôi dạt di truyền mạnh mẽ. Đối với cá vồ đém, khi xem xét 2 nhóm quần thể dựa trên phân tích STRUCTURE và PCA, Ne được ghi nhận lần lượt là 347,2 ở nhóm thượng lưu-LMB và hạ lưu-LMB, và 116,7 ở trung lưu-LMB. Bên cạnh đó, giá trị Ne được ước tính chung ở LMB là 430,6; cho thấy tiềm năng di truyền chung của loài này ở lưu vực vẫn ở mức cao.

3.2.3. Dự đoán mô hình di cư của các loài cá ở LMB

3.2.3.1. Mô hình di cư hiện tại của các loài cá

Mô hình di cư hiện tại được suy luận từ phương pháp divMigrate của cá ét mọi và cá vồ đém ở LMB với độ tương đồng cosin cao nhất được thể hiện ở **Hình 3.9**. Nhìn chung, cả hai loài cá đều thực hiện quá trình di cư xuôi dòng và ngược dòng, tuy nhiên, cá vồ đém thể hiện sự di cư mạnh mẽ hơn với tỉ lệ di cư tương đối từ 0,5 (SR đến UB-MK - ngược dòng, qua thác Khôn) đến 0,9 (AG đến DT - di cư ngang giữa 2 nhánh sông ở ĐBSCL), trong khi tỉ lệ này ở cá ét mọi dao động từ 0,2 (LP đến AG - xuôi dòng) đến 1 (RE đến PE – di cư ngang).



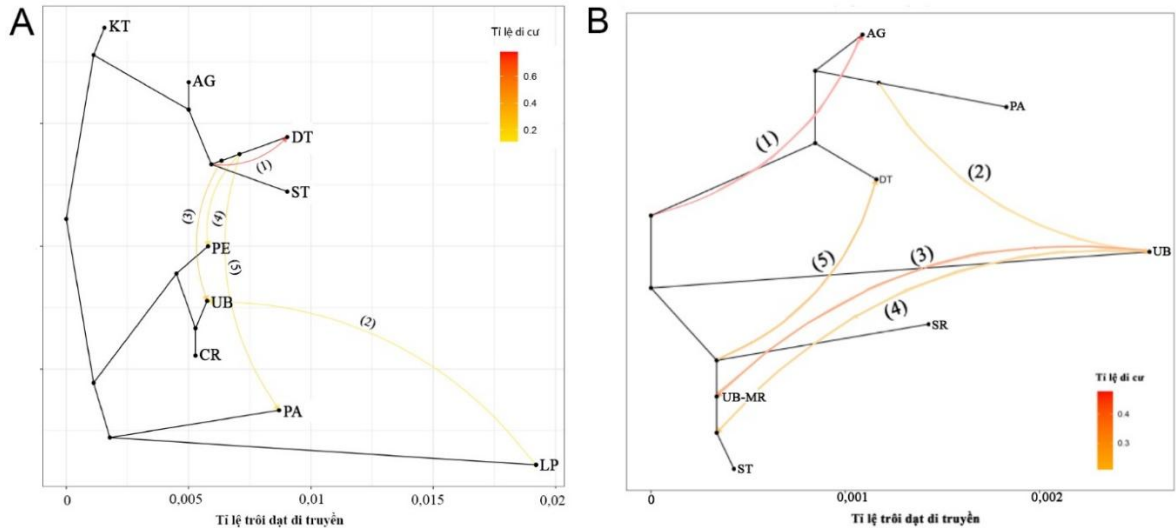
Hình 3.9. Mô hình di cư hiện tại của cá ét mọi (A) và cá vồ đém (B) ở hạ lưu sông Mekong. Mũi tên chỉ hướng di chuyển dòng gen và các con số biểu thị tỉ lệ di cư tương đối, độ đậm nhạt của mũi tên tương ứng với tỉ lệ di cư tương đối từ cao đến thấp

Đối với cá ét mọi, tại hệ thống di cư thượng lưu-LMB, quần thể LP đóng vai trò là nguồn phát tán di cư (xuôi dòng) với tỉ lệ thấp ($N_m < 0,5$), trong khi quần thể PA là điểm tiếp nhận nguồn gen (ngược dòng) từ tất cả các quần thể khác (tỉ lệ di cư tương đối $N_m = 0,2-0,5$) và quần thể PE vừa là điểm tiếp nhận vừa là điểm phát tán nguồn gen. Tại hệ thống di cư trung lưu-LMB, quần thể UB-MK là điểm tiếp nhận từ hầu hết các quần thể và có sự di cư ngược dòng lên PA ($N_m = 0,5$) và xuôi dòng đến ST ($N_m = 0,48$), trong khi đó, quần thể RE chủ yếu là điểm phát tán nguồn gen đến các quần thể PA, PE, UB-MK, ST, KT, AG và tiếp nhận nguồn gen từ LP và DT. Tại hệ thống di cư hạ lưu-LMB, quần thể ST thể hiện là điểm tiếp nhận nguồn gen của hầu hết các quần thể và có sự di cư ngược dòng đến RE. Quần thể KT là điểm tiếp nhận nguồn gen từ LP, RE, AG, DT và di cư ngược dòng đến các quần thể PA, PE, UB-MK, ST. Ở ĐBSCL, trong khi AG là điểm tiếp nhận nguồn gen từ 3 quần thể (LP, RE và DT) và phát tán đến 5 quần thể khác (PA, PE, UB-MK, ST, KT) thì DT chủ yếu phát tán nguồn gen đến hầu hết các quần thể, ngoại trừ LP. Đáng chú ý, nghiên cứu ghi nhận khả năng di cư ngược dòng vượt thác Khôn của cá ét mọi, thể hiện qua các đường di cư từ KT và ĐBSCL lên phía thượng nguồn (PA, PE và UB-MK) với tỉ lệ từ 0,35 đến 0,85 (**Hình 3.9A**).

Ở cá vồ đém, các quần thể PA (thượng lưu-LMB), SR (Biển Hồ) và UB-MR (trung lưu-LMB) đóng vai trò là nguồn phát tán gen đến tất cả các quần thể ở lưu vực. Tại hệ thống trung lưu-LMB, UB-MK là địa điểm tiếp nhận chính từ nhiều nguồn và có sự di cư xuôi dòng mạnh đến DT ($N_m = 0,77$), trong khi ST ở hệ thống di cư hạ lưu-LMB là điểm nhận gen và có xu hướng di cư ngược dòng về UB-MK ($N_m = 0,83$). Khác với cá ét mọi, quần thể DT ở cá vồ đém chủ yếu đóng vai trò là điểm nhận gen, còn AG duy trì sự trao đổi gen, vừa tiếp nhận từ 3 quần thể (PA, UB-MR và SR) và phát tán đến 3 quần thể khác (UB-MK, ST và DT). Đáng chú ý, sự di cư ngược dòng vượt thác Khôn ở cá vồ đém rất rõ nét với tỉ lệ cao, điển hình là các đường di cư từ AG và ST ngược lên UB-MK với tỉ lệ đạt từ 0,64 đến 0,83 và từ SR lên PA với tỉ lệ 0,49 (**Hình 3.9B**).

3.2.3.2. Lịch sử phân tách và pha trộn quần thể của các loài cá

Dựa trên phương pháp OptM, số sự kiện tối ưu được dự đoán của cả hai loài cá ét mọi và vô đém ở LMB được xác định là 5. Sơ đồ hình cây của các quần thể xây dựng bằng thuật toán hà tiện tối đa (ML) được thể hiện ở **Hình 3.10**.



Hình 3.10. Sơ đồ hình cây về lịch sử phân tách và pha trộn quần thể cá ét mọi (A) và vô đém (B) sử dụng Treemix. *Mũi tên chỉ hướng của dòng gen, màu sắc và độ dày của đường biểu thị tỉ lệ di cư*

Đối với cá ét mọi (**Hình 3.10A**), nghiên cứu ghi nhận sự kết hợp giữa các luồng di cư xuôi dòng và ngược dòng. Sự kiện di cư mạnh mẽ nhất diễn ra xuôi dòng từ điểm trung gian phía trên AG đến DT (1), với tỉ lệ di cư $w = 0,6$. Sự kiện di cư xuôi dòng thứ 2 được ghi nhận từ LP đến UB-MK ($w = 0,22$). Đặc biệt, mô hình xác định 3 sự kiện di cư ngược dòng từ điểm trung gian giữa DT – ST lên các quần thể phía trên thác Khôn, gồm UB-MK (3), PE (4) và PA (5), đều có cùng tỉ lệ $w = 0,2$.

Ở cá vô đém (**Hình 3.10B**), các sự kiện di cư ở loài này thể hiện mạng lưới kết nối phức tạp giữa dòng chính và các dòng nhánh. Sự kiện di cư chính (1) từ điểm trung gian trên thác (UB-MK và PA) và dưới thác Khôn (AG và DT) xuôi dòng về AG ($w = 0,48$). Tại khu vực thượng lưu, sự kiện di cư xuôi dòng từ PA đến UB-MK (2) được ghi nhận với $w = 0,21$. Đáng chú ý là các sự kiện di cư ngang và ngược dòng giữa dòng chính và dòng nhánh từ UB-MK đến UB-MR (3; $w = 0,4$) và xuôi dòng đến điểm trung gian của UB-MR và ST (4; $w = 0,24$). Cuối cùng, sự kiện di cư (5) từ

điểm trung gian trên thác Khôn (UB-MK và UB-MR) và dưới thác Khôn (SR và ST) đổ về DT ($w = 0,27$).

3.2.4. Thảo luận

❖ Sự tương tác giữa tập tính di cư và vòng đời phát triển trong việc hình thành biến thiên di truyền quần thể

Kết quả nghiên cứu hiện tại đã làm sáng tỏ mối tương quan giữa các đặc điểm sinh học (tập tính di cư và vòng đời phát triển) trong việc hình thành các biến thiên di truyền và khả năng duy trì dòng gen giữa các khu vực địa lý. Tập tính di cư đóng vai trò trao đổi vật chất di truyền, thể hiện qua mối tương quan thuận với mức độ đa dạng di truyền (**Bảng 3.4**) và tương quan nghịch với hệ số cận huyết (**Bảng 3.9**). Ở loài cá vồ đém, chiến lược di cư chằng dãi kết hợp với phạm vi di chuyển rộng lớn đã tạo điều kiện tối ưu cho dòng gen trao đổi liên tục, giúp duy trì mức đa dạng cao nhất ($H_o/H_e = 0,220/0,266$) và hệ số cận huyết thấp nhất ($F_{IS} = 0,173$). Ngược lại, sự hạn chế di chuyển ở loài không di cư như cá chạch lá tre đã tạo ra các rào cản sinh học dẫn đến hạn chế trao đổi gen, khiến mức độ đa dạng thấp nhất ($H_o/H_e = 0,140/0,176$) và hệ số cận huyết cao nhất ($F_{IS} = 0,40-0,61$). Đáng chú ý, tại dòng nhánh sông Mun (UB-MR), dù kích thước quần thể hiệu quả khá cao ($N_e = 454,3$), nhưng hệ số cận huyết vẫn đạt mức cao nhất ($F_{IS} = 0,61$) (**Bảng 3.9**). Điều này cho thấy nếu thiếu hụt dòng gen từ tập tính di cư, sự biến thiên di truyền sẽ bị thu hẹp đáng kể do tình trạng giao phối cận huyết và trôi dạt gen ngẫu nhiên. Kết quả này phù hợp với các mô hình tiến hóa về “hiệu ứng người sáng lập” và “trôi dạt gen”, nơi các alen có lợi dễ bị mất đi ngẫu nhiên trong các quần thể nhỏ và cô lập [233, 235].

Sự tương tác này còn được củng cố bởi các giai đoạn trong vòng đời phát triển, đặc biệt là quá trình phát tán thụ động của trứng và ấu trùng. Với giai đoạn phù du kéo dài (22 ngày ở cá chạch lá tre [63], 30 – 50 ngày ở cá ét mọi [68] và 33 – 35 ngày ở cá vồ đém [81]), dòng chảy sông Mekong đóng vai trò phát tán nguồn gen từ thượng lưu xuống hạ lưu. Sự kết hợp giữa di cư chủ động và phát tán thụ động giải thích cho xu hướng tăng dần đa dạng di truyền dọc theo dòng chính, được thể hiện rõ ràng ở cá chạch lá tre (**Bảng 3.4**). Xu hướng này cũng được quan sát thấy ở các loài cá khác

thuộc sông Mekong (như cá thát lát *N. notopterus* [236], cá rô đồng *A. testudineus* [20], cá linh *H. lobatus*) và các hệ thống sông khác (như cá bảy màu *P. reticulata* ở sông Marianne-Trinidad [237], cá nhác *S. macrops* ở sông Min-Trung Quốc [238] và cá may *G. cambodgiensis* ở sông Chao Phraya-Thái Lan) [239]. Sự phân hóa này còn chịu tác động bởi cấu trúc cảnh quan và đặc điểm thủy văn dọc sông Mekong [235, 240, 241]. Khu vực thượng lưu với địa hình phân mảnh, dòng chảy xiết tạo rào cản vật lý gây cô lập, trong khi hạ lưu với lòng sông mở rộng và môi trường đa dạng đã tạo điều kiện hội tụ và duy trì biến thiên di truyền cao [34, 241].

Hơn nữa, nghiên cứu ghi nhận các hợp lưu như (Mekong – 3S) và Kratié (Mekong – Biển Hồ) là những khu vực hội tụ nguồn gen với mức độ đa dạng di truyền cao hơn các quần thể khác (**Bảng 3.4**). Kết quả này tương đồng với các nghiên cứu trên loài cá lăng (*H. spilopterus*) tại hệ thống sông 3S và cá rô đồng (*A. testudineus*) tại Stung Treng, khẳng định vai trò của các dòng nhánh trong việc cung cấp và bổ sung biến dị di truyền cho dòng chính. Ngược lại, tại các dòng nhánh như sông Khan (LP), sông Mun (UB-MR) và sông Chi (RE), mức độ đa dạng di truyền và kích thước quần thể hiệu quả ghi nhận ở mức thấp và hệ số cận huyết cao (**Bảng 3.4 và 3.9**). Đây là tín hiệu báo động về sự suy giảm sức sống di truyền, tương tự như hiện trạng của cá chạch bùn *M. favius* ở Luang Prabang [158]. Điều này cho thấy cấu trúc cảnh quan sông đóng vai trò then chốt trong việc định hình đặc điểm di truyền và duy trì khả năng thích nghi của các loài cá nước ngọt [237, 242].

Mặc dù các quần thể cá sông Mekong trong nghiên cứu hiện tại vẫn duy trì mức độ đa dạng di truyền ở ngưỡng ổn định, nhưng các áp lực nhân tạo đang tạo ra những thách thức nghiêm trọng. Việc xây dựng hàng loạt đập thủy điện trên cả dòng chính và dòng nhánh không chỉ làm gián đoạn chu kỳ di cư của cá trưởng thành mà còn ảnh hưởng đến quá trình phát tán thụ động của trứng và ấu trùng xuống hạ lưu, dẫn đến hiện tượng phân mảnh quần thể và gia tăng cận huyết [243, 244]. Kết hợp với sự thay đổi nhịp điệu thủy văn và tác động từ hoạt động nuôi trồng thủy sản, các yếu tố này đang làm suy giảm kích thước quần thể hiệu quả và tính thích nghi của các quần thể bản địa [245, 246]. Hiện tượng suy giảm kích thước quần thể hiệu quả đã được ghi

nhận ở quần thể cá lăng *H. spilopterus* tại lưu vực sông Serepok [15] và cá phèn vàng *P. melanochir* ở ĐBSCL [16]. Bên cạnh đó, việc thả thoát cá nuôi vào môi trường tự nhiên cũng có thể làm suy giảm tính thích nghi của các quần thể bản địa, như trường hợp đã được ghi nhận ở cá tra *Pn. hypophthalmus* [18, 247].

Từ những kết quả này, nghiên cứu nhấn mạnh tầm quan trọng của việc bảo tồn các hành lang di cư nhằm duy trì sự kết nối giữa các quần thể, bảo vệ các khu vực có đa dạng di truyền cao và quản lý nghiêm ngặt các hoạt động khai thác cũng như nuôi trồng thủy sản là vô cùng cần thiết. Việc thực thi đồng bộ các giải pháp này không chỉ đảm bảo khả năng thích ứng dài hạn mà còn duy trì giá trị đa dạng sinh học đặc hữu của khu hệ cá sông Mekong trước bối cảnh biến đổi khí hậu và áp lực phát triển kinh tế mạnh mẽ như hiện nay.

❖ Tác động tích lũy của rào cản cảnh quan và biến đổi môi trường lên cấu trúc di truyền của các loài cá

Kết quả nghiên cứu về sự khác biệt di truyền (F_{ST}) và cấu trúc quần thể (STRUCTURE, PCA và AMOVA) đã cung cấp bằng chứng thực nghiệm cho thấy cấu trúc di truyền hiện nay phản ánh những tác động tích lũy từ các rào cản cảnh quan và biến đổi môi trường tại LMB. Sự phân hóa quần thể rõ rệt nhất được ghi nhận ở loài không di cư (cá chạch lá tre - 4 nhóm) (**Hình 3.4**) và di cư chặng ngắn, tùy nghi (cá ét mọi - 3 nhóm) (**Hình 3.5**), tương ứng với các khu vực sinh thái thượng, trung và hạ lưu. Điều này minh chứng cho sự thích nghi cục bộ của chúng trước các điều kiện thủy văn và địa hình đặc trưng [34]. Trong đó, thác Khôn đóng vai trò là rào cản tự nhiên, gây ra sự sai khác di truyền rõ rệt giữa nhóm quần thể phía trên và dưới thác Khôn của cá chạch lá tre (**Hình 3.3A**) và sự phân tách của 2 nhóm quần thể tương ứng này ở cá chạch lá tre và cá ét mọi (**Hình 3.4 và 3.5**). Kết quả tương tự cũng được ghi nhận ở nhóm cá không di cư (cá thát lát *N. notopterus* [236], cá rô đồng *A. testudineus* [20], cá lóc *C. striata* [159] và cá chạch lấu *M. favus* [157, 158]) và nhóm di cư chặng ngắn (cá trôi *C. molitorella* [248]). Ngay cả với các loài di cư chặng dài như cá linh *H. lobatus* và cá xác *H. leptorhynchus*, tác động của thác Khôn cũng được ghi nhận trong việc định hình cấu trúc quần thể [17].

Tập tính di cư đóng vai trò quyết định mức độ phân hóa di truyền, thể hiện qua việc tỉ lệ sai khác giữa các quần thể (AMOVA) giảm dần từ loài không di cư (30%) (**Bảng 3.6**) xuống loài di cư chặng ngắn và tùy nghi (2-6%) (**Bảng 3.7**) và đạt mức không đáng kể ở loài di cư đường dài (<1%) (**Bảng 3.8**). Kết quả phân tích IBD càng củng cố thêm nhận định này khi cho thấy mối tương quan thuận giữa khoảng cách địa lý và sự khác biệt di truyền ở các loài không di cư và di cư chặng ngắn (**Hình 3.7A và B**), ngược lại, khả năng di chuyển xa đã giúp các loài di cư đường dài như cá vô đém vượt qua tác động của rào cản địa lý (**Hình 3.7C**). Kết quả này cũng phù hợp với các nghiên cứu trước đây về cá sông Mekong sử dụng chỉ thị DNA ti thể hoặc nhân [12, 19, 157, 165, 248]. Mặc dù thời gian phát tán trứng và ấu trùng của ba loài cá được biết là tương đương nhau (22 ngày ở cá chạch lá tre, 30 – 50 ngày ở cá ét mọi và 33 – 35 ngày ở cá vô đém), các yếu tố như tuổi trưởng thành, kích thước trưởng thành, thời gian thế hệ, thời gian nở, kích thước trứng và ấu trùng, ... có thể ảnh hưởng đến khả năng và khoảng cách phát tán của mỗi loài [249]. Các nghiên cứu trên cá biển cũng đã chỉ ra mối liên hệ giữa các yếu tố phát tán và mô hình IBD [282], tuy nhiên, mối liên hệ này cần được nghiên cứu thêm trên đối tượng cá nước ngọt.

Ngoài khoảng cách sông, phân tích dbMEM (**Hình 3.8**) đã chỉ ra rằng cấu trúc không gian phức tạp của môi trường sống, bao gồm sự phân nhánh và các rào cản vật lý, có ảnh hưởng đến sự cấu trúc di truyền mà các kiểm định Mantel có thể bỏ qua. Các nghiên cứu trước đây sử dụng cả hai phương pháp này đã cho thấy kết quả đa dạng, có thể tương đồng [250, 251] hoặc mâu thuẫn [252]. Điều này cho thấy việc chỉ dựa vào khoảng cách địa lý là không đủ để giải thích sự khác biệt di truyền giữa các quần thể cá. Cần phải xem xét đến cả cấu trúc không gian của môi trường sống để có cái nhìn đầy đủ hơn về sự phân bố và di truyền của các loài.

Bên cạnh yếu tố tự nhiên, các thay đổi môi trường nhân tạo đang gây ra những ảnh hưởng tiêu cực đến tính liên tục sinh thái và dòng gen của cá. Nghiên cứu ghi nhận sự tách biệt di truyền của quần thể cá chạch lá tre tại nhánh sông Mun (UB-MR) (**Hình 3.4**) là hệ quả của việc hạn chế trao đổi gen giữa dòng chính và dòng nhánh, khẳng định tầm quan trọng của tính liên tục sinh thái đối với cả những loài không di

cư. Tác động của các rào cản nhân tạo được thể hiện rõ qua trường hợp đập Pak Mun (cao 17m) trên sông Mun, cách dòng chính sông Mekong 5,5km và vận hành đã làm suy giảm nghiêm trọng sản lượng và thành phần loài ở thượng nguồn [253]. Để giảm thiểu tác động tiêu cực, Chính phủ Thái Lan đã quyết định mở cửa đập 4 tháng/năm (tháng 6 – tháng 9) nhằm tạo điều kiện cho sự trao đổi gen của các loài cá [254]. Mặc dù các nghiên cứu hiện tại chưa có dẫn liệu cụ thể về ảnh hưởng của đập thủy điện đến các quần thể cá nhưng những bằng chứng về sự cô lập quần thể cá lăng (*H. spilopterus*) tại sông Sesan do hệ thống đập thượng nguồn là lời cảnh báo về những tác động tiêu cực có thể tích lũy và xảy ra trong tương lai gần. Đáng chú ý, dựa trên các phân tích sử dụng các loci đáp ứng, mặc dù sự thích ứng của các quần thể cá với những thay đổi của môi trường chưa thể hiện rõ ràng, tuy nhiên, nghiên cứu hiện tại ghi nhận sự kết nối rộng rãi của các quần thể cá chạch lá tre ở phía trên thác Khôn và không có sự khác biệt nào ở dòng nhánh sông Mun (quần thể UB-MR) (**Hình 2 - Phụ lục 12**). Điều này cho thấy sự khác biệt về điều kiện sinh thái ở phía trên và dưới thác Khôn, đặc biệt là sự phân bố của các vùng ngập, tạo nên những thích nghi của các loài sinh vật ít di chuyển. Mặc dù, vùng ngập phân bố dọc theo hầu hết chiều dài ở LMB, nhưng ở phía dưới thác Khôn, với địa hình bằng phẳng và chu kỳ ngập lũ hằng năm của hệ thống Biển Hồ, tạo điều kiện cho vùng ngập ở khu vực này rộng lớn hơn so với khu vực trên thác Khôn.

Tuy nhiên, các kết quả STRUCTURE cũng chỉ ra sự sẻ chia thông tin di truyền dọc sông Mekong nhờ quá trình phát tán thụ động của ấu trùng và di cư xuôi dòng của cá trưởng thành sau mùa sinh sản. Bằng chứng rõ ràng nhất là sự pha trộn thông tin di truyền giữa khu vực thượng lưu, trung và hạ lưu ở LMB, cũng như sự trộn lẫn thông tin di truyền ở các quần thể hợp lưu (Stung Treng và Kratié). Cụ thể, quần thể cá chạch lá tre ở ST mang thông tin di truyền có nguồn gốc từ các quần thể phía trên thác Khôn (màu hồng), sông Mun (màu vàng) và phía dưới thác Khôn (màu xanh dương và xanh lá); quần thể Kratié có thành phần di truyền chủ yếu là từ ST và ĐBSCL, và sự kết nối di truyền của hai quần thể ở ĐBSCL - An Giang và Đồng Tháp (**Hình 3.4A**). Ở cá ét mọi, vật liệu di truyền có nguồn gốc từ thượng lưu được thể

hiện ở trung và hạ lưu (màu đỏ, **Hình 3.5A**), nhiều cá thể ở hạ lưu mang thông tin di truyền từ trung lưu (màu xanh dương, **Hình 3.5A**). Ở cá vô đốm, sự kết nối quần thể rộng rãi được thể hiện ở trung lưu và hạ lưu (**Hình 3.6A**). Qua đây có thể thấy các quá trình sinh học như phát tán ấu trùng và di cư đóng vai trò quan trọng trong việc duy trì sự đa dạng di truyền và liên kết quần thể của các loài cá trong khu vực. Vì vậy, để đảm bảo sự kết nối di truyền và sự tồn tại của các loài cá, việc duy trì dòng chảy tự nhiên của sông Mekong và hạn chế tối đa việc xây dựng các đập thủy điện là điều cần thiết.

Việc kết hợp phân tích mô hình di cư lịch sử (TreeMix) và dòng gen hiện tại (divMigrate) đã góp phần gợi mở những thích nghi và xung đột trong tập tính của các loài cá trước biến động của môi trường sông Mekong. Kết quả từ divMigrate cho thấy cả cá ét mọi và cá vô đốm đều thể hiện khả năng di cư, gồm di cư xuôi dòng và ngược dòng, trong đó cá vô đốm thể hiện xu hướng di chuyển mạnh mẽ hơn (**Hình 3.9**). Điều này hoàn toàn phù hợp với tập tính di cư của mỗi loài, khi cá vô đốm được biết đến là loài di cư đường dài, còn cá ét mọi thuộc nhóm di cư chặng ngắn và tùy nghi.

Đáng chú ý, kết quả phân tích gợi mở giả thuyết về khả năng di cư ngược dòng đường dài lên tới 1.200 km của cá ét mọi, từ khu vực ĐBSCL lên đến Paksan (Lào) (**Hình 3.9A**), phù hợp với dự đoán của Mashyaka và Duong (2021) sử dụng chỉ thị phân tử microsatellite [70]. Việc ghi nhận di cư chặng dài ở cá ét mọi đã phần nào hỗ trợ cho những tranh cãi trước đây về việc loài này di cư chặng dài thay vì chặng ngắn và tùy nghi. Tuy nhiên, đây chỉ là những suy luận dựa trên dữ liệu hiện tại, và cần có những dẫn liệu thực tế hơn để kiểm chứng giả thuyết này. Đồng thời, nghiên cứu không ghi nhận việc di cư ngược dòng lên đến Luang Prabang ở cá ét mọi (**Hình 3.9A**), điều này góp phần giải thích cho sự khác biệt di truyền và phân tách của quần thể này (**Hình 3.3B và Hình 3.5**). Đối với cá vô đốm, khoảng cách di cư ngược dòng xa nhất ghi nhận được là hơn 1.000 km, từ Siem Riệp đến Paksan (**Hình 3.9B**), tương tự như các loài cá trong họ Pangasiidae đã được ghi nhận bằng phương pháp phân tích thành phần hóa học trong đá tai [136]. Tuy nhiên, mặc dù được ghi nhận là loài di cư chặng dài, kết quả nghiên cứu lại chỉ ra xu hướng không liên tục trong tuyến di

cư của cá vô đém, chỉ giới hạn giữa hạ lưu-AG và trung lưu – UB-MK (1.000km). Do nghiên cứu hiện tại không thu thập được quần thể cá vô đém ở LP, nghiên cứu chưa thể cung cấp dữ liệu để dự đoán liệu tuyến di cư của loài này có kéo dài lên đến khu vực Thượng lưu hay không. Các nghiên cứu tiếp theo, bao gồm mở rộng khu vực nghiên cứu, theo dõi di cư trực tiếp và phân tích di truyền sâu hơn, là cần thiết để làm sáng tỏ mô hình di cư thực sự của loài cá này.

Bên cạnh đó, nghiên cứu ghi nhận sự thay đổi giữa mô hình lịch sử và hiện tại của loài cá vô đém tại hệ thống sông Mun-Chi thuộc lưu vực Korat. Mô hình TreeMix chỉ ra rằng trong lịch sử, cá vô đém có xu hướng di cư ngược dòng từ dòng chính sông Mekong vào sông Mun để sinh sản (**Hình 3.10B**). Ngược lại, mô hình divMigrate hiện tại lại cho thấy cả cá ét mọi và cá vô đém chủ yếu thực hiện hành vi di cư xuôi dòng từ sông Mun và Chi ra dòng chính (**Hình 3.10A và B**). Sự thay đổi này có thể được giải thích bởi tác động của đập Pak Mun. Trước khi có đập, cá vô đém có thể tự do di chuyển vào sông Mun, nhưng rào cản nhân tạo cao 17m này có thể đã ngăn chặn tuyến đường ngược dòng và chiến lược di cư xuôi dòng thực hiện bởi cá trưởng thành để tìm kiếm thức ăn hoặc phát tán thụ động trứng và ấu trùng.

Tầm quan trọng của các tuyến đường di cư vượt thác Khôn cũng được khẳng định qua việc cả cá vô đém và cá ét mọi đều thực hiện các cuộc di cư ngược dòng từ hạ lưu lên trung lưu vượt qua khu vực này để sinh sản. Những vùng ngập ngay trên và dưới thác Khôn được dự đoán là bãi đẻ quan trọng. Phát hiện này tương đồng với các thông tin khảo sát kiến thức địa phương của ngư dân [7, 255]. Việc hoàn thành đập Don Sahong trên kênh Hou Sahong – con đường di cư duy nhất vượt thác Khôn – đang gây ra những lo ngại sâu sắc về việc ngăn chặn tuyến đường di cư này, đe dọa trực tiếp đến sự duy trì đa dạng di truyền của các quần thể cá [17].

Tóm lại, nghiên cứu về cấu trúc quần thể của các loài cá ở hạ lưu lưu vực sông Mekong đã chỉ ra sự khác biệt rõ rệt về mức độ phân hóa di truyền giữa các nhóm có tập tính sinh học khác nhau. Sự định hình cấu trúc di truyền này là kết quả của các tác động tích lũy từ tập tính di cư, khoảng cách địa lý, rào cản tự nhiên (như Thác Khôn) cho đến các áp lực nhân tạo (như hệ thống đập thủy điện). Sự đan xen và xung

đột trong mô hình dòng gen giữa loài di cư chặng ngắn/tùy nghi và di cư chặng dài đã phản ánh tính đa dạng, phức tạp của khu hệ sinh thái sông Mekong. Những phát hiện này khẳng định vai trò then chốt của di truyền học cảnh quan trong việc đánh giá các động lực tiến hóa và khả năng thích ứng cục bộ của sinh vật nước ngọt.

Mặc dù nghiên cứu hiện tại cung cấp những dẫn liệu quan trọng về mối tương quan giữa tập tính sinh học, đặc điểm vùng sinh thái và cấu trúc di truyền quần thể của các loài cá, sự chênh lệch về cỡ mẫu và số lượng quần thể giữa 3 loài cá (239 cá thể từ 9 quần thể cá chạch lá tre; 232 cá thể từ 9 quần thể cá ét mọi và 160 cá thể từ 7 quần thể cá vồ đém) cũng phần nào đặt ra giới hạn đối với các phép so sánh liên loài. Cụ thể, tại Biển Hồ - hồ chứa nước lũ tự nhiên vận hành đồng thời là nơi phát tán nguồn gen và tiếp nhận các quần thể cá di cư trong mùa sinh sản, việc thiếu mẫu cá chạch lá tre và cá ét mọi khiến nghiên cứu chưa thể làm sáng tỏ vai trò trao đổi di truyền đối với hai loài này cũng như việc đánh giá các hệ thống di cư ngang. Tương tự, tại khu vực dòng nhánh sông Khan (quần thể Luang Prabang), việc thiếu dữ liệu của cá vồ đém khiến tuyến đường di cư lý thuyết của loài này lên vùng thượng nguồn phía trên thác Khôn bị ngắt quãng.

Nhằm hoàn thiện bức tranh di truyền học cảnh quan cá sông Mekong, các nghiên cứu tiếp theo cần mở rộng phạm vi thu thập dữ liệu toàn diện trong lưu vực sông Mekong. Đồng thời, việc ứng dụng các chỉ thị phân tử thế hệ mới có độ phân giải cao hơn cần được kết hợp chặt chẽ với các công nghệ định vị sinh học trực tiếp – như gắn thẻ điện tử hoặc phân tích đồng vị bền đá tai – nhằm theo dõi liên tục quỹ đạo di chuyển của cá theo thời gian thực. Về mặt sinh thái học hành vi và môi trường, cần triển khai các nghiên cứu tích hợp giữa quan sát tự nhiên và thực nghiệm nhân tạo để làm sáng tỏ các yếu tố nội sinh ảnh hưởng đến tập tính di cư. Đặc biệt, việc đánh giá các tác động tích lũy từ áp lực nhân sinh như ô nhiễm môi trường, biến đổi khí hậu và xu hướng phân mảnh quần thể do hệ thống đập thủy điện là vô cùng cấp bách, nhằm cung cấp cái nhìn toàn diện và thực chất hơn cho các chiến lược bảo tồn đa dạng sinh học xuyên biên giới.

❖ ***Định hướng quản lý và bảo tồn nguồn lợi ở sông Mekong***

Các phát hiện về cấu trúc quần thể và mô hình di cư trong nghiên cứu này mang lại những bằng chứng khoa học cho việc định hình các chiến lược bảo tồn đa dạng di truyền và quản lý nghề cá bền vững. Đối với khu hệ cá sông Mekong, nếu các chính sách quản lý không phù hợp sẽ dễ dẫn đến hiện tượng khai thác quá mức, làm suy giảm đa dạng thích nghi và đẩy nhanh quá trình suy thoái nguồn lợi [104]. Dựa trên các thông tin di truyền cụ thể của từng loài, các định hướng quản lý phân hóa theo vùng sinh thái được đề xuất như sau:

Đối với khu vực vùng sinh thái trên và dưới Thác Khôn: Kết quả phân tích cấu trúc di truyền đã chỉ ra thác Khôn vận hành như một rào cản sinh thái nghiêm ngặt đối với các loài có tập tính di chuyển hạn chế, nhưng lại là hành lang mở đối với loài di cư chằng dài. Do đó, mô hình bảo tồn tại đây cần được thiết kế riêng biệt dựa trên tập tính sinh học của loài. Ở loài không di cư như cá chạch lá tre, sự phân biệt cấu trúc di truyền rõ rệt giữa hai phía thác Khôn kết hợp với kích thước quần thể hiệu quả nhỏ ở dòng chính thượng nguồn ($N_e = 109-292$) là cơ sở để công nhận các Đơn vị Tiến hóa Quan trọng (*Evolutionary Significant Units*) độc lập [256]. Khuyến nghị cơ quan quản lý không được di dời hoặc thả giống tái tạo liên vùng giữa trên và dưới thác nhằm bảo tồn nguyên vẹn tính thích nghi cục bộ, đồng thời ưu tiên thiết lập các khu bảo tồn nguồn gen nghiêm ngặt tại chỗ tại vùng Thượng lưu – nơi có áp lực quần thể nhỏ – để tránh hiện tượng trôi dạt gen ngẫu nhiên. Ngược lại, đối với cá ét mọi và cá vô đém, các mô hình phân tích đều chỉ ra dòng gen xuyên biên giới vẫn được duy trì và có khả năng vượt qua Thác Khôn, góp phần duy trì chỉ số dị hợp mong đợi ở mức cao toàn lưu vực ($H_e = 0,233$ và $0,266$). Do đó, vùng sinh thái này cần được quản lý như một Đơn vị Quản lý chung (*Management Unit*) xuyên quốc gia. Cần lưu ý rằng, sự xuất hiện của các dự án đập thủy điện dòng chính (như Don Sahong) đã và đang phá vỡ tính liên thông sinh thái này, biến tuyến đường di cư chằng dài vốn liên tục của cá vô đém trở nên bị ngắt quãng về mặt không gian, đe dọa trực tiếp đến tính toàn vẹn di truyền của toàn lưu vực.

Khu vực các dòng nhánh chịu tác động bởi đập thủy điện: Sự phân hóa được ghi nhận tại các dòng nhánh (sông Mun, sông Chi, sông Khan) là minh chứng cho hệ quả của sự phân mảnh sinh cảnh nhân tạo lên cấu trúc di truyền quần thể của các quần thể hoang dã. Tại hệ thống sông Mun-Chi (thuộc Cao nguyên Korat), mặc dù kích thước quần thể hiệu quả của cá chạch lá tre khá cao ($N_e = 454,3$), hệ số cận huyết của loài này lại thấp nhất ($F_{IS} = 0,61$). Đồng thời, mô hình lịch sử TreeMix dự đoán cá vồ đém từng di cư ngược dòng vào sông Mun để sinh sản, nhưng mô hình hiện tại tại divMigrate cho thấy dòng gen bị đảo ngược thành xu hướng một chiều ra dòng chính. Bằng chứng di truyền này phản ánh tác động của đập Pak Mun trong việc hạn chế tuyến đường di cư sinh sản lịch sử, dẫn đến xu hướng gia tăng hiện tượng cận huyết cục bộ trong quần thể. Nhằm duy trì và bảo tồn nguồn gen tại vùng sinh thái này, cần thiết phải áp dụng các biện pháp giảm thiểu tác động (chẳng hạn như vận hành mở cửa đập phù hợp vào mùa sinh sản) để đảm bảo hành lang di cư của cá. Sự kết nối di truyền liên tục này đóng vai trò sống còn trong việc duy trì mức độ đa dạng sinh học và tăng cường khả năng phục hồi của các quần thể trước những áp lực môi trường và khai thác ngày càng gia tăng. Riêng tại dòng nhánh sông Khan (Luang Prabang), kết quả ghi nhận một cấu trúc di truyền biệt lập, tương thích chặt chẽ với đặc trưng sinh thái độc đáo của vùng thượng nguồn. Do đó, đây là cơ sở khoa học vững chắc để đề xuất thiết lập các Đơn vị Tiến hóa Quan trọng (ESUs) riêng biệt nhằm bảo tồn nguyên vẹn nguồn gen bản địa thích nghi đặc thù.

Quản lý nghề cá đa loài xuyên biên giới và thích ứng môi trường: Chiến lược quản lý nghề cá sinh thái đa loài mang tính liên vùng cần được thiết lập nhằm kết nối hành lang bảo tồn giữa các “nguồn” phát tán ấu trùng ở vùng Trung lưu (như hệ thống sông 3S, khu vực trên Thác Khôn) và các “đích” tiếp nhận ở vùng Hạ lưu (Tonlé Sap, ĐBSCL). Song song đó, trước xu hướng suy giảm kích thước quần thể hiệu quả của các loài nghiên cứu do phân mảnh thủy văn, mọi quy hoạch phát triển kinh tế lưu vực cần đặt tính liên thông sinh thái và dòng chảy tự nhiên lên trên các giải pháp công trình kỹ thuật. Đây là điều kiện nhằm duy trì dòng gen liên tục, bảo lưu tính toàn vẹn

di truyền và đảm bảo tiềm năng tiến hóa dài hạn của nguồn lợi thủy sản trước bối cảnh áp lực khai thác và biến đổi khí hậu ngày càng gia tăng [257–260].

3.3. Lắp ráp và chú giải hệ gen ti thể, xác định mã vạch RAD và khảo sát cấu trúc di truyền quần thể cá ét mọi *Labeo chrysophekadion*

3.3.1. Lắp ráp và chú giải hệ gen ti thể của cá ét mọi

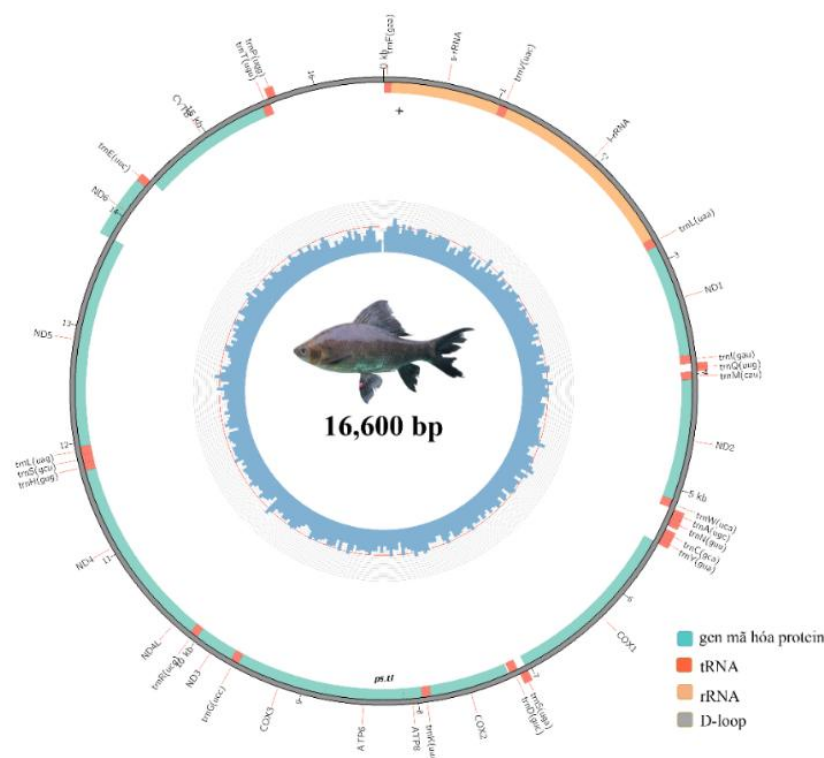
Tổng cộng có 1.026.721.048 đoạn đọc chất lượng cao (912 – 36.260 bp) từ 255 thư viện cá ét mọi được sử dụng để lắp ráp hệ gen ti thể. Tỷ lệ đoạn đọc được sắp xếp thành công vào hệ gen ti thể dao động từ 0,06 – 0,12%. Chiều dài trình tự đồng thuận của 255 cá thể dao động từ 436 - 16.600 bp. Trong đó, 08 cá thể có chiều dài trình tự ngắn (436 – 5.973 bp) và tỷ lệ sắp xếp dưới 50% (2,62 – 35,93%) đã bị loại. Cuối cùng, 247 trình tự được sử dụng cho các bước phân tích tiếp theo.

Trình tự của cá thể thu thập ở Thái Lan (UB-MK503) có độ tương đồng cao nhất (99,8%) với trình tự cùng loài trên Ngân hàng Gen (**Bảng 3.10 và Phụ lục 13**). được lựa chọn để mã hóa và chú giải hệ gen ti thể của cá ét mọi.

Bảng 3.10. Sự tương đồng trình tự hệ gen ti thể của UB-MK503 với dữ liệu từ Ngân hàng Gen

STT	Loài tương đồng	Chiều dài (bp)	Độ bao phủ	% tương đồng	Địa điểm thu mẫu	Mã số Genbank
1	<i>L. chrysophekadion</i>	16.602	100	99,8	Kandal, Campuchia	AP011199
2	<i>L. calbasu</i>	16.607	100	95,1	Châu Á (cá nuôi)	AP012143
3	<i>L. fimbriatus</i>	16.614	99	93,5	Ấn Độ	KP025676
4	<i>L. gonius</i>	16.614	99	93,4	Ấn Độ	KT001152
5	<i>L. rohita</i>	16.611	99	93,4	Ấn Độ	JN412817

Hệ gen ti thể của mẫu UB-MK503 (mã số OR637878) có chiều dài 16.600 bp, với tỉ lệ GC là 42,9% (A 32,5%, T 24,6%, G 15,5% và C 27,4%). Hệ gen được chú giải gồm 37 gen, với 13 gen mã hóa protein (PCGs), 22 gen RNA vận chuyển (tRNA), 02 gen RNA ribosome (rRNA) và 01 vùng không mã hóa (D-loop). Trong tổng số 37 gen, 28 gen được mã hóa trên chuỗi nặng (heavy-H strand), trong khi đó 1 gen mã hóa protein (*ND6*) và 8 gen RNA vận chuyển (*tRNA^{Ala}*, *tRNA^{Asn}*, *tRNA^{Cys}*, *tRNA^{Gln}*, *tRNA^{Glu}*, *tRNA^{Pro}*, *tRNA^{Ser(UCN)}* và *tRNA^{Tir}*) mã hóa trên chuỗi nhẹ (light-L strand). Trình tự sắp xếp các gen dạng mạch vòng được thể hiện ở **Hình 3.12**, vị trí và đặc điểm của từng gen trong hệ gen ti thể được trình bày ở **Bảng 1 - Phụ lục 13**. Đối với các gen mã hóa protein, 12/13 gen bắt đầu bằng codon ATG, riêng gen cytochrome coxidase subunit I - *COI* bắt đầu với codon GTG. Có ba loại codon kết thúc được ghi nhận, gồm TAA ở 6 gen (*ND1*, *COI-III*, *ATP6-8*, *ND4L*, *ND5* và *cytb*, TAG ở *ND2*) và codon không hoàn toàn (T--) ở *ND3-4* và *ND6*.



Hình 3.11. Hệ gen ti thể của cá ét mội *Labeo chrysophekadion*

Các gen hiển thị bên trong vòng tròn được mã hóa trên chuỗi nặng, trong khi các gen bên ngoài được mã hóa trên chuỗi nhẹ

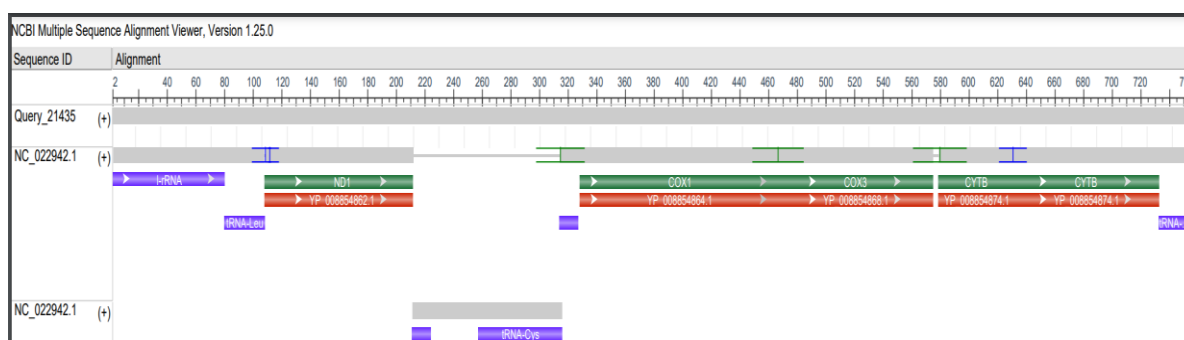
Cá ét mọi là loài có giá trị kinh tế quan trọng, phân bố rộng khắp lưu vực sông Mekong. Tuy nhiên, thông tin di truyền về loài cá này còn hạn chế, trước đây chỉ có một hệ gen ti thể được giải trình tự từ một cá thể ở Kandal, Campuchia, sử dụng phương pháp Sanger truyền thống. Nghiên cứu mới này đã bổ sung thêm một hệ gen ti thể của cá ét mọi từ Ubon Ratchathani (UB-MK), Thái Lan, đồng thời, cung cấp thông tin chi tiết về vị trí và đặc điểm gen so với nghiên cứu trước đó.

Điểm nổi bật của nghiên cứu là việc áp dụng kỹ thuật RAD-seq để lắp ráp hệ gen ti thể. RAD-seq mang lại nhiều ưu điểm so với phương pháp Sanger truyền thống và giải trình tự toàn bộ hệ gen (WGS), bao gồm hiệu quả kinh tế cao, tốc độ xử lý nhanh và độ phủ sâu ở vùng gen mục tiêu, đảm bảo độ chính xác của hệ gen ti thể [78, 79]. Việc sử dụng RAD-seq mở ra tiềm năng nghiên cứu hệ gen ti thể và di truyền quần thể cho nhiều loài thủy sản khác ở lưu vực sông Mekong, hỗ trợ công tác bảo tồn và quản lý nguồn lợi thủy sản bền vững. Hệ gen ti thể cá ét mọi từ UB-MK được ghi nhận gồm 37 gen (13 gen mã hóa protein, 2 gen rRNA và 22 gen tRNA) cùng với vùng D-loop, với trật tự sắp xếp gen hoàn toàn phù hợp với các nghiên cứu trước đây về họ cá chép Cyprinidae [261–264]. Khi so sánh với trình tự hệ gen ti thể đã được công bố (mã số AP011199), hệ gen ti thể cá ét mọi trong nghiên cứu này có kích thước ngắn hơn 2 nucleotide (C và A), tương ứng với vị trí thứ 16.601 và 16.602 ở vùng D-loop. Sự khác biệt về kích thước này là một đặc điểm thường thấy giữa các cá thể và các loài, được giải thích là do sự thay đổi kích thước vùng điều khiển và vùng lặp [265]. Dữ liệu này cũng được củng cố thêm bởi sự dao động kích thước hệ gen ti thể trong các loài cùng giống *Labeo*, với 16.596 bp ở *L. calbasu* (mã số AP012143) đến 16.766 bp ở *L. pierreii* (mã số NC_022943) (Dữ liệu được cập nhật trên Ngân hàng Gen vào tháng 01/2025). Việc xác định chi tiết vị trí và đặc điểm gen, cùng với sự khác biệt di truyền cấp độ nucleotide này, cung cấp thông tin quan trọng cho công tác bảo tồn, nghiên cứu đa dạng di truyền, cấu trúc quần thể, mối quan hệ tiến hóa và xác định loài, đặc biệt là các loài có giá trị kinh tế tại lưu vực sông Mekong.

3.3.2. Xác định mã vạch RAD thuộc hệ gen ti thể và khảo sát di truyền quần thể cá ét mọi ở LMB

3.3.2.1. Xác định mã vạch RAD thuộc hệ gen ti thể

Quy trình radBARCODER xác định được tập hợp các vùng mã vạch RAD có kích thước là 757 bp. Trong đó, 502 bp (66,2%) được phát hiện trong các PCGs (*ND1*, *COI*, *COIII* và *cytb*), 79 bp thuộc gen *16S rRNA* và phần còn lại thuộc các gen tRNA (*tRNA^{Leu}*, *tRNA^{Asn}*, *tRNA^{Cys}*, *tRNA^{Tir}* và *tRNA^{Thr}*) (Hình 3.12).



Hình 3.12. Giao diện đồ họa thể hiện vị trí của vùng mã vạch RAD trên hệ gen ti thể sử dụng tính năng BLAST (màu xanh dương thuộc gen rRNA và tRNA; màu xanh lá thuộc PCGs và màu đỏ là các acid amin tương ứng)

3.3.2.2. Mức độ đa dạng di truyền của các quần thể cá ét mọi ở LMB

Phân tích dữ liệu tập hợp mã vạch RAD (757 bp) cho thấy 9 quần thể cá ét mọi ở LMB có mức độ đa dạng haplotype cao ($H_d = 0,849 \pm 0,014$), với tổng số 49 haplotype (chiếm 19,8%) được ghi nhận. Số lượng haplotype ở mỗi quần thể dao động từ 6 haplotype/22 cá thể (27,3%) ở LP đến 12 haplotype/30 cá thể (40%) ở RE. Sự đa dạng haplotype dao động từ $0,71 \pm 0,071$ (LP) đến $0,871 \pm 0,046$ (UB-MK). Trong khi đó, sự đa dạng nucleotide ($0,005 \pm 0,0008$) và số lượng các vị trí đa hình (113 vị trí) thu được tương đối thấp. Về sự đa dạng nucleotide và số lượng các vị trí đa hình ở từng quần thể, Luang Prabang có giá trị cao nhất ($0,016 \pm 0,002$ và 45 vị trí), trong khi Đồng Tháp và Stung Treng có giá trị thấp nhất ($0,002 \pm 0,001$ và 9-10 vị trí). Nghiên cứu cũng ghi nhận sự khác biệt không đáng kể về đa dạng haplotype giữa các quần thể phân bố ở dòng chính, dòng nhánh và hợp lưu của sông Mekong.

Đồng thời, đa dạng nucleotide thấp được ghi nhận ở hai quần thể An Giang và Đồng Tháp thuộc ĐBSCL, và Stung Treng - hợp lưu Mekong và 3S (**Bảng 3.11**).

Bảng 3.11. Đa dạng di truyền của các quần thể cá ét mọi ở hạ lưu sông Mekong dựa trên dữ liệu mã vạch RAD

Vị trí trên sông Mekong	Quần thể	Nse	N _h	S	Hd	π
Dòng chính	PA	31	10	37	0,738±0,07	0,005±0,002
	PE	30	10	30	0,805±0,06	0,005±0,002
	UB-MK	26	11	19	0,871±0,05	0,004±0,002
	KT	27	10	55	0,835±0,05	0,009±0,003
	AG	24	10	16	0,822±0,06	0,004±0,002
	DT	29	10	10	0,842±0,05	0,002±0,001
Hợp lưu Mekong – 3S	ST	28	9	9	0,833±0,05	0,002±0,001
Dòng nhánh – sông Khan	LP	22	6	45	0,71±0,07	0,016±0,002
Dòng nhánh – sông Chi	RE	30	12	54	0,807±0,06	0,007±0,003
Tổng cộng		247	49	113	0,849±0,01	0,005±0,0008

Ghi chú: Nse: số trình tự mã vạch RAD, N_h: số haplotypes; S: số vị trí đa hình, Hd: đa dạng haplotype, π : đa dạng nucleotide, số liệu được trình bày dưới dạng giá trị trung bình ± độ lệch chuẩn.

3.3.2.3. Cấu trúc quần thể của cá ét mọi ở LMB

Sự khác biệt di truyền (F_{ST}) giữa các quần thể cá ét mọi ở LMB được ghi nhận là tương đối thấp. Trong đó, quần thể Luang Prabang thể hiện sự khác biệt di truyền có ý nghĩa thống kê so với các quần thể khác, với giá trị F_{ST} dao động từ 0,08 (LP/RE) đến 0,17 (LP/DT và LP/ST) ($P < 0,05$). Các quần thể còn lại cho thấy kết nối di truyền cao ($P > 0,05$), với giá trị F_{ST} từ $<0,001$ đến 0,04 (**Bảng 3.12**).

Mạng lưới haplotype được xây dựng dựa trên tập dữ liệu mã vạch RAD của 247 cá thể cá ét mọi thuộc 9 quần thể thu mẫu cùng với 1 trình tự tham chiếu từ Kandal - Campuchia (**Hình 3.13**). Kết quả cho thấy sự kết nối di truyền cao giữa các quần thể cá ét mọi ở LMB. Trong tổng số 49 haplotype được xác định, haplotype H1 thể hiện sự chia sẻ chung của tất cả 9 quần thể thu mẫu và trình tự tham chiếu ở Campuchia (70 trình tự, chiếm 28,2%). Hai haplotype (H7 và H10) được tìm thấy trong 8/9 quần thể, ngoại trừ quần thể Luang Prabang với số lượng trình tự tương ứng là 43 (17,3%)

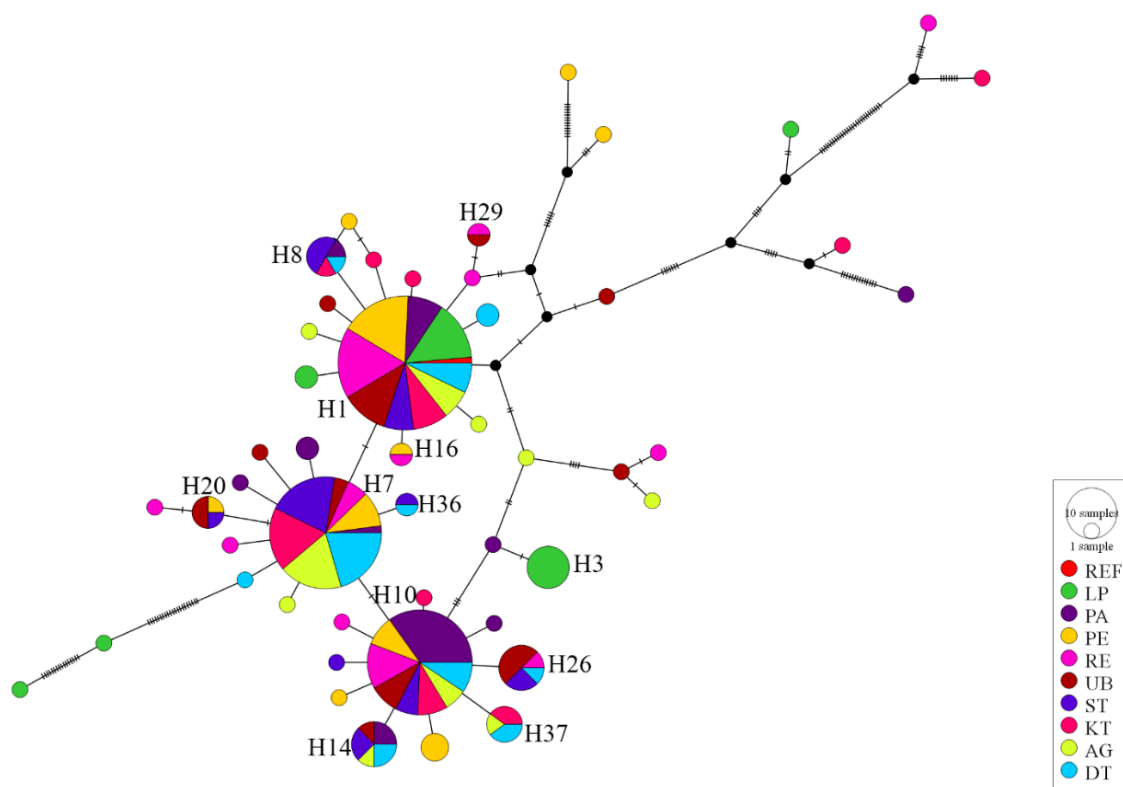
và 49 (19,8%). Bên cạnh đó, 5/9 quần thể (PA, UB-MK, ST, AG và DT) được ghi nhận ở haplotype H14 với 8 trình tự (3,2%). Haplotype H8 và H26 lần lượt được chia sẻ bởi 4/9 quần thể - PA, ST, KT, DT (6 trình tự, chiếm 2,4%) và RE, UB-MK, ST, DT (8 trình tự, chiếm 3,2%). Các haplotype H16, H20, H29, H36 và H37 ghi nhận các trình tự từ 2 – 3 quần thể. Đáng chú ý, một số lượng lớn haplotype đặc trưng (35/49), tức là các haplotype chỉ được tìm thấy ở một quần thể duy nhất, được ghi nhận ở tất cả quần thể. Riêng đối với quần thể Luang Prabang, ngoài haplotype H1 chung, nghiên cứu ghi nhận thêm 5 haplotype đặc trưng của quần thể này, trong đó haplotype H3 có 7/22 trình tự.

Bảng 3.12. Sự khác biệt di truyền (F_{ST}) giữa các quần thể cá ét mọi ở hạ lưu sông Mekong

Quần thể	LP	PA	PE	RE	UB-MK	ST	KT	AG	DT
LP	-								
PA	0,12	-							
PE	0,12	0,04	-						
RE	0,08	0,02	<0,001	-					
UB-MK	0,10	0,02	<0,001	<0,001	-				
ST	0,17	0,03	0,02	0,01	<0,001	-			
KT	0,09	0,01	<0,001	<0,001	0,01	<0,001	-		
AG	0,10	0,03	0,01	<0,001	<0,001	0,01	<0,001	-	
DT	0,17	0,02	0,02	0,01	0,01	<0,001	<0,001	0,01	-

*Ghi chú: Kí hiệu của các quần thể thu mẫu tương tự như **Bảng 3.11**. Giá trị in đậm là có ý nghĩa thống kê ($P < 0,05$)*

Dữ liệu mã vạch RAD, với chiều dài 757 bp, bao phủ cả vùng gen mã hóa và không mã hóa của hệ gen ti thể, đã chứng minh khả năng vượt trội trong việc phát hiện các vị trí đa hình với độ chính xác cao. Ưu điểm này cho phép phân tích hiệu quả sự đa dạng di truyền và sự khác biệt giữa các quần thể. Kết quả phân tích đa dạng di truyền bằng mã vạch RAD cho thấy sự kết nối quần thể cao ở hầu hết các khu vực hạ lưu sông Mekong, ngoại trừ khu vực thượng lưu-LMB (quần thể LP), nơi xuất hiện sự khác biệt rõ rệt và các haplotype đặc trưng quần thể.



Hình 3.13. Mạng lưới haplotype của các quần thể cá ét mọi ở hạ lưu sông Mekong sử dụng dữ liệu mã vạch RAD. Các đường nối giữa các vòng tròn thể hiện một bước đột biến. Đường gạch thể hiện việc cộng thêm các bước đột biến. Kích cỡ của vòng tròn thể hiện số lượng trình tự. Các màu sắc tương ứng với trình tự của các quần thể thu mẫu và trình tự từ Ngân hàng Gen [177]

Sự khác biệt nổi bật giữa kết quả mã vạch RAD và chỉ thị SNPs (vốn phân tách 3 nhóm quần thể riêng biệt) được giải thích bởi bản chất khác nhau của hai loại chỉ thị: mtDNA chỉ phản ánh cấu trúc di truyền dựa trên dòng mẹ, trong khi SNPs khảo sát biến dị trên toàn bộ hệ gen (gồm gen nhân và ti thể), cung cấp độ phân giải cao hơn. Sự đối lập này cho thấy các phân tách cục bộ trong quần thể cá ét mọi chỉ có thể được phát hiện bởi SNPs do khả năng khảo sát toàn hệ gen, nhấn mạnh tầm quan trọng của việc sử dụng kết hợp nhiều loại chỉ thị để nắm bắt toàn diện lịch sử tiến hóa và cấu trúc quần thể. Việc không phát hiện được cấu trúc quần thể rõ rệt ở hầu hết các khu vực LMB khi sử dụng mã vạch RAD tương đồng với nghiên cứu về hàu (*C. virginica*) ở Vịnh Mexico [150], nơi sự kết nối mạnh mẽ được quy cho giai đoạn trôi nổi kéo dài của ấu trùng (50-350 ngày), tạo điều kiện cho sự phát tán rộng rãi của các

cá thể. Tương tự, cá ét mọi cũng có giai đoạn trôi nổi của ấu trùng (35-50 ngày), góp phần vào sự kết nối quần thể. Tuy nhiên, đặc điểm di cư chằng ngấn/tùy nghi của cá ét mọi tạo ra các luồng gen cục bộ, và sự khác biệt giữa hai chỉ thị cung cấp một góc nhìn sâu sắc về lịch sử tiến hóa và khả năng thích nghi của loài cá di cư ở LMB.

Việc phát hiện các haplotype đặc trưng quần thể ở thượng lưu-LMB từ mã vạch RAD là một bước tiến quan trọng, mở ra tiềm năng ứng dụng trong nhiều lĩnh vực nghiên cứu. Các haplotype này có thể được sử dụng để theo dõi di cư, giúp xác định đường di chuyển của cá giữa các khu vực, đặc biệt hữu ích cho các loài di cư chằng ngấn và tùy nghi. Đồng thời, chúng cho phép xác định nguồn gốc quần thể, từ đó xác định các khu vực sinh sản và tuyến đường di cư quan trọng, hỗ trợ công tác bảo tồn và quản lý. Kết quả này cũng tương tự như nghiên cứu trước đây về quần thể cá lóc *C. striata* ở Biển Hồ, nơi ghi nhận các haplotype đặc trưng và sự khác biệt di truyền cao cho thấy sự hạn chế dòng gen giữa quần thể cá lóc ở Biển Hồ và các quần thể khác [159]. Điều này phù hợp với xu hướng chung trong sinh học bảo tồn, nơi mà các chỉ thị di truyền ngày càng được sử dụng để hỗ trợ các quyết định quản lý. Hơn nữa, sự biến đổi tần suất của haplotype theo thời gian có thể phản ánh tác động của các yếu tố môi trường như biến đổi khí hậu hay ô nhiễm, giúp đánh giá mức độ tổn thương và đưa ra biện pháp bảo tồn phù hợp. Cuối cùng, haplotype đặc trưng còn là công cụ để phân tích chi tiết di truyền quần thể, xác định mức độ biệt hóa di truyền và đánh giá các quá trình tiến hóa, từ đó cung cấp kiến thức sâu sắc về sinh học quần thể và hỗ trợ bảo tồn hiệu quả.

KẾT LUẬN VÀ KIẾN NGHỊ

Kết luận

1. Nghiên cứu đã thu thập và tạo thư viện từ ba loài cá đặc trưng của lưu vực sông Mekong: **272 mẫu (9 quần thể)** cá chạch lá tre, **255 mẫu (9 quần thể)** cá ét mọi, và **192 mẫu (7 quần thể)** cá vồ đém. Từ đó, hệ gen tham chiếu rút gọn của ba loài cá này được lắp ráp *de novo*. Kết quả phân tích SNPs xác định được **4.237 SNPs** đặc trưng cho cá chạch lá tre, **825 SNPs** cho cá ét mọi và **1.270 SNPs** cho cá vồ đém.

2. Mức độ đa dạng di truyền, khác biệt di truyền và cấu trúc quần thể của ba loài cá điển hình ở sông Mekong được xác định. Đồng thời, mô hình di cư của cá ét mọi và cá vồ đém ở LMB được dự đoán sử dụng chỉ thị SNPs. Cụ thể là:

- Tập tính di cư của các loài cá có mối tương quan thuận với mức độ đa dạng di truyền và tương quan nghịch với tỉ lệ giao phối cận huyết. Ở cá chạch lá tre không di cư, mức độ đa dạng di truyền của loài này thấp nhất ($H_o/H_e = 0,140/0,176$) và hệ số cận huyết cao nhất ($F_{IS} = 0,40-0,61$). Cá vồ đém - di cư chặng dài - có mức độ đa dạng di truyền cao nhất ($H_o/H_e = 0,220/0,266$) và hệ số cận huyết thấp nhất ($F_{IS} = 0,173$). Cá ét mọi - di cư chặng ngắn và tùy nghi - thể hiện mức độ đa dạng di truyền và giao phối cận huyết trung bình ($H_o/H_e = 0,185/0,233$, $F_{IS} = 0,225-0,299$).

- Kết quả phân tích sự khác biệt di truyền (F_{ST} và AMOVA) và cấu trúc quần thể ghi nhận: cá chạch lá tre và cá ét mọi thể hiện sự phân hóa quần thể rõ rệt theo khu vực địa lý - với lần lượt 4 nhóm (UMB, thượng và trung lưu-LMB, dòng nhánh trung lưu-LMB và hạ lưu-LMB) và 3 nhóm (thượng, trung, hạ lưu-LMB) - thì quần thể cá vồ đém lại duy trì sự kết nối di truyền cao trên toàn lưu vực LMB.

- Mô hình di cư (divMigrate) và lịch sử phân tách, pha trộn quần thể (Treemix) ghi nhận cả cá ét mọi và cá vồ đém ở LMB đều có khả năng di cư xuôi dòng, ngược dòng và vượt qua khu vực thác Khôn. Dựa trên dữ liệu di truyền, sự đối lập trong tập tính của hai loài được dự đoán: cá ét mọi (loài di cư chặng ngắn/tùy nghi) có xu hướng thực hiện một tuyến di cư dài khoảng 1.200 km; trong khi đó, cá vồ đém (loài di cư chặng dài) lại cho thấy mô hình không liên tục trong tuyến di cư, giới hạn giữa hạ lưu (AG) và trung lưu (UB-MK) với khoảng cách 1.000 km.

3. Nghiên cứu lắp ráp và chú giải hệ gen ti thể của cá ét mòi (*L. chrysophekadion*), đạt kích thước 16.600 bp và được đăng ký trên ngân hàng Gen với mã số OR637878. Vùng mã vạch RAD (757 bp) được xác định từ 247 cá thể cá ét mòi ở lưu vực sông Mekong. Kết quả cho thấy sự kết nối di truyền cao giữa các quần thể, với đa dạng haplotype đạt $0,849 \pm 0,014$ và đa dạng nucleotide đạt $0,005 \pm 0,0008$. Ngoại trừ quần thể Luang Prabang, không có sự khác biệt di truyền đáng kể nào được ghi nhận giữa các quần thể.

Kiến nghị

1. Về quản lý và bảo tồn:

- Chuyển đổi sang quản lý theo đơn vị di truyền: Thiết lập ranh giới bảo tồn dựa trên số lượng cụm quần thể đã được ghi nhận (04 vùng cho cá chạch lá tre, 03 vùng cho cá ét mòi và 01 vùng cho cá vô đém) để đảm bảo bảo vệ được tính đa dạng đặc hữu của từng khu vực địa lý.

- Duy trì tính kết nối tại các điểm giao sinh thái: Ưu tiên bảo vệ các đoạn sông xung quanh khu vực thác Khôn và các hành lang trung – hạ lưu để duy trì dòng gen xuyên biên giới, đặc biệt cho các loài cá trắng và cá xám.

2. Về định hướng nghiên cứu:

- Tích hợp các phương pháp luận và mở rộng quy mô: Kết hợp dữ liệu SNPs với phân tích đá tai và gắn thẻ điện tử, đồng thời mở rộng các địa điểm thu mẫu ở các khu vực sinh thái (như Siêm Riệp, Luang Prabang, hệ thống 3S) để so sánh giữa “dòng gen lịch sử” và “hành vi di cư thực tế” nhằm làm sáng tỏ các điểm mâu thuẫn về tuyến đường di cư đã phát hiện.

- Đánh giá tác động và giám sát lâu dài: Thiết lập chương trình giám sát định kỳ kích thước quần thể hiệu quả thông qua eDNA hoặc SNPs để đưa ra cảnh báo sớm về các quần thể có nguy cơ suy vong do cận huyết, đồng thời nghiên cứu sâu tác động của đập thủy điện và biến đổi khí hậu đến cấu trúc quần thể.

DANH SÁCH CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

Bài báo tạp chí ISI

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>. (SCIE, Q1)

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>. (SCIE, Q2)

Bài báo tạp chí khoa học trong nước

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>.

Bài báo hội nghị quốc tế

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 7th Asia Pacific International Modern Sciences Congress**, Jakarta, Indonesia, 363 – 372. ISBN 978-625-8246-59-9.

Bài báo hội nghị quốc gia

5. **Trương Thị Oanh**, Ngô Thái Bích Vân, Đặng Thúy 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.

Bài báo tạp chí Scopus

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. (Scopus, Q3).

TÀI LIỆU THAM KHẢO

1. Hortle KG (2009). Fishes of the Mekong: How many species are there? *Catch and Culture* **15**(2):4–12.
2. Valbo-Jørgensen J, Coates D, Hortle K (2009). Fish diversity in the Mekong River basin. In: Campbell IC (ed) *The Mekong*. Elsevier, Amsterdam, pp 161–196
3. Ziv G, Baran E, Nam S, Rodríguez-iturbe I, Levin SA (2012). Trading-off fish biodiversity, food security, and hydropower in the Mekong River Basin. *Proceedings of the National Academy of Sciences* **109**(15):5609–5614. <https://doi.org/10.1073/pnas.1201423109>.
4. Hughes K (2024). The Mekong's forgotten fishes and the emergency recovery plan to save them. WWF, Gland, Switzerland
5. Varis O, Kummu M, Salmivaara A (2012). Ten major rivers in monsoon Asia-Pacific: An assessment of vulnerability. *Applied Geography* **32**:441–454. <https://doi.org/10.1016/j.apgeog.2011.05.003>.
6. Poulsen AF, Valbo-Jørgensen J (2000). Fish migrations and spawning habits in the Mekong Mainstream - a survey using local knowledge. *AMFC Technical Report Mekong River Commission* :1–149.
7. Poulsen AF, Ouch P, Sintavong V, Ubolratana S, Nguyen TT (2002). Fish migrations of the Lower Mekong River basin: Implications for development, planning and environmental management. MRC Technical Paper No. 8, Mekong River Commission, Phnom Penh. 62 pp. ISSN: 1683-1489
8. Sor R, Ngor PB, Lek S, Chann K, Khoeun R, Chandra S, Hogan ZS, Null SE (2023). Fish biodiversity declines with dam development in the Lower Mekong Basin. *Scientific Reports* **13**:8571. <https://doi.org/10.1038/s41598-023-35665-9>.
9. Hurwood D (2005). Identifying stock structure of two *Henicorhynchus* species in the Mekong River using mitochondrial DNA. In: Proceeding 7th Tech. Symp. Mekong Fish. Thail. pp 253–263
10. So N, Maes GE, Volckaert FAM (2006). High genetic diversity in cryptic populations of the migratory sutchi catfish *Pangasianodon hypophthalmus* in the Mekong River. *Heredity* **96**:166–174. <https://doi.org/10.1038/sj.hdy.6800781>.
11. Ngamsiri T, Nakajima M, Sukmanomon S, Sukumasavin N, Kamonrat W, Nannakorn U, Taniguchi N (2007). Genetic diversity of wild Mekong giant catfish *Pangasianodon gigas* collected from Thailand and Cambodia. *Fisheries Science* **73**(4):792–799. <https://doi.org/10.1111/j.1444-2906.2007.01398.x>.
12. Adamson EAS, Hurwood DA, Baker AM, Mather PB (2009). Population subdivision in Siamese mud carp *Henicorhynchus siamensis* in the Mekong River basin: implications for management. *Journal of Fish Biology* **75**(6):1371–1392.
13. So N, Maes GE, Volckaert FAM (2006). Intra-annual genetic variation in the downstream larval drift of sutchi catfish (*Pangasianodon hypophthalmus*) in

- the Mekong river. *Biological Journal of the Linnean Society* **89**(4):719–728. <https://doi.org/10.1111/j.1095-8312.2006.00707.x>.
14. Iranawati F (2014). An assessment of the geographical scale of recurrent gene flow in wild populations of two species of Mekong River Carps (*Henicorhynchus* spp.). PhD Thesis, Queensland University
 15. Ackiss AS, Dang BT, Bird CE, Biesack EE, Chheng P, Phounvisouk L, Vu QHD, Uy S, Carpenter KE (2019). Cryptic lineages and a population dammed to incipient extinction? Insights into the genetic structure of a Mekong River catfish. *Journal of Heredity* **110**(5):535–547. <https://doi.org/10.1093/jhered/esz016>.
 16. Dang BT, Vu QHD, Biesack EE, Doan T V, Truong OT, Tran TL, Ackiss AS, Stockwell BL, Carpenter KE (2019). Population genomics of the peripheral freshwater fish *Polynemus melanochir* (Perciformes, Polynemidae) in a changing Mekong Delta. *Conservation Genetics* **20**:961–972. <https://doi.org/10.1007/s10592-019-01189-x>.
 17. Biesack EE, Dang BT, Ackiss AS, Bird CE, Chheng P, Phounvisouk L, Truong OT, Carpenter KE (2020). Evidence for population genetic structure in two exploited Mekong River fishes across a natural riverine barrier. *Journal of Fish Biology* **97**(3):696–707. <https://doi.org/10.1111/jfb.14424>.
 18. Vu NT, Ha TTT, Thuy VTB, Trang VT, Nguyen NH (2020). Population genomic analyses of wild and farmed striped catfish *Pangasianodon hypophthalmus* in the Lower Mekong River. *Journal of Marine Science and Engineering* **8**(6):471. <https://doi.org/10.3390/jmse8060471>.
 19. Hurwood DA, Adamson EAS, Mather PB (2007). Evidence for strong genetic structure in an economically important, highly vagile cyprinid (*Henicorhynchus lobatus*) in the Mekong River Basin. *Ecology of Freshwater Fish* **17**(2):273–283.
 20. Takagi AP, Ishikawa S, Nao T, Limsong S, Hort S, Thammavong K, Saphakdy B, Phomsouvanhm A, Nishida M, Kurokura H (2011). Population structure of the climbing perch, *Anabas testudineus*, in the lower Mekong River basin. *Fisheries Management and Ecology* **18**:145–153. <https://doi.org/10.1111/j.1365-2400.2010.00763.x>.
 21. Sunde J, Yildirim Y, Tibblin P, Forsman A (2020). Comparing the performance of Microsatellites and RADseq in population genetic studies: Analysis of data for Pike (*Esox lucius*) and a synthesis of previous studies. *Frontiers in Genetics* **11**(218):1–17. <https://doi.org/10.3389/fgene.2020.00218>.
 22. Kitada S, Kishino H (2021). Population structure of chum salmon and selection on the markers collected for stock identification. *Ecology and Evolution* **11**(20):13972–13985. <https://doi.org/10.1002/ece3.8102>.
 23. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**(10):e3376. <https://doi.org/10.1371/journal.pone.0003376>.

24. Liu S, Lu P, Liu D, Jin P, Wang W (2009). Pinpointing the sources and measuring the lengths of the principal rivers of the world. *International Journal of Digital Earth* **2**(1):80–87. <https://doi.org/10.1080/17538940902746082>.
25. Wang W, Lu H, Ruby Leung L, Li HY, Zhao J, Tian F, Yang K, Sothea K (2017). Dam construction in Lancang-Mekong River Basin could mitigate future flood risk from warming-induced intensified rainfall. *Geophysical Research Letters* **44**(20):10,378–10,386. <https://doi.org/10.1002/2017GL075037>.
26. Baird IG (2011). The Don Sahong Dam: Potential impacts on regional fish migrations, livelihoods and human health. *Critical Asian Studies POLIS Project on Ecological Governance, University of Victoria, PO Box 3060, University House 4, Victoria, BC, Canada V8W 3R4* ianbaird@shaw.ca www.polisproject.org **43**(2):211–235. <https://doi.org/10.1080/14672715.2011.570567>.
27. Abell R, Thieme ML, Revenga C, *et al* (2008). Freshwater ecoregions of the World: A new map of biogeographic units for freshwater biodiversity conservation. *BioScience* **58**(5):403–414.
28. Ruiz-Barradas A, Nigam S (2018). Hydroclimate variability and change over the Mekong River basin: Modeling and predictability and policy implications. *Journal of Hydrometeorology* **19**(5):849–869. <https://doi.org/10.1175/JHM-D-17-0195.1>.
29. Adamson PT, Rutherford ID, Peel MC, Conlan IA (2009). The hydrology of the Mekong River. In: Campbell IC (ed) *Mekong Biophys. Environ. an Int. River Basin*. Academic Press, San Diego, USA, pp 53–76
30. Hecht JS, Lacombe G, Arias ME, Dang TD, Piman T (2018). Hydropower dams of the Mekong River basin: a review of their hydrological impacts. *Journal of Hydrology* **568**:285–300. <https://doi.org/10.1016/j.jhydrol.2018.10.045>.
31. Poulsen AF, Hortle KG, Valbo-Jørgensen J, *et al* (2004). Distribution and ecology of some important riverine fish species of the Mekong River basin. MRC Technical Paper No. 10. ISSN: 1683-1489
32. MRC (2021). Recent fish migrations in Khone Falls (Lao PDR) according to local ecological knowledge. *MRC Technical Report Series, Vientiane* :69 pp.
33. Null SE, Farshid A, Goodrum G, Gray CA, Lohani S, Morrisett CN, Prudencio L, Sor R (2021). A meta-analysis of environmental tradeoffs of hydropower dams in the Sekong, Sesan, and Srepok (3S) rivers of the Lower Mekong Basin. *Water* **13**(63):1–18. <https://doi.org/10.3390/w13010063>.
34. Kang B, Huang X (2021). Mekong fishes: Biogeography, migration, resources, threats, and conservation. *Reviews in Fisheries Science and Aquaculture* **30**(2):170–194. <https://doi.org/10.1080/23308249.2021.1906843>.
35. Stimson Center (2023). Mekong dam monitor, sharing data, empowering people. <https://monitor.mekongwater.org/home/?v=1680811357217>.
36. MRC (2010). State of the Basin Report. *Mekong River Commission, Vientiane*,

Lao PDR :123.

37. Hortle KG, Bamrungrach P (2015). Fisheries habitat and yield in the Lower Mekong basin. Mekong River Commission, Phnom Penh, Cambodia
38. Cowx IG, Vu A V., Hogan Z, Mallen-Cooper M, Baumgartner LJ, Lai TQ, Grill G, Sayer CA (2024). Understanding the threats to fish migration: Applying the global swimways concept to the Lower Mekong. *Reviews in Fisheries Science & Aquaculture* :1–29. <https://doi.org/10.1080/23308249.2024.2401018>.
39. Robinson W, Ning N, Baumgartner L, *et al* (2023). Fish passage monitoring in Khone Falls for the Joint Environment Monitoring program pilots (JEM). CSU, Wagga Wagga
40. Welcomme RL (2001). Inland Fisheries: Ecology and Management. <https://doi.org/10.1002/9780470995693>
41. Baran E, Samadee S, Jiau TS, Tran TC (2013). Fish and fisheries in the Sekong, Sesan and Srepok basins (3S rivers, Mekong watershed), with special reference to the Sesan river. ICEM – International Centre for Environmental Management, Hanoi
42. Arthington AH, Welcomme RL (1995). The condition of large river systems of the world. In: N. B. Armantrout (ed) *Cond. world's Aquat. habitats*, Proc. World Fish. Congr. Science Publishers Inc., Lebanon, USA, pp 44–75
43. Gonzalez F, Barria P, Ponce F, Mora S (2023). Population genetic structure of marine fishes. *Popul Genet DNA to Evol Biol*. <https://doi.org/10.5772/intechopen.112694>
44. Gao Y, Sarker S, Sarker T, Leta OT (2022). Analyzing the critical locations in response of constructed and planned dams on the Mekong River basin for environmental integrity. *Environ Res Commun*. <https://doi.org/10.1088/2515-7620/ac9459>
45. Roberts TR (2001). On the river of no returns: Thailand's Pak Mun Dam and its fish ladder. *Natural History Bulletin of the Siam Society* **49**:189.
46. Roessig JM, Woodley CM, Cech JJ, Hansen LJ (2004). Effects of global climate change on marine and estuarine fishes and fisheries. *Reviews in Fish Biology and Fisheries* **14**(2):251–275. <https://doi.org/10.1007/s11160-004-6749-0>.
47. Eastham J, Mpelasoka F, Mainuddin M, Ticehurst C, Dyce P, Hodgson G, Ali R, Kirby M (2008). Mekong River Basin water resources assessment: Impacts of climate change. *CSIRO: Water for a Healthy Country National Research Flagship* :153 pp.
48. MRC (2009). Adaptation to climate change in the countries of the Lower Mekong basin. *MRC Management Information Booklet Series* :1–8.
49. Evers J, Pathirana A (2018). Adaptation to climate change in the Mekong River Basin: introduction to the special issue. *Climatic Change* **149**:1–11. <https://doi.org/10.1007/s10584-018-2242-y>.
50. Mai Đình Yên, Nguyễn Văn Trọng, Nguyễn Văn Thiện, Lê Hoàng Yên, Hứa

- Bạch Loan (1992). Định loại các loài cá nước ngọt Nam Bộ. Nhà xuất bản Khoa học và Kỹ thuật, Hà Nội
51. Trương Thủ Khoa, Trần Thị Thu Hương (1993). Định loại cá nước ngọt vùng đồng bằng sông Cửu Long. Nhà xuất bản Trường Đại học Cần Thơ
 52. Tran DD, Shibukawa K, Nguyen PT, Ha HP, Tran LX, Mai H Van, Utsugi K (2013). Mô tả định loại cá Đồng bằng sông Cửu Long, Việt Nam. Nhà xuất bản Đại học Cần Thơ, Việt Nam
 53. Robert TR (1980). A revision of the Asian Mastacembelid fish genus *Macrognathus*. *Copeia* **3**:385–391.
 54. Rainboth WJ (1996). Fishes of the Cambodian Mekong. *FAO Species Identification Field Guide for Fishery Purposes, Rome* :265, 27 colour plates.
 55. Kline JL, Loftus WF, Kotun K, Trexler JC, Rehage JS, Lorenz JJ, Robinson M (2014). Recent fish introductions into Everglades National Park: An unforeseen consequence of water management? *Wetlands* **34**:175–187. <https://doi.org/10.1007/s13157-012-0362-0>.
 56. Schofield PJ (2020). Expert bioblitzes facilitate non-native fish tracking and interagency partnerships. *Management of Biological Invasions* **11**(1):139–154. <https://doi.org/10.3391/mbi.2020.11.1.10>.
 57. Hui TH, Peng KKL, Huan LJ, Wei LB, Hing RLB, Beng JKT, Yeo DCJ (2020). The non-native freshwater fishes of Singapore: An annotated compilation. *Raffles Bulletin of Zoology* **68**:150–195. <https://doi.org/10.26107/RBZ-2020-0016>.
 58. U.S. Fish & Wildlife Service (2018). Peacock eel (*Macrognathus siamensis*) ecological risk screening summary. 1–12.
 59. Sokheng C, Chhuon Kim Chhea, Valbo-Jørgensen J (2000). Lateral fish migrations between Tonle Sap river and its floodplain. In: 3th Tech. Symp. Mekong Fish. Phnom Penh. pp 102–114
 60. Phomikong P, M. Fukushima, Sricharoendham B, Nohara S, Jutagate T (2015). Diversity and community structure of fishes in the regulated versus unregulated tributaries of the Mekong River. *River Research and Applications* **31**:1262–1275.
 61. Nuon V, Lek S, Ngor PB, So N, Grenouillet G (2020). Fish community responses to human-induced stresses in the Lower Mekong basin. *Water* **12**:1–21. <https://doi.org/10.3390/w12123522>.
 62. Hoàng Đức Đạt, Nguyễn Xuân Thu, Thái Ngọc Trí, Nguyễn Xuân Đồng (2007). Đặc điểm sinh học cá chạch lá tre (*Macrognathus siamensis* Gunther) ở Đồng bằng sông Cửu Long. In: Hội nghị Khoa học và Công nghệ. pp 583–587
 63. Saowakoon H, Saowakoon S (2007). Breeding and nursing of spotted spiny eel (*Macrognathus siamensis*; Gunther, 1861). In: Proc. 45th Kasetsart Univ. Annu. Conf. Fish. Sect. pp 712–721
 64. Ngô Trọng Lư, Nguyễn Kim Độ (2006). Kỹ thuật nuôi thủy đặc sản nước ngọt. Nhà xuất bản Nông nghiệp, Tp. Hồ Chí Minh

65. Jutagate T, Krudpan C (2004). Population dynamics of the Black shark-minnow *Morulius chrysophekadion* (Bleeker): Cyprinidae, in a “Run-of-the-River” reservoir in Thailand. *Asian Fisheries Science* **17**:1–8. <https://doi.org/10.33997/j.afs.2004.17.1.001>.
66. Baran E, Myschowoda C (2008). Have fish catches been declining in the Mekong River basin? *Modern Myths of the Mekong - Part II: Making living out of nature* :55–64.
67. Baran E, Baird IG, Cans G (2005). Fisheries bioecology at the Khone Falls (Mekong River, Southern Laos). WorldFish Center
68. Hortle KG, Khounsavanh O, Chanthasone P, Phommachanh P, Singhanouvong D, Viravong S (2015). Larval and juvenile fish in the Mekong River in Northern Lao PDR. MRC Technical Paper No.46. Mekong River Commission, Phnom Penh, Cambodia. 87 pp.
69. Chan B, Ngor PB, So N, Lek S (2017). Spatial and temporal changes in fish yields and fish communities in the largest tropical floodplain lake in Asia. *International Journal of Limnology* **53**:485–493. <https://doi.org/10.1051/limn/2017027>.
70. Mashyaka A, Duong T-Y (2021). Genetic diversity analysis revealed possible long migration of Black sharkminnow (*Labeo chrysophekadion*) along the Mekong River. *Songklanakarin Journal of Science and Technology* **43**(4):955–960.
71. Pin K, Nut S, Hogan ZS, Chandra S, Saray S, Touch B, Chheng P, Ngor PB (2020). 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* **12**:1–25. <https://doi.org/10.3390/w12092506>.
72. Halls AS, Conlan I, Wisesjindawat W, Phouthavongs K, Viravong S, Chan S, Vu VA (2013). Atlas of deep pools in the Lower Mekong River and some of its tributaries. Mekong River Commission, Phnom Penh, Cambodia
73. Pongsirijan S, Rungtongbaisuree S, Pongjanyakun T (2001). Induced breeding of Black-ear catfish, *Pangasius larnaudii* Bocourt, 1866. Tech. Pap. Freshw. Fish. Div. Dep. Fish. Minist. Agric. Coop.
74. Singanouvong D, Soulignavong C, Vonghachak K, Saadsy B, Warren TJ (1996). The main wet-season migration through Hoo Som Yai, a steep-gradient channel at the great fault line on the Mekong River, Champassack Province, Southern Lao PDR. *Indigenous Fishery Development Project, Fisheries Ecology Technical Report No 4* :115.
75. Bardach J (1959). Report on Fisheries in Cambodia. USOM, Cambodia, Phnom Penh
76. U.S. Fish & Wildlife Service (2020). Spot pangasius (*Pangasius larnaudii*) ecological risk screening summary. :1–12.
77. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007). Rapid and cost-effective polymorphism identification and genotyping using

- restriction site associated DNA (RAD) markers. *Genome research* **17**(2):240–248. <https://doi.org/10.1101/gr.5681207>.
78. Davey JW, Blaxter ML (2011). RADSeq: Next-generation population genetics. *Briefings in Functional Genomics* **9**(5):416–423. <https://doi.org/10.1093/bfgp/elq031>.
 79. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**(2):81–92. <https://doi.org/10.1038/nrg.2015.28>.
 80. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**(5):e37135. <https://doi.org/10.1371/journal.pone.0037135>.
 81. Wang S, Meyer E, McKay JK, Matz M V (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods* **9**(8):808–810.
 82. Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-silva I, Andrews KR, Bird CE (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* **1**: e203:1–15. <https://doi.org/10.7717/peerj.203>.
 83. Elshire RJ, Glaubitz JC, Sun Q, Poland J a, Kawamoto K, Buckler ES, Mitchell SE (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* **6**(5):e19379. <https://doi.org/10.1371/journal.pone.0019379>.
 84. Miller JR, Koren S, Sutton G (2010). Assembly algorithms for next-generation sequencing data. *Genomics* **95**(6):315–327. <https://doi.org/10.1016/j.ygeno.2010.03.001>.
 85. LaCava MEF, Aikens EO, Megna LC, Randolph G, Hubbard C, Buerkle CA (2019). Accuracy of de novo assembly of DNA sequences from double-digest libraries varies substantially among software. *Molecular Ecology Resources* **20**(2):360–370. <https://doi.org/10.1111/1755-0998.13108>.
 86. El-Metwally S, Hamza T, Zakaria M, Helmy M (2013). Next-Generation sequence assembly: Four stages of data processing and computational challenges. *PLOS Computational Biology* **9**(12):e1003345.
 87. Huang X, Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Research* **9**(9):868–877. <https://doi.org/10.1101/gr.9.9.868>.
 88. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembien LA, Berka J, Braverman MS, Chen YJ, Chen Z (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
 89. Zerbino DR, Birney E (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome research* **18**(5):821–829. <https://doi.org/10.1101/gr.074492.107>.
 90. Luo R, Liu B, Xie Y, *et al* (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1**(1):18. <https://doi.org/10.1186/2047-217X-1-18>.
 91. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011).

- Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**(3):171–182. <https://doi.org/10.1534/g3.111.000240>.
92. Ayling M, Clark MD, Leggett RM (2020). New approaches for metagenome assembly with short reads. *Briefings in Bioinformatics* **21**(2):584–594.
 93. Wojcieszek M, Pawełkiewicz M, Nowak R, Przybecki Z (2014). Genomes correction and assembling - present methods and tools. SPIE - Int. Soc. Opt. Eng.
 94. Warren RL, Sutton GG, Jones SJM, Holt RA (2007). Assembling millions of short DNA sequences using SSAKE. *Bioinformatics (Oxford, England)* **23**(4):500–501. <https://doi.org/10.1093/bioinformatics/btl629>.
 95. Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2007). SHARCGS, a fast and highly accurate short-read assembly algorithm for de novo genomic sequencing. *Genome research* **17**(11):1697–1706. <https://doi.org/10.1101/gr.6435207>.
 96. Chong Z, Ruan J, Wu C-I (2012). Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics* **28**(21):2732–2737. <https://doi.org/10.1093/bioinformatics/bts482>.
 97. Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**:e2584. <https://doi.org/10.7717/peerj.2584>.
 98. Liu ZJ, Cordes JF (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**(1):1–37. <https://doi.org/10.1016/j.aquaculture.2004.05.027>.
 99. Pierce BA (2003). Genetics: A conceptual approach. Worth Publishers Inc.,US
 100. Nguyễn Đức Thành (2014). Các kỹ thuật chỉ thị DNA trong nghiên cứu và chọn lọc thực vật. *Tạp chí Sinh học* **36**(3):265–294. <https://doi.org/10.15625/0866-7160/v36n3.5974>.
 101. Jehan T, Lakhanpaul S (2006). Single nucleotide polymorphism (SNP) - methods and applications in plant genetics: A review. *Indian Journal of Biotechnology* **5**(4):435–459.
 102. Morin PA, Luikart G, Wayne RK (2004). SNPs in ecology , evolution and conservation. *Trends in Ecology and Evolution* **19**(4):208–216. <https://doi.org/10.1016/j.tree.2004.01.009>.
 103. Reed DH, Frankham R (2003). Correlation between fitness and genetic diversity. *Conservation Biology* **17**(1):230–237. <https://doi.org/10.1046/j.1523-1739.2003.01236.x>.
 104. Allendorf FW, Hohenlohe P a, Luikart G (2010). Genomics and the future of conservation genetics. *Nature reviews Genetics* **11**(10):697–709. <https://doi.org/10.1038/nrg2844>.
 105. Hughes JB, Daily GC, Ehrlich PR (1997). Population diversity: its extent and extinction. *Science* **278**(5338):689–692. <https://doi.org/10.1126/science.278.5338.689>.

106. Mukhopadhyay T, Bhattacharjee S (2016). Genetic diversity: Its importance and measurements. *Conserving biological diversity: A multiscaled approach New Delhi, India: Research India Publications* :251–295.
107. Goudet J (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**(6):485–486. <https://doi.org/10.1093/oxfordjournals.jhered.a111627>.
108. Peakall R, Smouse PE (2012). GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**(19):2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>.
109. Meirmans PG, Van Tienderen PH (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**(4):792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>.
110. Purcell S, Neale B, Todd-Brown K, *et al* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**(3):559–575. <https://doi.org/10.1086/519795>.
111. Mahesh BA, Kannan E, Davis GDJ, Venkatesan P, Ragunath PK (2020). Genpop-An online tool to analyze human population genetic data. *Bioinformation* **16**(2):149–152. <https://doi.org/10.6026/97320630016149>.
112. Toro M, Fernández J, Caballero A (2009). Molecular characterization of breeds and its use in conservation. *Livestock Science* **120**(3):174–195. <https://doi.org/10.1016/j.livsci.2008.07.003>.
113. Hänfling B, Brandl R (1998). Genetic variability, population size and isolation of distinct populations in the freshwater fish *Cottus gobio* L. *Molecular Ecology* **7**(12):1625–1632. <https://doi.org/10.1046/j.1365-294x.1998.00465.x>.
114. Frankham R, Ballou JD, Briscoe DA (2002). Introduction to conservation genetics. <https://doi.org/10.1017/CBO9780511808999>
115. Frankham R (1996). Relationship of genetic variation to population size in wildlife. *Conservation Biology* **10**(6):1500–1508.
116. Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**(2):945–959. <https://doi.org/10.1093/genetics/155.2.945>.
117. Jombart T, Collins C (2015). A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. :1–43.
118. Ward RD (2000). Genetics in fisheries management. *Hydrobiologia* **420**(1):191–201. <https://doi.org/10.1023/A:1003928327503>.
119. Piorski NM, Sanches A, Carvalho-Costa LF, Hatanaka T, Carrillo-Avila M, Freitas PD, Galetti PM (2008). Contribution of conservation genetics in assessing neotropical freshwater fish biodiversity. *Brazilian Journal of Biology* **68**:1039–1050. <https://doi.org/10.1590/S1519-69842008000500011>.
120. Baran E, Zalinge N Van, Bun NP, Baird I, Coates D (2001). Fish resource and hydrobiological modelling approaches in the Mekong Basin. *ICLARM*,

- Penang, Malaysia and the Mekong River Commission Secretariat (May 2014):60 pp.
121. West G (1990). Methods of assessing ovarian development in fishes: A review. *Marine and Freshwater Research* **41**(2):199–222.
 122. Smokorowski KE, Pratt TC (2007). Effect of a change in physical structure and cover on fish and fish habitat in freshwater ecosystems – a review and meta-analysis. *Environmental Reviews* **15**(NA):15–41. <https://doi.org/10.1139/a06-007>.
 123. Baumgartner LJ, Boys CA, Barlow C, Roy M (2017). Lower Mekong fish passage conference: Applying innovation to secure fisheries productivity. *Ecological Management and Restoration* **18**(3):E8–E12. <https://doi.org/10.1111/emr.12272>.
 124. Campana SE (1999). Chemistry and composition of fish otoliths: Pathways, mechanisms and applications. *Marine Ecology Progress Series* **188**:263–297. <https://doi.org/10.3354/meps188263>.
 125. Gillanders BM (2005). Otolith chemistry to determine movements of diadromous and freshwater fish. *Aquatic Living Resources* **18**:291–300.
 126. Vũ Vi An, Nguyễn Nguyễn Du, Nguyễn Văn Phụng (2021). Khái quát về nghiên cứu cá di cư bằng phương pháp phân tích thành phần hóa học trong đá tai và khả năng áp dụng ở Việt Nam. *Tạp chí Nông nghiệp và Phát triển Nông thôn* **1**(4/2021):101–108.
 127. Brooks JL, Chapman JM, Barkley AN, *et al* (2018). Biotelemetry informing management: case studies exploring successful integration of biotelemetry data into fisheries and habitat management. *Canadian Journal of Fisheries and Aquatic Sciences* **76**(7):1238–1252. <https://doi.org/10.1139/cjfas-2017-0530>.
 128. Hogan Z, Em S, Tach P, Hortle KG (2006). Tagging fish - a case study from the Tonle Sap, Cambodia. *MRC Technical Paper No 12* (12):34 pp.
 129. Mitamura H, Mitsunaga Y, Arai N, Viputhanumas T (2008). Movements of immature hatchery-reared Mekong giant catfish *Pangasianodon gigas* released in the Mekong River, measured using acoustic telemetry. *Fisheries Science* **74**(5):1034–1039. <https://doi.org/10.1111/j.1444-2906.2008.01621.x>.
 130. Mitamura H, Mitsunaga Y, Arai N, Viputhanumas T (2006). Comparison of two methods of attaching telemetry transmitters to the Mekong giant catfish, *Pangasianodon gigas*. *Zoological science* **23**(3):235–238. <https://doi.org/10.2108/zsj.23.235>.
 131. Loury EK, Elliott VL, Ainsley SM, *et al* (2021). Priority knowledge needs for management of migratory fish species in Cambodia. *Fisheries Management and Ecology* **28**(5):393–416. <https://doi.org/10.1111/fme.12483>.
 132. Hawkins PR, Hortle KG, Phommanivong S, Singsua Y (2018). Underwater video monitoring of fish passage in the Mekong River at Sadam Channel, Khone Falls, Laos. *River Research and Applications* **34**(3):232–243. <https://doi.org/10.1002/rra.3239>.
 133. Wang J, Whitlock MC (2003). Estimating effective population size and

- migration rates from genetic samples over space and time. *Genetics* **163**(1):429–446. <https://doi.org/10.1093/genetics/163.1.429>.
134. Sundqvist L, Keenan K, Zackrisson M, Prodöhl P, Kleinhans D (2016). Directional genetic differentiation and relative migration. *Ecology and evolution* **6**(11):3461–3475. <https://doi.org/10.1002/ece3.2096>.
 135. Pickrell JK, Pritchard JK (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8**(11):e1002967. <https://doi.org/10.1371/journal.pgen.1002967>.
 136. Vu A V, Baumgartner LJ, Limburg KE, *et al* (2022). Life history strategies of Mekong pangasiid catfishes revealed by otolith microchemistry. *Fisheries Research* **249**:106239. <https://doi.org/10.1016/j.fishres.2022.106239>.
 137. Taanman JW (1999). The mitochondrial genome: Structure, transcription, translation and replication. *Biochimica et Biophysica Acta - Bioenergetics* **1410**(2):103–123. [https://doi.org/10.1016/S0005-2728\(98\)00161-3](https://doi.org/10.1016/S0005-2728(98)00161-3).
 138. Boore JL (1999). Animal mitochondrial genomes. *Nucleic Acids Research* **27**(8):1767–1780. <https://doi.org/10.1093/nar/27.8.1767>.
 139. Curole J, Kocher T (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology & Evolution* **14**(10):394–398. [https://doi.org/10.1016/s0169-5347\(99\)01660-2](https://doi.org/10.1016/s0169-5347(99)01660-2).
 140. Ballard và Whitlock, Ballard J, Whitlock M (2004). The incomplete natural history of mitochondria. *Molecular Ecology* **13**(4):729–744. <https://doi.org/10.1046/j.1365-294x.2003.02063.x>.
 141. Cameron SL (2014). How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Systematic Entomology* **39**(3):400–411. <https://doi.org/10.1111/syen.12071>.
 142. Cheng S, Chang S, Gravitt P, Respass R (1994). Long PCR. *Nature* **369**:684–685. <https://doi.org/10.1038/369684a0>.
 143. Sverdlöv E, Azhikina T (2005). Primer walking. *Encycl Life Sci*. <https://doi.org/10.1038/npg.els.0005382>
 144. Li R, Wang G, Wen ZY, Zou YC, Qin CJ, Luo Y, Wang J, Chen GH (2019). Complete mitochondrial genome of a kind of snakehead fish *Channa siamensis* and its phylogenetic consideration. *Genes and Genomics* **41**(2):147–157. <https://doi.org/10.1007/s13258-018-0746-5>.
 145. Gastineau R, Otis C, Boyle B, Lemieux C, Turmel M, St-Cyr J, Koken M (2023). The mitochondrial genome of the bioluminescent fish *Malacosteus niger* Ayres, 1848 (Stomiidae, Actinopterygii) is large and complex, and contains an inverted-repeat structure. *ZooKeys* **2023**(1157):177–191. <https://doi.org/10.3897/zookeys.1157.97921>.
 146. Sturk-Andreaggi K (2022). 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
 147. Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003). Biological

- identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270**:313–321. <https://doi.org/10.1098/rspb.2002.2218>.
148. Schroeter JC, Maloy AP, Rees CB, Bartron ML (2019). Fish mitochondrial genome sequencing: expanding genetic resources to support species detection and biodiversity monitoring using environmental DNA. *Conservation Genetics Resources* **12**(3):433–446. <https://doi.org/10.1007/s12686-019-01111-0>.
 149. Tisthammer KH, Forsman ZH, Sindorf VL, Massey TL, Bielecki CR, Toonen RJ (2016). The complete mitochondrial genome of the lobe coral *Porites lobata* (Anthozoa: Scleractinia) sequenced using ezRAD. *Mitochondrial DNA Part B Resouce* **1**(1):247–249. <https://doi.org/10.1080/23802359.2016.1157770>.
 150. Rue CR, Selwyn JD, Cockett PM, *et al* (2021). Genetic diversity across the mitochondrial genome of eastern oysters (*Crassostrea virginica*) in the northern Gulf of Mexico. *PeerJ*. <https://doi.org/10.7717/peerj.12205>
 151. Whitney JL, Coleman RR (2023). The complete mitochondrial genome of the Reef Manta Ray, *Mobula alfredi*, from Hawaii. *Mitochondrial DNA Part B* **8**(2):197–203. <https://doi.org/10.1080/23802359.2023.2167475>.
 152. Bird CE (2021). radBARCODER: A pipeline to assemble, align, and filter plastid genomes from nucleotide data in binary alignment maps. In: <https://github.com/cbirdlab/radBARCODER>.
 153. Baran E, Chum N, Fukushima M, Hand T, Hortle KG, Jutagate T, Kang B (2012). Fish biodiversity research in the Mekong basin. In: Biodivers. Obs. Netw. Asia-Pacific Reg. Ministry of Environment in Japan and Kyoto University, p 16.
 154. Willing EM, Hoffmann M, Klein JD, Weigel D, Dreyer C (2011). Paired-end RAD-seq for *de novo* assembly and marker design without available reference. *Bioinformatics* **27**(16):2187–2193. <https://doi.org/10.1093/bioinformatics/btr346>.
 155. Phan HTT, Tran SQ, Truong OT, Phounvisouk L, Grudpan C, Vu QHD, Dang BT (2022). De novo genome assembly and SNPs detection for population genetics of *Pangasius macronema* Bleeker, 1851 (Siluriformes: Pangasiidae) at Lower Mekong River. In: 5th Int. Eur. Conf. Interdiscip. Sci. Res. pp 342–355
 156. Vu QDH, Bird CE, Truong OT, Tran SQ, E. Ellen B, Ackiss AS, Carpenter KE, Dang BT (2025). Genome-wide SNP analyses reveal high gene flow of endemic smallscale croaker (*Boesemania microlepis*) in the Lower Mekong Basin. *Zoological Studies* **64**(18):1–13. <https://doi.org/10.6620/ZS.2025.64-18>.
 157. Jamaluddin JAF, So N, Tam BM, Ahmad A, Grudpan C, Page LM, Khaironizam MZ, Mohd Nor SA (2019). Genetic variation, demographic history and phylogeography of tire track eel, *Mastacembelus favus* (Synbranchiformes: Mastacembelidae) in Southeast Asia. *Hydrobiologia* **838**:163–182. <https://doi.org/10.1007/s10750-019-03987-3>.
 158. Jamaluddin JAF, So N, Tam BM, Ahmad A, Grudpan C, Sah ASRM, Nor SAM

- (2021). Genetic diversity of the tire track eel *Mastacembelus favus* in Southeast Asia inferred from microsatellite markers. *Ichthyological Research* **68**:529–540. <https://doi.org/10.1007/s10228-021-00807-4>.
159. Duong TY, Uy S, Chheng P, So N, Tran THT, Thi NT, Pomeroy R, Eгна H (2019). Genetic diversity and structure of striped snakehead (*Channa striata*) in the Lower Mekong basin: Implications for aquaculture and fisheries management. *Fisheries Research* **218**:166–173. <https://doi.org/10.1016/j.fishres.2019.05.014>.
 160. Pham TTN, Duong TY (2014). Genetic diversity of climbing perch (*Anabas testudineus*, Bloch 1792) populations based on RAPD and ISSR markers. *CTU Journal of Science Special is*(1):101–108.
 161. Dương Thúy Yên, Nguyễn Phương Thảo, Tiêu Văn Út, Trần Đắc Định (2018). Đa dạng di truyền của cá hường (*Helostoma temminckii*) ở Đồng bằng sông Cửu Long. *Tạp chí Khoa học Trường Đại học Cần Thơ* **54**(7B):86–93. <https://doi.org/10.22144/ctu.jvn.2018.144>.
 162. Duong TY, Nguyen TN-T, Tran DD (2020). Genetic diversity of Malayan leaf-fish (*Pristolepis fasciata* Bleeker, 1851) in the Mekong Delta. *CTU Journal of Science* **56**(Species issue: Fisheries):200–206. <https://doi.org/10.22144/ctu.jsi.2020.023>.
 163. Sawasawa W, Duong T-Y (2020). Genetic diversity of endangered snakehead *Channa lucius* (Cuvier, 1831) in the Mekong Delta inferred from ISSR markers. *Asian Fisheries Science* **33**:266–273.
 164. Takagi AP, Ishikawa S, Nao T, Song SL, Hort S, Thammavong K, Saphakdy B, Phomsouvanhm A, Nishida M, Kurokura H (2011). Genetic differentiation of *Macrognathus siamensis* within the Mekong River between Laos and Cambodia. *Journal of Applied Ichthyology* **27**(5):1150–1154. <https://doi.org/10.1111/j.1439-0426.2011.01774.x>.
 165. Duong T-Y, Nguyen T-V (2019). Genetic diversity of *Pangasius krempfi* in the Mekong River estuaries. *Can Tho University Journal of Science* **11**(2):81–88. <https://doi.org/10.22144/ctu.jen.2019.027>.
 166. Nguyen MT Le, Duong YT (2022). Genetic diversity of the migratory Mekong endemic catfish species *Pangasius macronema* populating along the Hau and Tien Rivers. *Science & Technology Development Journal* **25**(1):1–8.
 167. Yodsiri S, Wongpakam K, Ardarn A, Senakun C, Khumkratok S (2017). Population genetic structure and genetic diversity in Twisted-jaw fish, *Belodontichthys truncatus* Kottelat & Ng, 1999 (Siluriformes: Siluridae), from Mekong Basin. *International Journal of Zoology* **2017**:1–7. <https://doi.org/10.1155/2017/5976421>.
 168. Robinson W, Baumgartner LJ, Homsombath K, et al (2024). PIT tagging systems are suitable for assessing cumulative impacts of Mekong River hydropower plants on (upstream) fish migrations in Lao PDR. *Fisheries Research* **274**:106995.
 169. Eschenroeder JC, Lee D, Ounboundisane S, et al (2025). Mekong migrations:

- Insights into fish movement in the Lower Mekong from a large-scale acoustic telemetry study. *Fisheries Management and Ecology* :e12806. <https://doi.org/10.1111/fme.12806>.
170. Tran NT, Labonne M, Hoang HD, Panfili J (2019). Changes in environmental salinity during the life of *Pangasius krempfi* in the Mekong Delta (Vietnam) estimated from otolith Sr: Ca ratios. *Marine and Freshwater Research* **70**(12):1734–1746. <https://doi.org/10.1071/MF18269>.
 171. Hogan Z, Baird IG, Radtke R, Vander Zanden MJ (2007). Long distance migration and marine habitation in the tropical Asian catfish, *Pangasius krempfi*. *Journal of Fish Biology* **71**:818–832.
 172. Vu A V., Baumgartner LJ, Mallen-Cooper M, *et al* (2022). Diverse migration tactics of fishes within the large tropical Mekong River system. *Fisheries Management and Ecology* **29**(5):708–723. <https://doi.org/10.1111/fme.12566>.
 173. Nelson JS (2006). *Fishes of the world*, Fourth edi. <https://doi.org/10.15713/ins.mmj.3>
 174. Sahoo B, Das G, Sahoo L, Mahapatra K Das, Meher PK, Udit UK, Sundaray JK, Das P (2021). The complete mitochondrial genome of *Labeo catla* (Hamilton, 1822) using long read sequencing. *Mitochondrial DNA Part B: Resources* **6**(2):402–403. <https://doi.org/10.1080/23802359.2020.1870879>.
 175. Sahoo L, Bit A, Meher PK, Murmu K, Sundaray JK, Das P (2017). Rapid recovery of complete mitogenome of Indian major carp, *Catla catla* from low depth paired end Illumina sequencing. *Mitochondrial DNA Part B: Resources* **2**(1):155–156. <https://doi.org/10.1080/23802359.2017.1298413>.
 176. Jahan H, Chakraborty M, Alam MS, Begum RA (2024). Characterization of complete mitochondrial genome of *Labeo rohita* from Bangladesh. *Bioresearch Communications* **10**(2):1539–1544. <https://doi.org/10.3329/brc.v10i2.74578>.
 177. Yang L, Arunachalam M, Sado T, *et al* (2012). Molecular phylogeny of the cyprinid tribe Labeonini (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution* **65**(2):362–379. <https://doi.org/10.1016/j.ympev.2012.06.007>.
 178. Ferguson JW, Healey M, Dugan P, Barlow C (2011). Potential effects of dams on migratory fish in the Mekong River: Lessons from salmon in the Fraser and Columbia rivers. *Environmental Management* **47**:141–159. <https://doi.org/10.1007/s00267-010-9563-6>.
 179. Sabo JL, Ruhi A, Holtgrieve GW, Elliott V, Arias ME, Ngor PB, Räsänen TA, Nam S (2017). Designing river flows to improve food security futures in the Lower Mekong Basin. *Science* **358**(1270):1–11. <https://doi.org/10.1126/science.aao1053>.
 180. Ngor PB (2018). Fish assemblages dynamic in the tropical flood-pulse system of the Lower Mekong River basin. PhD Thesis, University of Toulouse, France
 181. Edwards S V., Robin V V., Ferrand N, Moritz C (2022). The evolution of comparative phylogeography: Putting the geography (and more) into comparative population genomics. *Genome Biology and Evolution* **14**(1):1–16.

- <https://doi.org/10.1093/gbe/evab176>.
182. Gauthier J, Moudén C, Suchan T, Alvarez N, Arrigo N, Riou C, Lemaitre C, Peterlongo P (2020). DiscoSnp-RAD: *de novo* detection of small variants for RAD-Seq population genomics. *PeerJ* **8**:e9291. <https://doi.org/10.7717/peerj.9291>.
 183. Hansen MJ, Douglas T, Beard J, Hayes DB (2007). Sampling and experimental design. In: Guy CS, Brown ML (eds) *Anal. Interpret. Freshw. Fish. Data*. American Fisheries Society, pp 51–120
 184. Andrews S (2010). FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
 185. Ewels P, Magnusson M, Lundin S, Käller M (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**(19):3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
 186. Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 187. Puritz JB, Hollenbeck CM, Gold JR (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**:e431. <https://doi.org/10.7717/peerj.431>.
 188. Eaton DAR, Overcast I (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**(8):2592–2594. <https://doi.org/10.1093/bioinformatics/btz966>.
 189. Bird CE (2022). dDocentHPC: A pipeline to trim, assemble, map, and genotype reduced representation genomic data. In: <https://github.com/cbirdlab/dDocentHPC>.
 190. Rochette NC, Rivera-Colón AG, Catchen JM (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* **28**(21):4737–4754. <https://doi.org/10.1111/mec.15253>.
 191. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
 192. Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
 193. Garrison E, Marth G (2012). Haplotype-based variant detection from short-read sequencing. arXiv 1207.
 194. Danecek P, Auton A, Abecasis G, *et al* (2011). The variant call format and VCFtools. *Bioinformatics* **27**(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
 195. Narum SR, Hess JE (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**:184–194. <https://doi.org/10.1111/>

- j.1755-0998.2011.02987.x.
196. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008). LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC bioinformatics* **9**:323. <https://doi.org/10.1186/1471-2105-9-323>.
 197. Foll M, Gaggiotti O (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**(2):977–993. <https://doi.org/10.1534/genetics.108.092221>.
 198. Lischer HEL, Excoffier L (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics (Oxford, England)* **28**(2):298–299. <https://doi.org/10.1093/bioinformatics/btr642>.
 199. Helyar SJ, Hemmer-Hansen J, Bekkevold D, *et al* (2011). Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**:123–136. <https://doi.org/10.1111/j.1755-0998.2010.02943.x>.
 200. Nei M (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70**(12):3321–3323. <https://doi.org/10.1073/pnas.70.12.3321>.
 201. Wright S (1943). Isolation by distance. *Genetics* **28**:114–138.
 202. Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**:564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>.
 203. Benjamini Y, Hochberg Y (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
 204. Waskom M (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software* **6**(60):3021. <https://doi.org/10.21105/joss.03021>.
 205. Earl D, Vonholdt B (2012). Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**:1–3. <https://doi.org/10.1007/s12686-011-9548-7>.
 206. Patterson N, Price AL, Reich D (2006). Population structure and eigenanalysis. *PLoS Genetics* **2**(12):2074–2093. <https://doi.org/10.1371/journal.pgen.0020190>.
 207. Jombart T, Ahmed I (2011). Adegnet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics (Oxford, England)* **27**(21):3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>.
 208. Oksanen A, Blanchet F, Friendly M, *et al* (2016). vegan: Community ecology package. R package version 2.4.3. <https://cran.r-project.org/package=vegan>
 209. Legendre P, Fortin M, Borcard D (2015). Should the Mantel test be used in

- spatial analysis? *Methods in Ecology and Evolution* **6**:1239–1247. <https://doi.org/10.1111/2041-210X.12425>.
210. Dray S, Bauman D, Blanchet G, Borcard D, Clappe S, Guillaume, Guenard Jombart G, Legendre P, Madi N, Wagner H (2020). Adespatial: Multivariate multiscale spatial analysis.
 211. Waples RS (2014). Testing for Hardy – Weinberg proportions: Have we lost the plot? *Journal of Heredity* **106**(1):1–19. <https://doi.org/10.1093/jhered/esu062>.
 212. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **14**(1):209–214. <https://doi.org/10.1111/1755-0998.12157>.
 213. Franklin IR (1980). Evolutionary changes in small populations. An evolutionary-ecological perspective, sinauer, Sunderland. :135–149.
 214. Epskamp S, Cramer A, Waldorp L, Schmittmann, VD Borsboom D (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software* **48**(4):1–18.
 215. Fitak RR (2021). OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols* **6**(1):1–6. <https://doi.org/10.1093/biomethods/bpab017>.
 216. Meng G, Li Y, Yang C, Liu S (2019). MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz173>
 217. Li D, Liu CM, Luo R, Sadakane K, Lam TW (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* **31**(10):1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
 218. Wheeler TJ, Eddy SR (2013). Nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**(19):2487–2489. <https://doi.org/10.1093/bioinformatics/btt403>.
 219. Birney E, Clamp M, Durbin R (2004). GeneWise and Genomewise. *Genome research* **14**(5):988–995. <https://doi.org/10.1101/gr.1865504>.
 220. Jühling F, Pütz J, Bernt M, Donath A, Middendorf M, Florentz C, Stadler PF (2012). 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 research* **40**(7):2833–2845. <https://doi.org/10.1093/nar/gkr1131>.
 221. Nawrocki EP, Eddy SR (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics (Oxford, England)* **29**(22):2933–2935. <https://doi.org/10.1093/bioinformatics/btt509>.
 222. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009). Circos: an information aesthetic for comparative genomics. *Genome research* **19**(9):1639–1645. <https://doi.org/10.1101/gr>.

- 092759.109.
223. 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. <https://doi.org/10.1093/molbev/msy096>.
 224. Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics*. <https://doi.org/10.6084/M9.FIGSHARE.963153.V1>.
 225. Löytynoja A, Vilella AJ, Goldman N (2012). Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics (Oxford, England)* **28**(13):1684–1691. <https://doi.org/10.1093/bioinformatics/bts198>.
 226. Gouy M, Guindon S, Gascuel O (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution* **27**(2):221–224. <https://doi.org/10.1093/molbev/msp259>.
 227. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**(18):2496–2497. <https://doi.org/10.1093/bioinformatics/btg359>.
 228. Grant WS, Bowen BW (1998). Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. *Journal of Heredity* **89**(5):415–426. <https://doi.org/10.1093/jhered/89.5.415>.
 229. Leigh JW, Bryant D (2015). POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**:1110–1116. <https://doi.org/10.1111/2041-210X.12410>.
 230. Li YL, Xue DX, Zhang BD, Liu JX (2018). An optimized approach for local de novo assembly of overlapping paired-end RAD reads from multiple individuals. *R Soc Open Sci*. <https://doi.org/10.1098/rsos.171589>
 231. Randhawa SS, Pawar R (2021). Fish genomes: Sequencing trends, taxonomy and influence of taxonomy on genome attributes. *Applied Ichthyology* **4**:553–562.
 232. Corral-Lou A, Perea S, Doadrio I (2021). High genetic differentiation in the endemic and endangered freshwater fish *Achondrostoma salmantinum* Doadrio and Elvira, 2007 from Spain, as revealed by mitochondrial and SNP markers. *Conservation Genetics* **22**: 585–600. <https://doi.org/10.1007/s10592-021-01381-y>.
 233. Escobar MD, Barroco L, Martínez JG, Bertuol F, Pouilly M, Freitas CE, Hrbek T (2024). How do hydroelectric dams affect non-migratory fish?: Genomic evidence for *Cichla temensis* (Perciformes : Cichlidae) in the Uatumã River, Amazonas, Brazil. *Biological Journal of the Linnean Society* **143**(1):blae078. <https://doi.org/10.1093/biolinnean/blae078>.
 234. Cure K, Thomas L, Hobbs JA, Fairclough D V, Kennington WJ (2017). Genomic signatures of local adaptation reveal source-sink dynamics in a high

- gene flow fish species. *Scientific Reports*: 1–10. <https://doi.org/10.1038/s41598-017-09224-y>.
235. Paz-Vinas I, Loot G, Stevens VM, Blanchet S (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology* **24**:4586–4604. <https://doi.org/10.1111/mec.13345>.
 236. Takagi APAP, Ishikawa S, Nao T, Song SLSL, Hort S, Thammavong K, Saphakdy B, Phomsouvanhm A, Nishida M, Kurokura H (2010). Genetic differentiation and distribution routes of the bronze featherback *Notopterus notopterus* (Osteoglossiformes: Notopteridae) in Indochina. *Biological journal of the Linnean Society* **101**(3):575–582.
 237. Crispo E, Bentzen P, Reznick DN, Kinnison MT, Hendry AP (2006). The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology* **15**(1):49–62. <https://doi.org/10.1111/j.1365-294X.2005.02764.x>.
 238. Zhao L, Chenoweth EL, Liu Q (2018). Population structure and genetic diversity of *Sinibrama macrops* from Ou River and Ling River based on mtDNA D-loop region analysis, China. *Mitochondrial DNA Part A* **29**(2):303–311. <https://doi.org/10.1080/24701394.2016.1278533>.
 239. Jaisuk C, Senanan W (2018). Effects of landscape features on population genetic variation of a tropical stream fish, Stone lapping minnow, *Garra cambodgiensis*, in the upper Nan River drainage basin, northern Thailand. *PeerJ* **6**:e4487. <https://doi.org/10.7717/peerj.4487>.
 240. Gupta A (2005). Landforms of Southeast Asia. In: Phys. Geogr. Southeast Asia. Oxford University Press New York, p 38.
 241. Grill G, Ouellet Dallaire C, Fluet Chouinard E, Sindorf N, Lehner B (2014). Development of new indicators to evaluate river fragmentation and flow regulation at large scales: A case study for the Mekong River basin. *Ecological Indicators* **45**:148–159. <https://doi.org/10.1016/j.ecolind.2014.03.026>.
 242. Thomaz AT, Christie MR, Knowles LL (2016). The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution* **70**(3):731–739. <https://doi.org/10.1111/evo.12883>.
 243. Chhuoy S, Hogan ZS, Chan B, Chandra S, Touch B, Sor R, Lek S, Ngor PB (2023). Declines in the Mekong’s megadiverse larval fish assemblages: Implications for sustainable development. *Sustainability (Switzerland)* **15**(18):1–19. <https://doi.org/10.3390/su151813535>.
 244. Song Y, Cheng F, Murphy BR, Xie S (2018). Downstream effects of the three gorges dam on larval dispersal, spatial distribution, and growth of the four major Chinese carps call for reprioritizing conservation measures. *Canadian Journal of Fisheries and Aquatic Sciences* **75**(1):141–151. <https://doi.org/10.1139/cjfas-2016-0278>.
 245. Lu XX, Siew RY (2006). Water discharge and sediment flux changes over the past decades in the Lower Mekong River: Possible impacts of the Chinese dams. *Hydrology and Earth System Sciences* **10**(2):181–195. <https://doi.org/>

- 10.5194/hess-10-181-2006.
246. Ngor PB, Oberdorff T, Phen C, Baehr C, Grenouillet G, Lek S (2018). Fish assemblage responses to flow seasonality and predictability in a tropical flood pulse system. *Ecosphere*. <https://doi.org/10.1002/ecs2.2366>
 247. Ye Y, Ren W, Zhang S, Zhao L, Tang J, Hu L, Chen X (2022). Genetic diversity of fish in aquaculture and of common carp (*Cyprinus carpio*) in traditional rice–fish coculture. *Agriculture (Switzerland)* **12**(997):1–17. <https://doi.org/10.3390/agriculture12070997>.
 248. Nguyen TTT, Sunnucks P (2012). Strong population genetic structure and its management implications in the mud carp *Cirrhinus molitorella*, an indigenous freshwater species subject to an aquaculture and culture-based fishery. *Journal of Fish Biology* **80**:651–68.
 249. Bradbury RI, Benzen P (2007). Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations. *Marine Ecology Progress Series* **340**:245–257. <https://doi.org/10.3354/meps340245>.
 250. Jensen AM, O’Neil NP, Iwaniuk AN, Burg TM (2019). Landscape effects on the contemporary genetic structure of Ruffed Grouse (*Bonasa umbellus*) populations. *Ecology and Evolution* **9**(10):5572–5592. <https://doi.org/10.1002/ece3.5112>.
 251. Maduna SN, Jónsdóttir ÓDB, Imsland AKD, Gíslason D, Reynolds P, Kapari L, Hangstad TA, Meier K, Hagen SB (2023). Genomic signatures of local adaptation under high gene flow in Lumpfish-Implications for broodstock provenance sourcing and larval production. *Genes* **14**:1–34. <https://doi.org/10.3390/genes14101870>.
 252. Graham CF, Boreham DR, Manzon RG, Wilson JY, Somers CM (2022). Population structure of lake whitefish (*Coregonus clupeaformis*) from the Mississippian lineage in North America. *Facets* **7**:853–874. <https://doi.org/10.1139/FACETS-2021-0191>.
 253. Roberts TR (2001). On the River of No Returns: Thailand’s Pak Mun Dam and its Fish Ladder. *Natural History Bulletin of the Siam Society* **49**:189–230.
 254. Baird IG, Manorom K, Phenow A, Gaja-Svasti S (2020). Opening the gates of the Pak Mun Dam: Fish migrations, domestic water supply, irrigation projects and politics. *Water Alternatives* **13**(1):141–159.
 255. Baird IG, Flaherty MS, Phylavanh B (2004). Mekong River Pangasiidae catfish migrations and the Khone Falls wing trap fishery in southern Laos. *Natural History Bulletin of the Siam Society* **52**:81–109.
 256. Hallerman E (2015). The role of genetics in managing declining fisheries resources. *Journal of Fisheries and Environment* **39**(2):40–74.
 257. Gandra M, Assis J, Martins MR, Abecasis D (2021). Reduced global genetic differentiation of exploited marine fish species. *Molecular Biology and Evolution* **38**(4):1402–1412. <https://doi.org/10.1093/molbev/msaa299>.
 258. Krause DJ, Bonin CA, Goebel ME, Reiss CS, Watters GM (2022). The rapid

- population collapse of a key marine predator in the northern antarctic peninsula endangers genetic diversity and resilience to climate change. *Frontiers in Marine Science* **8**(January):1–18. <https://doi.org/10.3389/fmars.2021.796488>.
259. Petrou EL, Fuentes-Pardo AP, Rogers LA, *et al* (2021). Functional genetic diversity in an exploited marine species and its relevance to fisheries management. *Proceedings of the Royal Society B: Biological Sciences* **288**:20202398. <https://doi.org/10.1098/rspb.2020.2398>.
 260. Plaisted HK, Novak AB, Weigel S, Klein AS, Short FT (2020). Eelgrass genetic diversity influences resilience to stresses associated with eutrophication. *Estuaries and Coasts* **43**(6):1425–1438. <https://doi.org/10.1007/s12237-019-00669-0>.
 261. Islam MN, Sultana S, Alam MJ (2020). Sequencing and annotation of the complete mitochondrial genome of a threatened labeonine fish, *Cirrhinus reba*. *Genomics and Informatics* **18**(3):1–7. <https://doi.org/10.5808/GI.2020.18.3.e32>.
 262. Duong TY, Pham LTK, Le XTK, Nguyen NTT, Nor SAM, Le TH (2023). Mitophylogeny of Pangasiid catfishes and its taxonomic implications for Pangasiidae and the suborder Siluroidei. *Zoological studies* **62**(48):1–30. <https://doi.org/10.6620/ZS.2023.62-48>.
 263. Laopichienpong N, Ahmad SF, Singchat W, *et al* (2021). Complete mitochondrial genome of Mekong fighting fish, *Betta smaragdina* (Teleostei: Osphronemidae). *Mitochondrial DNA Part B: Resources* **6**(3):776–778. <https://doi.org/10.1080/23802359.2021.1882893>.
 264. Zhang R, Deng L, Lv X, Tang Q (2022). Complete mitochondrial genomes of two catfishes (Siluriformes, Bagridae) and their phylogenetic implications. *ZooKeys* **1115**:103–116. <https://doi.org/10.3897/ZOOKEYS.1115.85249>.
 265. Liu ZQ, Shen Y zhong, Zhao M, Wang W, Chen W, Ma CY, Zhang FY, Ma LB (2020). Characterization of the complete mitochondrial DNA sequence of the *Ariosoma meeki* (Anguilliformes, Congridae). *Mitochondrial DNA Part B: Resources* **5**(3):2539–2541. <https://doi.org/10.1080/23802359.2020.1781557>.