

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
DIỄN HÌNH Ở LƯU VỰC HẠ LƯU SÔNG MEKONG

Chuyên ngành: Công nghệ Sinh học
Mã ngành: 9420201

TÓM TẮT LUẬN ÁN TIẾN SĨ KỸ THUẬT

Đà Nẵng - 2026

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH TẠI
TRƯỜNG ĐẠI HỌC BÁCH KHOA –
ĐẠI HỌC ĐÀ NẴNG**

Người hướng dẫn chính:

1. PGS.TS. Đặng Thúy Bình

2. TS. Ngô Thái Bích Vân

Phản biện 1:

Phản biện 2:

Luận án sẽ được bảo vệ trước hội đồng chấm luận án tiến sĩ cấp
cơ sở tại: Trường Đại học Bách Khoa

Vào lúc giờ ngày tháng năm

Có thể tìm hiểu luận án tại thư viện:

- Thư viện Quốc gia Việt Nam

- Trung tâm Thông tin – Học liệu & Truyền thông, Đại học Đà Nẵng

MỞ ĐẦU

1. Lý do chọn đề tài

Lưu vực sông Mekong (MRB), một trong những điểm nóng về đa dạng sinh học, đang phải đối mặt với các mối đe dọa ngày càng tăng từ sự suy thoái môi trường và các hoạt động của con người, đặc biệt là việc xây dựng đập thủy điện. Những tác động này không chỉ là rào cản vật lý ngăn cản sự di cư, làm thay đổi hoặc chia cắt môi trường sống tự nhiên của cá, gây ảnh hưởng đến tập tính sống của nhiều loài thủy sản mà còn làm thay đổi dòng chảy tự nhiên, nhịp lũ và lượng nước vào thời điểm giao mùa. Những thay đổi trên đã dẫn đến nguồn lợi thủy sản suy giảm và tăng nguy cơ tuyệt chủng của các loài cá. Do đó, nghiên cứu chuyên sâu về đa dạng di truyền và cấu trúc quần thể các loài cá đại diện cho các khu hệ sinh thái đặc trưng và tập tính sống đóng vai trò quan trọng trong công tác quản lý nghề cá, tái tạo quần đàn và hướng tới việc duy trì hệ sinh thái bền vững.

Cho đến nay, đa dạng các chỉ thị phân tử (DNA ti thể, đa hình đoạn khuếch đại ngẫu nhiên-RAPD, microsatellites, đa hình nucleotide đơn-SNPs) được sử dụng trong nghiên cứu di truyền quần thể của các loài cá sông Mekong. Việc hàng ngàn chỉ thị SNPs phân bố trên toàn bộ hệ gen được phát hiện từ dữ liệu giải trình tự DNA dựa trên vị trí cắt giới hạn (*Restriction site Associated DNA Sequencing*, RAD-seq) góp phần làm sáng tỏ cấu trúc quần thể của các sinh vật, đặc biệt là các loài chưa có thông tin về hệ gen.

Với đặc tính di cư đa dạng của các loài cá sông Mekong trong một môi trường biến động và thay đổi, ba loài cá có 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 chằng dãi) phân bố ở MRB được thu thập. Nghiên cứu hướng tới giải quyết các vấn đề 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

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 LMB 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.

3. Nội dung nghiên cứu

1) Lắp ráp *de novo* hệ gen của 03 loài cá điển hình ở LMB và xác định các chỉ thị phân tử SNPs.

2) Khảo sát và so sánh đa dạng di truyền, cấu trúc quần thể, ước lượng kích thước quần thể hiệu quả và dự đoán mô hình di cư của 03 loài cá ở LMB.

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 mọi.

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 MRB 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 vồ 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 vồ đém ở LMB, gọi 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. Phần mở đầu 05 trang, kết luận và kiến nghị 02 trang, các công trình đã công bố 01 trang, tài liệu tham khảo 22 trang và phụ lục. Nội dung chính của luận án chia làm 03 chương: Chương 1: Tổng quan gồm 36 trang, Chương 2: Đối tượng, phạm vi và phương pháp nghiên cứu gồm 24 trang; Chương 3: Kết quả nghiên cứu và thảo luận gồm 42 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. Đặc điểm khu vực hạ lưu sông Mekong

Sông Mekong có chiều dài là 4.909 km, chảy qua địa phận của Trung Quốc và Myanmar (phần thượng lưu, UMB) và bốn quốc gia Lào, Thái Lan, Campuchia và Việt Nam (phần hạ lưu, LMB). MRB là một trong những điểm nóng toàn cầu về đa dạng sinh học và là hệ sinh thái có đa dạng sinh học cao thứ ba trên thế giới sau sông Amazon và Congo, với khoảng 1393 loài cá được biết, trong đó 293 loài (21%) được xác định là loài di cư. Theo các đặc điểm sinh thái và tập tính di cư, các loài cá ở MRB được chia thành 3 nhóm là cá trắng, cá đen và cá xám. Nhiều loài cá sông Mekong được biết đến là loài di cư theo chiều dọc (di cư lên-xuống thượng và hạ nguồn) hoặc/và chiều ngang (di cư từ sông vào các vùng ngập) và giữa các nơi cư trú. Quá trình di cư thường diễn ra ở tất cả các giai đoạn phát triển của cá và liên quan đến việc trú ẩn trong mùa khô, kiếm mồi trong mùa lũ và di cư để sinh sản cũng như tránh các điều kiện môi trường bất lợi. Ba hệ thống di cư đường dài (thượng lưu, trung lưu và hạ lưu) và hai hệ thống di cư ngang (Biển Hồ, Sekong-Sesan-Srepok và Korat) được xác định. Các yếu tố thủy văn được xem như là tín hiệu cho sự di cư của các loài cá.

Tuy nhiên, MRB đang phải đối mặt với những tác động từ các hoạt động của con người (xây dựng đập thủy điện, khai thác quá mức, tăng dân số, đô thị hóa và thay đổi mục đích sử dụng đất, ô nhiễm môi trường) và những ảnh hưởng của biến đổi khí hậu (sự thay đổi nhiệt độ và lượng mưa, xâm nhập mặn). Những tác động này không chỉ là rào cản vật lý ngăn cản sự di cư, làm thay đổi hoặc chia cắt môi trường sống tự nhiên của cá, gây ảnh hưởng đến tập tính sống của nhiều loài thủy sản mà còn làm thay đổi dòng chảy tự nhiên, nhịp lũ và lượng nước vào thời điểm giao mùa. Từ những thay đổi trên, nguồn lợi thủy

sản suy giảm và nguy cơ tuyệt chủng của các loài cá tầng. Do đó, việc nghiên cứu chuyên sâu về đa dạng di truyền và cấu trúc quần thể các loài cá đại diện cho các khu hệ sinh thái đặc trưng và tập tính sống đa dạng đóng vai trò quan trọng trong công tác quản lý nghề cá, tái tạo quần đàn, hướng tới duy trì hệ sinh thái bền vững cũng như theo dõi sự thích nghi đáp ứng với môi trường thay đổi của sinh vật.

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

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

Cá chạch lá tre là loài bản địa ở MRB, Mae Klong, Chao Phraya (Thái Lan), bán đảo Mã Lai (Malaysia), là loài ngoại lai ở Hoa Kỳ và Singapore. Chúng thuộc nhóm cá đen-không di cư và sống ở tầng đáy.

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

Cá ét mọi phân bố chủ yếu ở Đông Nam Á, gồm MRB, Chao Phraya (Thái Lan), bán đảo Mã Lai (Malaysia), Sumatra, Java và Borneo (Indonesia). Cá thường sống thành đàn, tập trung ở tầng giữa và tầng đáy. Chúng là loài di cư tùy nghi, có thể thuộc nhóm cá xám – di cư chặng ngắn hoặc cá trắng – di cư chặng dài. Nhiều quần đàn cá ét mọi phân bố ở các dòng nhánh và một số hồ chứa ở Thái Lan.

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

Cá vồ đém phân bố trên các sông vừa và lớn và các vùng ngập ở MRB và Chao Phraya (Thái Lan). Chúng thuộc nhóm cá trắng và thường di cư vào tháng 4 – 7 để tìm kiếm thức ăn và sinh sản. Một quần đàn cá vồ đém ở hạ lưu được ghi nhận, bắt đầu từ Pakse (Lào) đến Đồng bằng sông Cửu Long-ĐBSCL (Việt Nam), kể cả hệ thống sông Tonle Sap và không có thông tin về cấu trúc đàn ở thượng lưu.

1.3. Tổng quan về phương pháp nghiên cứu

1.3.1. Kỹ thuật giải trình tự DNA dựa trên vị trí cắt giới hạn (RAD-seq)

Kỹ thuật RAD-seq được phát triển nhằm làm giảm độ phức tạp và

kích thước của hệ gen bằng cách sử dụng các enzyme giới hạn để cắt hệ gen thành những phân đoạn DNA ngắn và tiến hành giải trình tự thế hệ mới (*Next generation sequencing*-NGS). Từ đó, hàng ngàn chỉ thị phân tử đa hình nucleotide đơn (SNPs) - sự khác biệt trình tự DNA của hai cá thể của loài ở mức độ một cặp base, khi một nucleotide đơn (A, T, G, C) trong hệ gen bị thay thế - được xác định. Trong số các kỹ thuật RAD-seq, ezRAD được áp dụng thành công trong các nghiên cứu về di truyền quần thể ở các loài sinh vật chưa có thông tin về hệ gen. Việc tạo thư viện gen được tối ưu hóa bằng cách sử dụng linh hoạt các enzyme cắt giới hạn và bộ kit thương mại, do đó, kỹ thuật này được lựa chọn trong nghiên cứu hiện tại.

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, lắp ráp *de novo* trong RAD-seq không nhằm xây dựng hệ gen hoàn chỉnh mà tạo ra một hệ gen tham chiếu rút gọn. Tập hợp các trình tự liên kề vị trí cắt hạn chế này đóng vai trò làm khung căn chỉnh các đoạn đọc để xác định SNPs phục vụ phân tích di truyền quần thể. Để lắp ráp *de novo* hệ gen, các bộ công cụ được phát triển dựa trên thuật toán đồ thị (*graph-based*) và tham lam (*greedy-based*).

1.3.3. Các phương pháp đánh giá đa dạng di truyền, khác biệt di truyền và cấu trúc quần thể

Đa dạng di truyền được đánh giá qua các thông số như số lượng alen trên mỗi locus, số lượng alen hiệu quả, dị hợp tử quan sát và mong đợi, tần suất alen thay thế và hệ số cận huyết. Sự biến dị di truyền của quần thể được xác định bằng tần số alen (tỉ lệ các locus đa hình trong hệ gen) và tỉ lệ dị hợp tử (tỉ lệ cá thể trong quần thể dị hợp tử tại một locus cụ thể). Việc xác định các nhóm cấu trúc quần thể - tập hợp những cá thể trong cùng một loài có sự khác biệt di truyền với

các cá thể khác trong một quần thể, là cơ sở cho việc xác định chiến lược quản lý nguồn lợi thủy sản.

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

Đối với phương pháp phụ thuộc vào nghề cá, khảo sát kiến thức sinh thái của người dân địa phương và giám sát các hoạt động đánh bắt cá là hai phương pháp phổ biến nhằm ghi nhận xu hướng quần thể, đánh giá sự đa dạng loài và suy ra thời gian di cư.

Đối với phương pháp không phụ thuộc vào nghề cá, sự di cư của cá được xác định thông qua nghiên cứu đặc điểm sinh học sinh sản, nuôi nhốt trong điều kiện thí nghiệm, phân tích thành phần hóa học trong đá tai, gắn thẻ, sử dụng các thiết bị hiện đại và công cụ di truyền.

1.3.5. Phương pháp lắp ráp hệ gen ti thể và xác định mã vạch RAD

Để giải trình tự và lắp ráp hệ gen ti thể, một số phương pháp được thực hiện như (i) khuếch đại các phân đoạn DNA bằng kỹ thuật PCR và giải trình tự Sanger, (ii) khuếch đại các phân đoạn DNA đoạn dài (từ 2kb) và giải trình tự Sanger, (iii) khuếch đại các phân đoạn DNA đoạn dài và NGS, (iv) tạo thư viện gen sử dụng enzyme cắt giới hạn và NGS.

Mã vạch RAD là tập hợp những phân đoạn DNA ngắn nằm trong hệ gen ti thể của sinh vật được xác định bằng cách dóng hàng dữ liệu RAD-seq với trình tự hệ gen ti thể của loài nhằm định loại sinh vật hoặc/và xác định các biến thể địa lý.

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

Lưu vực sông Mekong là hệ sinh thái có đa dạng sinh học cao, tuy nhiên, thông tin về hệ gen của các loài cá vẫn còn hạn chế. Với sự phát triển của NGS và ứng dụng các công cụ tin sinh học, hệ gen của một số loài cá ở MRB được xây dựng thành công (như cá tra *Pn. hypophthalmus*, cá xác sọc *P. macronema*). Hệ gen của một số loài cá cũng được lắp ráp *de novo* trong các nghiên cứu di truyền quần thể.

Các nghiên cứu về đa dạng và di truyền quần thể của các loài cá sông Mekong đã được thực hiện. Ở nhóm cá không di cư, sự đa dạng di truyền cao và cấu trúc quần thể có sự phân tách tương ứng với các khu vực địa lý ở các dòng chính của sông Mekong thuộc Campuchia và Lào (như cá rô đồng *Anabas testudineus* và cá chạch lá tre) và ở khu vực ĐBSCL (như cá rô đồng, cá hường *Helostoma temminckii*, cá rô biển *Pristolepis fasciata* và cá phèn vàng *Polynemus melanochir*) được ghi nhận. Ở nhóm cá di cư chặng ngắn, Ackiss *et al.* (2019) ghi nhận sự phân tách quần thể cá lăng *Hemibagrus spilopterus* ở trên (Paske, Lào) và dưới thác Khôn (Kratie, Stung Treng, Campuchia) và giữa các nhánh sông 3S tại Lào (Sekong), Campuchia (Sesan) và Việt Nam (Serepok). Hệ số cận huyết cao và kích thước quần thể hiệu quả thấp ở Đắc Lắc (sông Serepok) được ghi nhận, và nguyên nhân dự đoán là do sự phát triển của đập thủy điện. Sự đa dạng di truyền cao và khoảng cách di truyền thấp giữa các quần thể cá ét mọi ở Lào và Việt Nam, Mashyaka and Duong (2021) giả thuyết đây là loài cá di cư đường dài. Ở nhóm cá di cư chặng dài, sự phân tách quần thể phụ thuộc vào đối tượng nghiên cứu, khu vực phân bố, hoặc/và các chỉ thị phân tử được sử dụng. Đối với cá tra, không có sự khác biệt giữa các quần thể thu thập ở Campuchia khi sử dụng chỉ thị RFLP hoặc ba nhóm quần thể được phát hiện ở Campuchia và Việt Nam sử dụng chỉ thị microsatellite, hoặc hai nhóm quần thể ở phía trên và dưới thác Khôn sử dụng chỉ thị SNPs. Sự kết nối quần thể cũng được ghi nhận ở một số loài cá da trơn khác thuộc họ Pangasiidae (cá basa *P. bocourti*, cá tra dầu *Pn. gigas*, cá bông lau *P. krempfi* và cá sát sọc *P. macronema*). Đối với cá linh *Henycorhynchus siamensis*, sự phân tách các quần thể ở phía trên và dưới thác Khôn được thể hiện khi sử dụng chỉ thị DNA ti thể, tuy nhiên, sử dụng chỉ thị microsatellite

không có sự khác biệt giữa hai khu vực này. Đối với cá linh *H. lobatus*, sự kết nối quần thể ở dòng chính sông Mekong được xác định khi sử dụng chỉ thị DNA ti thể và microsatellite, trong khi chỉ thị SNPs cho thấy sự khác biệt của quần thể phía trên và dưới thác Khôn.

Từ những dẫn liệu tổng quan, các nghiên cứu về đa dạng và di truyền quần thể của các loài cá sông Mekong được thực hiện trên từng đối tượng riêng lẻ hoặc kết hợp 2 đối tượng nhưng với quy mô nhỏ và số lượng mẫu hạn chế. Kết quả của những nghiên cứu chỉ ra các mức độ khác nhau về cấu trúc quần thể cá ở các dòng chính của sông Mekong, cũng như giữa dòng chính và các dòng nhánh. Đồng thời, sự suy giảm đa dạng di truyền, tỉ lệ cận huyết cao, sự phân tách giữa các quần thể tự nhiên của các loài cá sông Mekong cũng được ghi nhận và nguyên nhân dự đoán là do tác động của sự khai thác quá mức, phát triển đập thủy điện và biến đổi khí hậu. Bên cạnh đó, đa dạng các chỉ thị phân tử như DNA ti thể, microsatellite và SNPs được sử dụng.

Trong một hệ sinh thái tổng hợp và đa loài như lưu vực sông Mekong thì việc chỉ quản lý đơn loài là không khả thi. Phương pháp tiếp cận địa sinh học so sánh và so sánh hệ gen quần thể được sử dụng bằng cách so sánh cấu trúc quần thể dựa trên thông tin di truyền (ở mức độ gen và hệ gen) của các loài theo môi trường phân bố. Các phân tích này góp phần làm rõ các tác động dẫn đến biến dị di truyền, cũng như các dòng tiến hóa theo đặc trưng phân bố địa lý, lịch sử phát triển và biến động môi trường sống của các loài sinh vật.

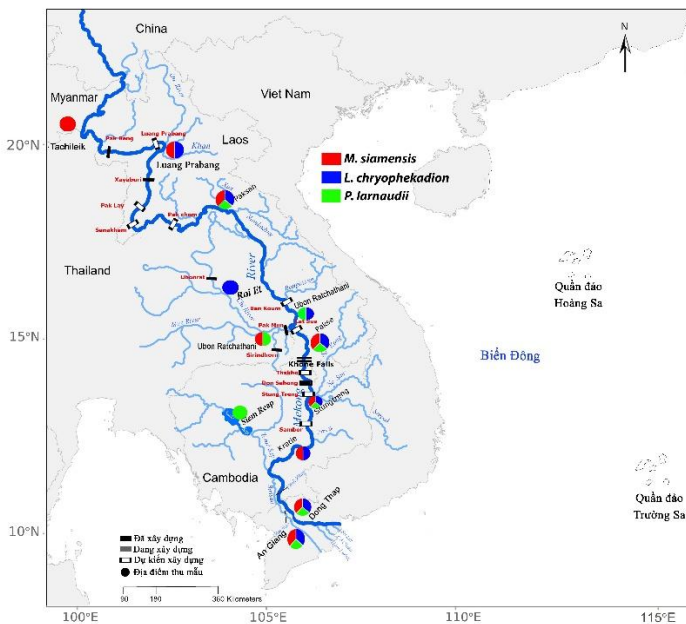
Cho đến nay, chỉ có một trình tự hệ gen ti thể của cá ét mọi được công bố và sử dụng để khảo sát mối quan hệ phát sinh loài ở tông Labeonini (họ Cyprinidae). Nhằm bổ sung những thông tin khoa học về cá ét mọi, nghiên cứu này tiến hành lắp ráp và chú giải thêm một hệ gen

ti thể cá ét mọi, xác định mã vạch RAD thuộc hệ gen ti thể để khảo sát sự đa dạng di truyền và cấu trúc quần thể của cá ét mọi ở LMB.

CHƯƠNG 2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

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

Cá chạch lá tre, cá ét mọi và cá vô đốm được thu thập từ các ngư dân đánh bắt, chợ địa phương hoặc cảng cá ở MRB từ 2017 – 2021 (**Hình 2.2**). Mẫu cá được định loại tại thực địa dựa vào các đặc điểm hình thái theo Rainboth (1996) và Trần Đức Định và cộng sự (2013).



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 (Khone Falls).

Nghiên cứu thu thập được 272 mẫu cá chạch lá tre (9 quần thể), 263 mẫu cá ét mọi (10 quần thể), 193 mẫu cá vô đốm (8 quần thể) được thu thập (**Bảng 2.1**). Tuy nhiên, tại Tacheilek, số lượng mẫu cá

ét mọi và cá vô đém thu được không đủ để phân tích di truyền quần thể, do đó các mẫu này đã bị loại khỏi nghiên cứu.

Bảng 2.1. Số lượng cá thể của các loài cá thu được tại các địa điểm thu mẫu ở lưu vực sông Mekong

Quốc gia	Vị trí trên sông Mekong		Địa điểm thu mẫu (Kí hiệu)	Cá chạch lá tre	Cá ét mọi	Cá vô đếm
Myamnar	Thượng lưu-UMP	Dòng nhánh	Tacheilek (TK)	32	8*	1*
Lào	Thượng lưu - LMB	Dòng nhánh	Luang Prabang (LP)	32	22	--
		Trung lưu - LMB	Dòng chính	Paksan (PA)	34	32
	Dòng chính		Pakse (PE)	27	32	--
Thái Lan	Dòng chính		Ubon Ratchathani (UB-MK)	--	28	29
	Dòng nhánh		Ubon Ratchathani (UB-MR)	32	--	24
	Dòng nhánh		Roi Et (RE)	--	30	--
Campuchia	Hạ lưu - LMB	Hợp lưu Mekong-3S	Strung Treng (ST)	32	28	32
		Dòng chính	Kratié (KT)	28	27	--
		Dòng nhánh	Siêm Riệp (SR)	--	--	30
Việt Nam		Dòng chính	Đồng Tháp (DT)	25	32	32
		Dòng chính	An Giang (AG)	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 và giải trình tự

DNA được tách chiết từ 30 mg mẫu cơ của từng cá thể của mỗi loài cá sử dụng bộ kit Wizard® Genomic DNA Purification (Promega, Mỹ). Nồng độ DNA được xác định bằng máy đo huỳnh quang Qubit® 2.0 Fluorometer (Invitrogen, USA). Thư viện DNA hệ gen của các loài cá được thu nhận bằng kỹ thuật ezRAD sử dụng bộ kit TruSeq Nano HT Library Preparation (Illumina) theo Toonen *et al.* (2013) và Dang *et al.* (2019). Thư viện DNA được giải trình tự thế hệ mới HiSeq 4000

tại phòng Thí nghiệm Di truyền, Trường Đại học Texas A&M Corpus Christi, Hoa Kỳ. Thông tin về mẫu vật và trình tự đoạn đọc thô của các loài cá được đăng ký mã số trên Ngân hàng Gen.

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

Trình tự đoạn đọc thô (file FASTQ) của ba loài cá được đánh giá chất lượng sử dụng công cụ FastQC và MultiQC, loại bỏ các trình tự và base chất lượng thấp bằng Trimmomatic v0.33. Sử dụng công cụ zcat v1.10, các đoạn đọc chất lượng cao được kết nối dựa trên vùng chồng lắp tạo thành các contigs – dữ liệu đầu vào để lắp ráp *de novo* hệ gen.

Sử dụng tập hợp các đoạn contigs của cá vồ đém, 03 bộ công cụ lắp ráp *de novo* hệ gen được khảo sát gồm dDocent 4.5, ipyrad 0.9.84 – đều sử dụng thuật toán tham lam và Stack 2.4 – sử dụng thuật toán đồ thị.

Để đánh giá kết quả lắp ráp *de novo*, các thông số lắp ráp (số lượng các đoạn contig, độ dài N50 của contigs, số lượng, kích thước lớn nhất và kích thước trung bình của cụm (cluster) và kích thước của hệ gen) và tỉ lệ các đoạn đọc được đóng hàng chính xác vào hệ gen được xác định lần lượt bằng công cụ awk v5.1.0, grep v3.4 và Samtools v1.9. 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 cá chạch lá tre và cá ét mọi.

Sau khi có hệ gen của 3 loài cá, chỉ thị SNPs thô được xác định bằng công cụ Freebayes và tiến hành lọc sử dụng vcftools và vcfFilter nhằm thu nhận các SNPs có ý nghĩa di truyền. Sau đó, các loci ngoại vi được xác định thông qua Lositan Selection Workbench và BayeScan v2.1 và bị loại bỏ khỏi tập dữ liệu để thu được các SNPs trung tính. 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 các loài cá.

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 LMB

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

Sử dụng dữ liệu SNPs trung tính, sự đa dạng di truyền của các quần thể cá được đánh giá thông qua các thông số gồm dị hợp tử quan sát (H_o) và dị hợp tử mong đợi (H_e) cho mỗi quần thể và gộp chung tất cả các cá thể bằng GenAlEx v6.5.2.

2.3.2. Xây dựng cấu trúc di truyền quần thể các loài 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}), khác biệt di truyền theo cặp (F_{ST}) giữa các quần thể và giá trị độ tin cậy được tính toán bằng Alrequin v3.5. Cấu trúc di truyền quần thể của các loài cá ở LMB được xác định sử dụng phương pháp phân tích nhóm STRUCTURE và phân tích PCA. Hệ số cận huyết (F_{IS}) được xác định bằng GenoDive v2.0b27 dựa vào kết quả phân nhóm quần thể được xác định bằng STRUCTURE và PCA. Đồng thời, mức độ khác biệt bên trong, giữa các quần thể và giữa các phân nhóm của ba loài cá điển hình được ước lượng qua phân tích phương sai phân tử (AMOVA) bằng Alrequin v3.5. Sự liên kết của các quần thể cá theo khoảng cách địa lý được kiểm tra theo phương pháp: thử nghiệm Mantel test và phân tích bản đồ vectơ eigen Moran dựa trên khoảng cách (db-MEM).

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

Nhằm kiểm chứng giả thuyết về cấu trúc quần đàn theo các hệ thống di cư đã được ghi nhận, nghiên cứu tiến hành dự đoán mô hình di cư của 02 loài cá có tập tính di cư khác nhau (cá ét mọi-di cư chặng ngắn, tùy nghi và cá vồ đếm-di cư chặng dài). Dữ liệu thông tin di truyền (chủ yếu là tần số alen) được phân tích dựa trên 2 phương pháp

gồm mô hình di cư hiện tại (divMigrate) và lịch sử phân tách và pha trộn quần thể (Treemix v1.13).

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

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

Sử dụng các đoạn đọc chất lượng cao, hệ gen ti thể của cá ét mọi được lắp ráp và chú giải theo quy trình MitoZ v3.4. Các gen mã hóa protein, RNA vận chuyển và RNA ribosome được chú giải lần lượt bằng GeneWise v2.2, MiTFi v1.0 và infernal v1.1.1. Công cụ BWA v0.7.17 và Samtools v1.15.1 được sử dụng để sắp xếp các gen theo trật tự của hệ gen ti thể và cấu trúc dạng vòng được thể hiện bằng phần mềm Circos. Hệ gen ti thể được đăng ký mã số trên Ngân hàng Gen.

2.4.2. Xác định trình tự mã vạch RAD thuộc hệ gen ti thể và khảo sát đa dạng di truyền và cấu trúc quần thể của cá ét mọi ở LMB

Trình tự mã vạch RAD của các quần thể cá ét mọi được xác định theo quy trình radBARCODER. Sử dụng dữ liệu mã vạch RAD, nghiên cứu khảo sát sự đa dạng di truyền của các quần thể cá ét mọi thông qua các thông số như tổng số haplotype (N_h), đa dạng haplotype (H_d), đa dạng nucleotide (π) và số lượng của vị trí đa hình (S) bằng DnaSP v5. Sự khác biệt di truyền giữa các cặp quần thể cá ét mọi (F_{ST}) được xác định bằng Alerquin v3.5. Mạng lưới haplotype được xây dựng bằng thuật toán Templeton Crandall and Sing trên phần mềm PopART.

CHƯƠNG 3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Lắp ráp *de novo* hệ gen của ba loài cá điển hình ở LMB 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

Tổng cộng **272** thư viện DNA cá chạch lá tre, **255** thư viện cá ét mọi và **192** thư viện cá vồ đếm được giải trình tự và thu được lần lượt là

36.447.827, 1.062.049.264, và 313.556.27 đoạn đọc thô. Sau khi loại bỏ các nucleotide chất lượng thấp, adapter và đoạn đọc có chiều dài nhỏ hơn 50 bp, số lượng đoạn đọc chất lượng cao được thu nhận 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

Công cụ dDocent - sử dụng thuật toán tham lam được lựa chọn để lắp ráp *de novo* hệ gen của các loài cá. Kết quả hệ gen lắp ráp của cá chạch lá tre, cá ét mọi và cá vồ đém có kích thước lần lượt là **30.372.557 bp** (tỉ lệ GC là 36,1%), **32.975.632 bp** (tỉ lệ GC là 35,9%) và **35.729.239 bp** (tỉ lệ GC là 35,1%).

3.1.3. Xác định chỉ thị SNPs của ba loài cá điển hình ở LMB

Sau khi phát hiện và sàng lọc qua các thông số, nghiên cứu tuyển chọn được **4.237 SNPs** của 239 cá thể (9 quần thể) cá chạch lá tre, **825 SNPs** của 232 cá thể (9 quần thể) cá ét mọi và **1.270 SNPs** của 160 cá thể (7 quần thể) cá vồ đém. Kiểm định các loci ngoại vi bằng BayeScan và Lositan, nghiên cứu thu được tập dữ liệu trung tính gồm **3.736 SNPs** ở cá chạch lá tre, **760 SNPs** ở cá ét mọi và **1.176 SNPs** ở cá vồ đém. Tập dữ liệu các loci đáp ứng gồm **189, 23 và 11 loci** lần lượt ở cá chạch lá tre, ét mọi và vồ đém cũng được thu nhận.

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 LMB

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.

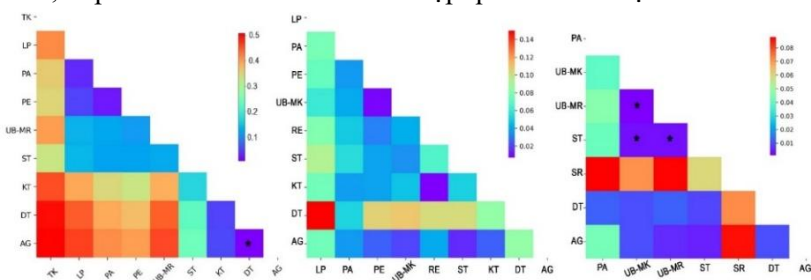
3.2.1. Sự đa dạng di truyền của các quần thể cá

Kết quả khảo sát đa dạng di truyền thể hiện giá trị dị hợp tử quan sát đều bằng hoặc nhỏ hơn dị hợp tử mong đợi ở tất cả các quần thể cá. Sự đa dạng di truyền cao nhất được thể hiện ở cá vồ đém (Ho/He = 0,220/0,266), tiếp theo là cá ét mọi (Ho/He = 0,185/0,233) và cuối cùng là cá chạch lá tre (Ho/He = 0,140/0,176) (**Bảng 3.4**).

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

Sự khác biệt di truyền chung (G_{ST}) thể hiện giá trị cao nhất ở cá chạch lá tre ($G_{ST}=0,303$), tiếp theo ở cá ét mọi ($G_{ST} = 0,038$) và thấp nhất ở cá vồ đém ($G_{ST} = 0,003$) và đều có ý nghĩa thống kê ($P < 0,001$).

Sự khác biệt di truyền theo cặp (F_{ST}) ở **Hình 3.3** thể hiện 27/28 cặp quần thể cá chạch lá tre, 36/36 cặp quần thể cá ét mọi, 18/21 cặp quần thể cá vồ đém thể hiện sự khác biệt có ý nghĩa thống kê. Các quần thể cá chạch lá tre ở dưới thác Khôn thể hiện sự khác biệt di truyền cao hơn so với phía trên thác Khôn. Cặp quần thể cá ét mọi Đồng Tháp và Luang Prabang thể hiện sự khác biệt di truyền lớn so với các cặp quần thể còn lại. Quần thể cá vồ đém ở Siêm Riệp thể hiện sự khác biệt cao nhất, tiếp theo là Paksan so với các cặp quần thể còn lại.



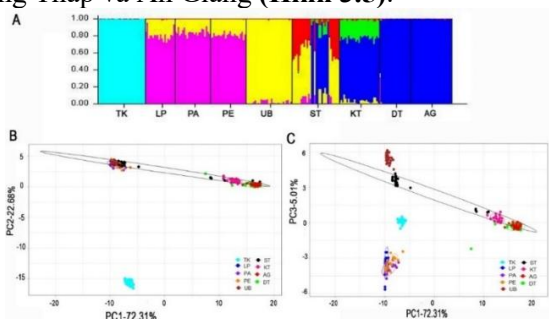
Hình 3.3. Biểu đồ heatmap thể hiện giá trị khác biệt di truyền theo cặp quần thể (F_{ST}) của cá chạch lá tre (A), cá ét mọi (B) và cá vồ đém (C). * chỉ ra sự khác biệt không có ý nghĩa thống kê sau khi hiệu chỉnh FDR

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

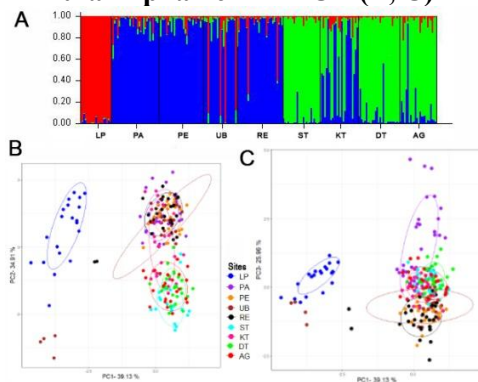
Vị trí trên sông Mekong			Địa điểm thu mẫu	Macrognathus siamensis			Labeo chrysophekadion			Pangasius larnaudii		
				Nse	Ho	He	Nse	Ho	He	Nse	Ho	He
UMB		Dòng nhánh	TK	32	0,140	0,140						
LMB	Thượng lưu -LMB	Dòng nhánh (Sông Khan)	LP	20	0,128	0,141	20	0,193	0,208			
	Trung lưu -LMB	Dòng chính	PK	24	0,110	0,142	31	0,243	0,248	20	0,263	0,298
			PE	24	0,108	0,146	29	0,217	0,245			
			UB-MK			22	0,214	0,245	26	0,259	0,271	
		Dòng nhánh (Sông Mun)	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	Hợp lưu giữa dòng chính và 3S	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	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/Trung bình				239	0,140	0,176	232	0,185	0,233	160	0,220	0,266

Ghi chú: Kí hiệu của các quần thể thu mẫu thể hiện ở bảng 3.1; Nse: Số lượng mẫu phân tích, Ho: Dị hợp tử quan sát, He: Dị hợp tử mong đợi

Dựa trên phân tích nhóm STRUCTURE và PCA, các quần thể cá chạch lá tre phân thành 5 nhóm, gồm: (1) Tachileik, (2) Luang Prabang, Paksan và Pakse, (3) Ubon Ratchathani thuộc sông Mun, (4) sự trộn lẫn thông tin di truyền từ các nhóm ở Strung Treng, và (5) Kratié đến ĐBSCL (An Giang và Đồng Tháp) (**Hình 3.4**). Trong khi đó, 3 phân nhóm cá ét mọi được ghi nhận như sau: (1) Luang Prabang, (2) Paksan, Pakse, Ubon Ratchathani và Roi Et, và (3) Stung Treng, Kratié, Dong Thap và An Giang (**Hình 3.5**).

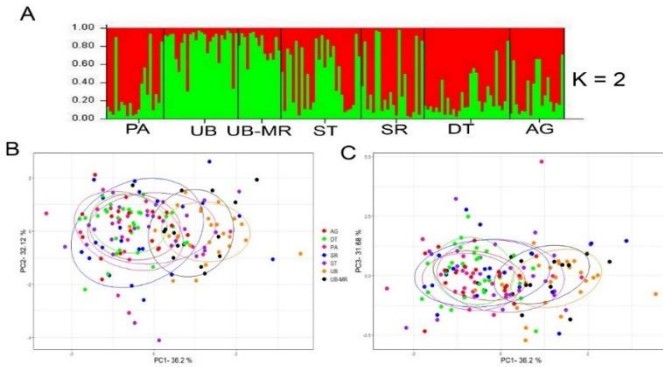


Hình 3.4. Cấu trúc quần thể cá chạch lá tre dựa trên 1,936 SNPs trung tính bằng phân tích nhóm Structure (A) và phân tích thành phần chính PCA (B, C)



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

Ở cá vồ đém, sự phân tách không rõ ràng của các quần thể được thể hiện, chứng tỏ có sự kết nối cao giữa các quần thể (**Hình 3.6**).



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

Về hệ số cận huyết, quần thể cá chạch lá tre ở UB-MR thể hiện giá trị cao nhất ($F_{IS} = 0,61$) và hạ lưu-LMB là thấp nhất ($F_{IS} = 0,40$). Trong khi đó, các phân nhóm quần thể cá ét mọi thể hiện hệ số cận huyết tương tự nhau, dao động từ 0,225 ở thượng lưu-LMB, tiếp theo là 0,245 ở trung lưu-LMB và 0,299 ở hạ lưu-LMB. Hệ số cận huyết được ước tính chung cho tất cả các quần thể cá vồ đém đạt 0,173.

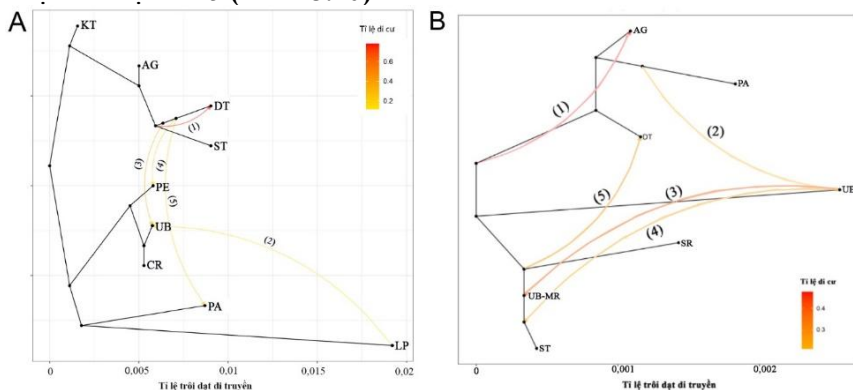
Phân tích Mantel test thể hiện mối tương quan có ý nghĩa thống kê giữa sự tăng khoảng cách địa lý dẫn đến sự khác biệt di truyền của các quần thể cá chạch lá tre ($P = 0,004$) và cá ét mọi ($P = 0,0005$), trong khi đó, các quần thể cá vồ đém không thể hiện mối tương quan này ($P = 0,4201$). Đối với phân tích dbMEM, trục MEM-1 thể hiện sự phân tách cấu trúc không gian của các quần thể ở cả 03 loài cá ($P < 0,01$), trong khi đó, các trục từ MEM 2-6 không ghi nhận được ($P > 0,05$).

Đối với cá chạch lá tre, phân tích AMOVA thể hiện sự khác biệt di truyền đáng kể giữa hai nhóm quần thể ở trên và dưới thác Khôn

(31,5%; $P = 0,03$), trong khi đó, các cấp độ khác (giữa các cá thể, giữa các cá thể trong quần thể và giữa các quần thể trong các nhóm) thể hiện tỉ lệ lần lượt là 49,1%; 9,5% và 9,8%, và tất cả các sự khác biệt đều có ý nghĩa thống kê ($P < 0,001$). Ở cá ét mọi, kết quả AMOVA của hai giả thuyết đều cho thấy nguồn biến dị chủ yếu từ sự khác biệt giữa các cá thể (50% và 84,49%) và có ý nghĩa thống kê ($P < 0,001$). Ở cá vồ đém, kết quả ghi nhận 57% khác biệt ở giữa các cá thể, 39% giữa các cá thể trong quần thể và chỉ 3% giữa các quần thể. Phân tích AMOVA cho thấy sự khác biệt di truyền không đáng kể giữa 2 nhóm (0,4%) và không có ý nghĩa thống kê ($P > 0,05$).

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

Sử dụng phương pháp divMigrate, sự di cư xuôi dòng, ngược dòng và vượt thác Khôn đều được ghi nhận ở cá ét mọi và cá vồ đém, đồng thời, tỉ lệ di cư mạnh mẽ hơn được thể hiện ở cá vồ đém. Sử dụng phân tích Treemix, số sự kiện di cư tối ưu của loài cá ét mọi và vồ đém đều được xác định là 5 (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 vồ 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

Ở cá ét mọi, (1) sự kiện di cư xuôi dòng từ điểm trung gian phía trên AG đến DT, (2) là di cư xuôi dòng từ LP đến UB, và ba sự kiện di cư ngược dòng từ điểm trung gian giữa DT – ST lên các quần thể phía trên thác Khôn gồm UB (3), PE (4) và PA (5). Ở cá vồ đém, (1) di cư xuôi dòng từ điểm trung gian trên và dưới thác Khôn đến AG, (2) di cư xuôi dòng từ điểm trung gian ở PA đến UB; (3) di cư ngang và ngược dòng từ UB đến UB-MR; (4) di cư ngang và xuôi dòng từ UB đến điểm trung gian của UB-MR và ST; và (5) di cư xuôi dòng từ điểm trung gian trên và dưới thác Khôn đến DT (**Hình 3.10**).

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 của cá ét mọi

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

Hệ gen ti thể của cá ét mọi thu thập ở Ubon Ratchathani lắp ráp thành công có kích thước là 16.600 bp (tỉ lệ GC là 42,9%) và thể hiện tỉ lệ tương đồng 99,8% với hệ gen ti thể cùng loài trên Ngân hàng Gen. Hệ gen này 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). Trình tự của hệ gen được đăng ký trên ngân hàng Gen với mã số là OR637878.

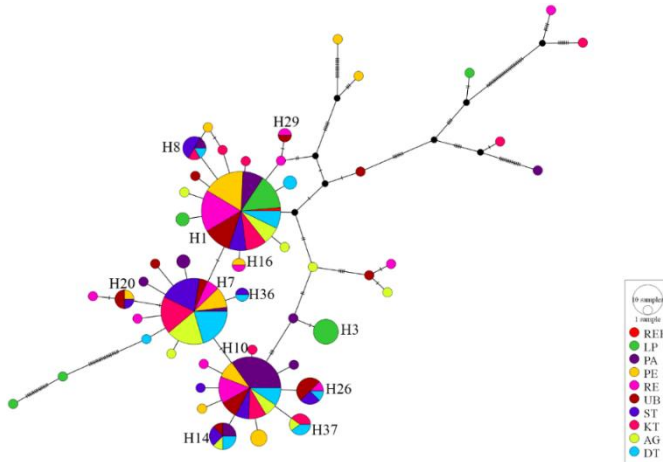
3.3.2. Xác định trình tự mã vạch RAD, khảo sát đa dạng di truyền và cấu trúc quần thể của cá ét mọi ở LMB

Nghiên cứu xác định được tập hợp các vùng mã vạch RAD có kích thước là 757 bp. Trong đó, 502 bp (66,2%) thuộc 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}*).

Tổng cộng 49 haplotype (chiếm 19,8%) được ghi nhận sử dụng dữ liệu tập hợp mã vạch RAD (757 bp) từ 247 cá thể cá ét mọi, dao động từ 6 haplotype/ 22 cá thể (27,3%) ở Luang Prabang đến 12 haplotype/

30 cá thể (40%) ở Roi Et. Kết quả thể hiện sự đa dạng haplotype cao với $H_d=0,849\pm0,014$, dao động từ $0,71\pm0,071$ (Luang Prabang) đến $0,871\pm0,046$ (Ubon Ratchathani). Sự khác biệt di truyền theo cặp thấp ($F_{ST}=0 - 0,17$) được ghi nhận giữa các quần thể cá ét mọi và không có sự khác biệt, ngoại trừ quần thể Luang Prabang.

Mạng lưới haplotype thể hiện sự kết nối di truyền cao giữa các quần thể cá ét mọi ở LMB. Trong đó, haplotype H1 thể hiện sự chia sẻ chung của tất cả các quần thể, hai haplotype (H7 và H10) được tìm thấy ở 8/9 quần thể (ngoại trừ Luang Prabang) và haplotype H14 chia sẻ bởi 5/9 quần thể. Haplotype H8 và H26 được chia sẻ bởi 4/9 quần thể và các haplotype H16, H20, H29, H36 và H37 ghi nhận các trình tự từ 2 – 3 quần thể. Số lượng lớn các haplotype đặc trưng (35/49) được tìm thấy ở tất cả quần thể (**Hình 3.13**).



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 từng quần thể thu mẫu và trình tự tham chiếu từ Ngân hàng Gen.

KẾT LUẬN VÀ KIẾN NGHỊ

Kết luận

1. Nghiên cứu đã thu thập và tạo thư viện từ ba loài cá điển hình của LMB: **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 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 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 chằng 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 chằng 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 mọi (*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 mọi ở lưu vực sông Mekong (LMB). 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

❖ 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 được ghi nhận (04 vùng cho cá chạch lá tre, 03 vùng cho cá ét mọi và 01 vùng cho cá vồ 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.

❖ 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 MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

1. Population genetics of *Macrognathus siamensis* (Synbranchiformes: Mastacembelidae): Implications for non-migratory fishery resources in the Mekong River basin. **Fisheries Research**, **281**:107210, (2025). <https://doi.org/10.1016/j.fishres.2024.107210>.

2. Population genetics and the importance of migration in the facultative migratory fish, *Labeo chrysophekadion*, in the Lower Mekong Basin. **Conservation Genetics**, **26**, 307–318, (2025). <https://doi.org/10.1007/s10592-024-01668-w>.

3. 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, (2024). <https://doi.org/10.32508/stdj.v26iSI.4199>.

4. 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, (2022). ISBN 978-625-8246-59-9.

5. 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, (2024). ISBN 978-604-489-393-8.

6. Genetic Differentiation and Isolation by Distance in Mekong River Fishes with Typical Migration Patterns. **Asian Fisheries Science** **38**: 131–142, (2025). <https://doi.org/10.33997/j.afs.2025.38.3.001>.