SỬ DỤNG CÔNG CỤ GIẢI TRÌNH TỰ ddRAD TRUY XUẤT NGUỒN GỐC DI TRUYỀN CỦA CÁ CHÉP CHÂU ÂU (*Cyprinus carpio carpio*) VÀ CÁ CHÉP AMUR (*Cyprinus carpio haematopterus*) LÊN QUẦN THỂ CÁ CHÉP TỰ NHIÊN CỦA VIỆT NAM

NGUYỄN VĂN QUÂN [1, 2], NEDOLUZHKO A. V. [3], PHẠM THẾ THƯ [1], NGUYỄN ĐỨC THẾ [1], PHẠM VĂN CHIẾN [1], RASTORGUEV S. M. [4], VŨ QUYẾT THÀNH [5]

1. ĐẶT VẤN ĐỀ

Theo dữ liệu hình thái học, bốn phân loài của cá chép phổ biến trên thế giới đã được xác định là *C. carpio carpio* (cá chép Châu Âu hoặc cá chép Ponto-Caspian), *C. carpio haematopterus* (lưu vực sông Amur phía nam Trung Quốc), *C. carpio viridiviolaceus* (cá chép phổ biến phía Bắc Việt Nam) và *C. carpio aralensis* (lưu vực sông vùng Trung Á) [1]. Các nghiên cứu di truyền dựa trên phân tích ADN ti thể và microsatellite đã không phân biệt được phân loài *C. c. aralensis,* vì nó rất giống với phân loài *C. carpio carpio* về mặt di truyền [2, 3].

Cá chép là loài rất đa dạng, nó bao gồm nhiều dạng quần thể khác nhau, biến đổi ở những nơi sinh sống khác nhau. Cá chép đã phân bố ở hầu hết các thủy vực có điều kiện môi trường thích hợp. Cá chép hiện nay là một trong những loài cá nước ngọt được nuôi rộng rãi nhất trên thế giới (FIGIS, 2020). Tổng sản lượng nuôi trồng cá chép tăng mạnh trong thời gian gần đây. Theo dữ liệu của FAO [4], tổng sản lượng của loài này đạt 4556621 tấn trong năm 2016, chiếm khoảng 10% sản lượng cá nuôi trên thế giới.

Ở Việt Nam, khai thác và nuôi trồng cá chép đóng vai trò kinh tế quan trọng, hiện có tám dòng cá chép địa phương và ba dòng cá chép *C. carpio*. Các dòng cá địa phương khác nhau về hình thái và màu sắc, như các dòng cá chép vảy trắng ở Bắc Cạn, Hồ Tây, Nam Hải Vân, dòng vây đỏ, tím, thân rộng và các loại vảy rải rác [5, 6]. Các dòng cá chép *C. carpio* gồm cá chép vảy Hungary, cá chép Hungary và vàng Indonesia. Cá chép vàng Indonesia du nhập vào Việt Nam trước năm 1975, lai với dòng *C. carpio* của Hungary du nhập năm 1970, và lai với dòng có vảy Hungary vào năm 1975. Trong điều kiện môi trường ở Việt Nam, cá chép Hungary dễ mắc bệnh và tỷ lệ sống thấp [5]. Do khả năng sống của cá chép Châu Âu ở Việt Nam thấp và tốc độ tăng trưởng chậm của các dòng cá chép địa phương, các dòng cá chép Hungary và Indonesia đã được sử dụng cho các chương trình lai tạo và chọn lọc hàng loạt với cá chép địa phương của Việt Nam. Để tạo ra con lai đó, việc lai tạo kết hợp với chọn lọc đã được thực hiện giữa ba nguồn gen - cá chép trắng Việt Nam, cá vảy Hungary và cá chép vàng Indonesia [5, 6].

Có sự quan tâm lớn về cơ sở khoa học và tầm quan trọng thực tế trong việc xác định và mô tả đặc điểm của cá di nhập bị thoát khỏi trang trại nuôi và con cháu của chúng trong môi trường tự nhiên ở Việt Nam. Trước đây, đã có một số công bố nghiên cứu quần thể cá chép Việt Nam bằng phương pháp phân tử như dùng chỉ thị SSCP, các đoạn DNA ty thể và các locus microsatellite [6, 7, 8], nhưng các phương

pháp này không đủ mạnh để xác định được chính xác tỷ lệ tác động của các nguồn gen. Gần đây, phương pháp phân tử dựa trên công nghệ giải trình tự thế hệ mới (NGS) cho phép ước tính chính xác các thông số di truyền, cũng như tỷ lệ nguồn gen lai nhập, trạng thái lai... Phương pháp này còn cho phép xác định hàng trăm nghìn locus SNP trong một mẫu và nó cho phép có thể phát hiện được số lượng xâm nhập di truyền rất nhỏ.

Một trong những phương pháp hiện đại là giải trình tự DNA liên kết với vị trí giới hạn (RAD-sequencing), cho phép sử dụng các ưu thế của công nghệ NGS cho các nghiên cứu trên toàn bộ quần thể [9]. Gần đây, các nhà khoa học đã cải tiến phương pháp RAD-sequencing cho phép ghép kênh mẫu ở quy mô lớn được gọi là giải trình tự ddRAD [10].

Một trong những lý do để ước tính tỷ lệ lai thành công trong nghiên cứu này là trên thực tế các giống cá chép bản địa và di nhập thuộc các phân loài khác nhau: các dòng bản địa thuộc phân loài *C.c. viridiviolaceus* (cá chép Việt Nam); các dòng Hungary thuộc phân loài Châu Âu (hoặc Ponto-Caspian) - *C.c.carpio,* và cá chép vàng Indonesia, theo trình tự vùng trung tâm ty thể và DNA microsatellite đã được công bố bởi tác giả Thai và cộng sự [6] là *C.c. haematopterus.* Tất cả các phân loài này đều có thể phân biệt rõ ràng dựa trên trình tự ddRAD.

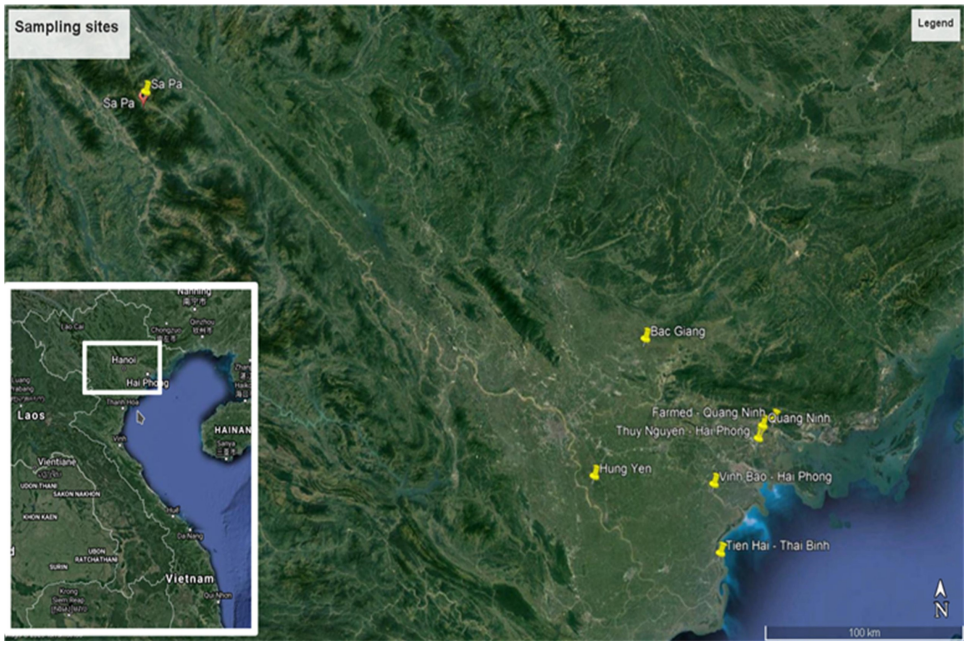
Mục tiêu của nghiên cứu là ước tính được tỷ lệ phối trộn gen giữa cá chép di nhập và cá chép nuôi thoát khỏi trại nuôi đối với quần thể cá chép tự nhiên của Việt Nam bằng phương pháp giải trình tự ddRAD. Kết quả nghiên cứu có thể làm sáng tỏ sự tác động di truyền của các giống cá chép khác nhau đối với cá chép tự nhiên Việt Nam và có thể đưa ra các biện pháp khai thác, bảo tồn các quần thể cá chép. Ngoài ra, kết quả có thể được sử dụng để phát triển các chỉ thị phân tử SNP đặc hiệu nhằm kiểm soát cá thoát ra khỏi trại nuôi và giám sát cá nuôi đã thoát ra ngoài.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

Các mẫu cá từ 5 quần thể được thu thập ngoài tự nhiên và từ 3 trại nuôi cá ở miền Bắc Việt Nam. Vị trí của các điểm thu thập mẫu được thể hiện trên hình 1. Số lượng mẫu vật cho mỗi quần thể được thể hiện trong bảng 1. Tất cả các mẫu cá sau khi thu ngoài hiện trường đã được đông lạnh ngay và bảo quản trong đá khô.

Năm mẫu vật của dòng cá chép nội địa Tata đã được sử dụng như là mẫu đối chiếu cho cá chép Hungary đã được du nhập vào môi trường tự nhiên Việt Nam vào những năm 1970. Các mẫu cá chép Hungary này được lấy từ ngân hàng gen Cá chép sống của Trung tâm Nghiên cứu và Đổi mới Nông nghiệp Quốc gia, Viện Nghiên cứu và Nuôi trồng Thủy sản (HAKI), Szarvas, Hungary. Ngoài ra, 5 mẫu cá chép Amur được lấy từ Bộ sưu tập Vật liệu Di truyền Tham khảo Quốc gia Nga (RNCRGM) của Viện Nghiên cứu Thủy sản và Hải dương học Liên bang Nga (VNIRO), Moscow, Nga. Tổng số có 215 mẫu vật đã được sử dụng trong nghiên cứu.



Hình 1. Vị trí và vùng thu mẫu cá ở miền Bắc Việt Nam

Bảng 1. Các mẫu cá chép của Việt Nam

TT	Tên nhóm mẫu	Thời gian thu	Địa điểm	Mẫu	Vị trí thu mẫu
1	Mẫu 4,1 (4) - 44 cá thể	4.6.2019	Tiền Hải, Thái Bình	Tự nhiên	20°28'28.90"N - 106°35'14.36"E
2	Mẫu 4,2 - 28 cá thể	4.6.2019	Vĩnh Bảo, Hải Phòng	Tự nhiên	20°45'11.10"N - 106°33'28.60"E
3	Mẫu 6,4 - 9 cá thể	6.6.2019	Quảng Ninh	Tự nhiên	20°59'5.70"N - 106°47'33.84"E
4	Mẫu 6,5 - 9 cá thể	6.6. 2019	Quảng Ninh	Nuôi	21° 0'24.66"N - 106°50'45.56"E
5	Mẫu 6,6 - 8 cá thể	7.6.2019	Thủy Nguyên, Hải Phòng	Tự nhiên	20°56'2.55"N - 106°46'16.20"E
6	Mẫu 6,7 - 21 cá thể	7.6.2019	Sa Pa	Nuôi	22°19'58.70"N - 103°50'38.83"E
7	Mẫu 6,8 - 14 cá thể	2.6.2019	Hưng Yên	Tự nhiên	20°47'24.35"N - 105°59'24.03"E
8	Mẫu Lai F2 - 20 cá thể	5.6.2019	Bắc Giang	Nuôi	21°20'18.56"N - 106°14'7.91"E
9	Mẫu Lai F3 - 22 cá thể	5.6.2019	Bắc Giang	Nuôi	21°20'18.56"N - 106°14'7.91"E
10	Mẫu Lai F5 - 19 cá thể	5.6.2019	Bắc Giang	Nuôi	21°20'18.56"N - 106°14'7.91"E
11	Mẫu 3X con lai F3 - 12 cá thể	5.6.2019	Bắc Giang	Nuôi	21°20'18.56"N - 106°14'7.91"E

2.2. Phương pháp nghiên cứu

2.2.1. Thu mẫu mô, tách chiết DNA và giải trình tự

Các mẫu mô vây cá được thu và ngâm bảo quản trong etanol 99%. DNA từ mẫu mô vây được tách chiết bằng enzyme proteinase K ở 50°C trong 16÷20 giờ, sau đó được tinh sạch bằng phenol-cloroform, kết tủa bằng etanol và bảo quản trong ddH_2O vô trùng [11]. DNA tinh sạch được định lượng bằng máy Qubit 2.0 (Invitrogen, Carlsbad, USA) và tính toàn vẹn của sản phẩm DNA được đánh giá thông qua điện di trên gel agarose.

Quy trình chuẩn bị thư viện theo các nguyên tắc chung của phương pháp quaddRAD [10]. DNA hệ gen được phân cắt bằng endonuclease giới hạn MspI và PstI với các adapter có 6 cặp base (bp) trong các trình tự chỉ thị và 4 cặp base ngẫu nhiên để loại bỏ bản sao PCR. Bước cắt nhỏ (digestion step) được tiến hành với ligase. Các thư viện ddRAD được tổng hợp với số lượng bằng nhau và được khuếch đại bằng cách sử dụng các đoạn mồi DNA đặc hiệu. Kích thước hạt Agencourt Ampure XP (Beckman Coulter, Mỹ) được sử dụng để lựa chọn các trình tự có chiều dài từ 600 đến 800 nucleotide cho thiết lập thư viện ddDNA. Việc chuẩn bị thư viện trình tự Illumina được thực hiện bởi "NEBNext® Ultra ™ II DNA Library Prep Kit" cho Illumina (NEB, Hoa Kỳ) theo hướng dẫn của nhà sản xuất.

Lưu lượng S2 của máy phân tích trình tự Illumina Novaseq6000 (Illumina, San Diego, CA) với các lần đọc đầu cuối chuỗi kép (độ dài 2 × 150 bp) được sử dụng trong trình tự của thư viện ddRAD.

2.2.2. Bản đồ đọc ddRAD, SNP và phân tích kiểu gen

Các lần đọc Illumina thô được phân kênh theo mã vạch bên ngoài và bên trong của chúng và được bản đồ hóa qua so sánh với bộ gen của cá chép có mã genome tham chiếu RefSeq: GCF_000951615.1 [12] bằng phần mềm Bowtie2 [13] với bộ tham số "very sensitive". Dữ liệu bản đồ hóa ở định dạng SAM được chuyển đổi sang định dạng nhị phân (BAM) và được sắp xếp lại; sau đó, những dữ liệu này được chỉ số hóa bằng cách sử dụng công cụ Samtools v 0.1.19. Độ rộng vùng phủ cũng được ước tính bằng công cụ độ phủ của samtools v 0.1.19 [14].

SNP được thực hiện bằng phần mềm Bcftools v 1.9 [14] với giá trị chất lượng tối thiểu là 30 (tham số --min-BQ). Các tham số của Bcftools được xử lý theo trình tự: mpileup --redo-BAQ --min-BQ 30 -per-sample-mF --annotate DP và sau đó hoàn thiện các bước: bcftools call -multiallelic-caller --variants-only -Ov. Các tệp VCF thu được bằng cách sử dụng Bcftools đã được lọc bởi tham số chất lượng kiểu gen để loại bỏ các locus có chất lượng kiểu gen trung bình dưới 30. File VCF đã lọc được tải vào phần mềm thống kê R (www.r-project.org) bằng package vcfR [15]. File VCF đã được lọc bổ sung theo thông số chất lượng kiểu gen trong vcfR [15]. Giá trị chất lượng thu được bằng lệnh getQUAL. Các locus có chất lượng tốt nhất (999) được sử dụng cho các bước tiếp theo. Các locus mà xuất hiện trong ít hơn 10% mẫu vật bị loại khỏi tập dữ liệu bằng cách sử dụng tập Perl tùy chỉnh. Các locus SNP đã được chuyển đổi thành định dạng genlight bằng package adegenet trong phần mềm R [16]. Phân tích thành phần chính (PCA) được thực hiện bằng cách sử dụng hàm glPca trong package adegenet.

Việc truy xuất tổ tiên riêng lẻ cho từng mẫu được thực hiện bằng cách sử dụng phần mềm NGSadmix [17]. Với mục đích thu được file kiểu gen ở định dạng beagle bằng cách sử dụng phần mềm ANGSD với việc sử dụng các file ở trên và xử lý bởi NGSadmix với số lượng quần thể (tham số -K) bằng 3, tương ứng với số lượng phân loài cá chép nghiên cứu. Hình ảnh hóa bằng cách sử dụng hàm barplot trong phần mềm thống kê R. Để hình dung tổ tiên của ba phân loài trong quần thể cá chép tự nhiên Việt Nam, chúng tôi đã xây dựng ba ô về phân bố nguồn gốc tổ tiên của từng phân loài được tính toán bằng phần mềm ANGSD cho từng quần thể cá tự nhiên.

2.3. Xác định các locus xâm nhập

Chúng tôi chọn các locus khác nhau về thành phần alen giữa phân loài cá chép Amur và cá chép Việt Nam để xác định locus gen của cá chép tự nhiên Việt Nam (phân loài cá chép Việt Nam) đã được xâm nhập vào hệ gen cá chép nuôi ở Việt Nam (phân loài cá chép Amur). Bộ dữ liệu RAD-seq đã công bố trước đây từ năm mẫu cá chép Amur tự nhiên [18] cũng như năm mẫu cá chép thu tại Vĩnh Bảo của Việt Nam có tác động di truyền thấp nhất từ cá chép Amur đã được sử dụng cho phân tích. 5 mẫu cá chép Vĩnh Bảo của Việt Nam được lựa chọn từ 28 mẫu dựa trên cơ sở không có xáo trộn di truyền với các dòng cá chép Amur và cá chép Châu Âu. Sau đó, chúng tôi phân loại dựa trên sự khác nhau giữa các locus, chỉ giữ lại những alen có ở cá chép Vĩnh Bảo nhưng không có trong các mẫu cá chép Amur tự nhiên.

Tất cả 84 mẫu vật của quần thể cá chép nuôi ở Quảng Ninh và Bắc Giang được sử dụng cho phân tích xác định sự xâm nhập di truyền của chúng vào cá chép tự nhiên của Việt Nam trong nuôi trồng các dòng cá chép tại địa phương. Đồng thời, quần thể cá chép nuôi ở Sa Pa đã bị loại khỏi phân tích này do có sự xâm nhập lớn về mặt di truyền từ phân loài cá chép Châu Âu. Chúng tôi đã tìm các locus khác biệt của cá chép ở Vĩnh Bảo trong các mẫu cá nuôi của Việt Nam và đưa tất cả các điểm trùng khớp SNP vào một tệp BED. Chỉ cần có thay thế của một alen đã được tính nhưng loại bỏ các đồng hợp tử/dị hợp tử.

SNPs cá chép tự nhiên Việt Nam được tìm thấy ở các quần thể nuôi ở Quảng Ninh và Bắc Giang được phân tích bằng cách sử dụng dữ liệu bộ gen cá chép thường (GCF_000951615.1) và bedtools intersectBed [19] để xác định vị trí các locus xâm nhập trong hệ gen *C. carpio* và xác định chính xác gen nào của cá chép đã bị ảnh hưởng. Danh sách tất cả các gen có alen "cá chép tự nhiên Việt Nam-wild Vietnamese carp" được kết hợp bằng cách sử dụng hệ thống tùy chỉnh - Perl script. Các gen được phân bố theo số lượng mẫu nuôi có chứa alen "cá chép tự nhiên Việt Nam".

3. KẾT QUẢ VÀ BÀN LUẬN

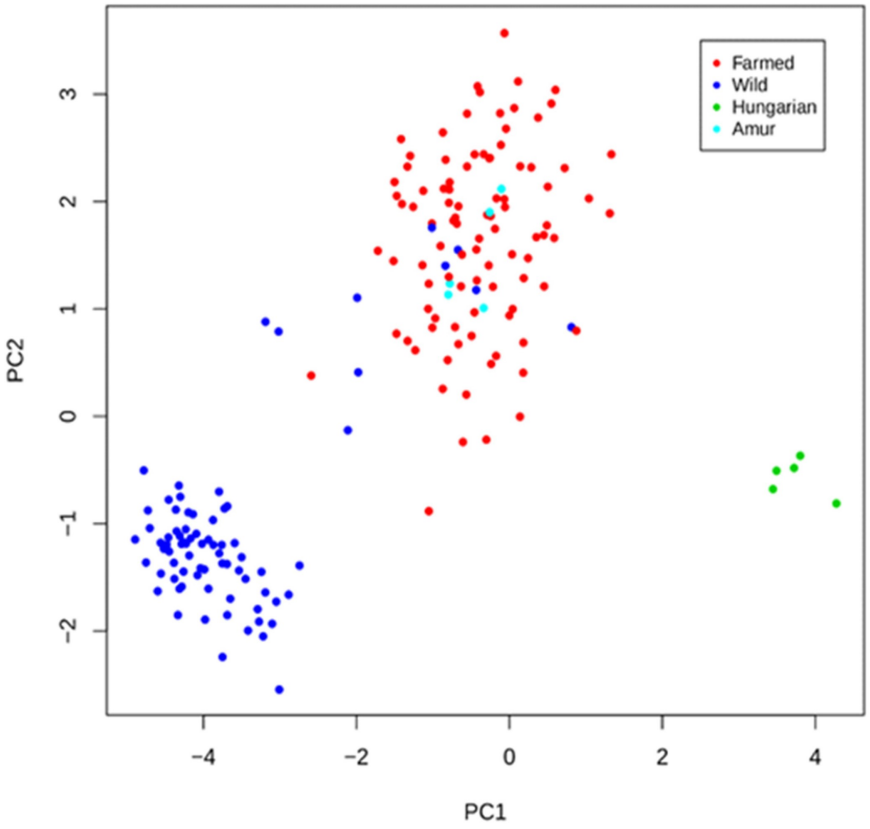
3.1. Kết quả phân tích các mẫu vật

Tổng cộng 205 mẫu cá chép Việt Nam và 10 mẫu cá chép Amur và Hungary đã được sử dụng trong phân tích. Tổng số lần đọc trình tự ddRAD với chiều dài 150 nucleotide là 1495640676. Thống kê và lập bản đồ trình tự cho từng nhóm cá chép được trình bày trong bảng 2.

Bảng 2. Số lượng mẫu, trình tự đọc và số lần đối chiếu

Quần thể	Số lượng mẫu	Tổng số lần đọc	Số lần đối chiếu
Cá chép tự nhiên Việt Nam	100	712 976 882	376 855 876
Cá chép nuôi Việt Nam (thuộc phân loài cá chép Amur)	105	740 036 570	420 954 370
Cá chép Hungary	5	17 650 186	12 317 233
Cá chép Amur	5	24 977 038	15 734 054

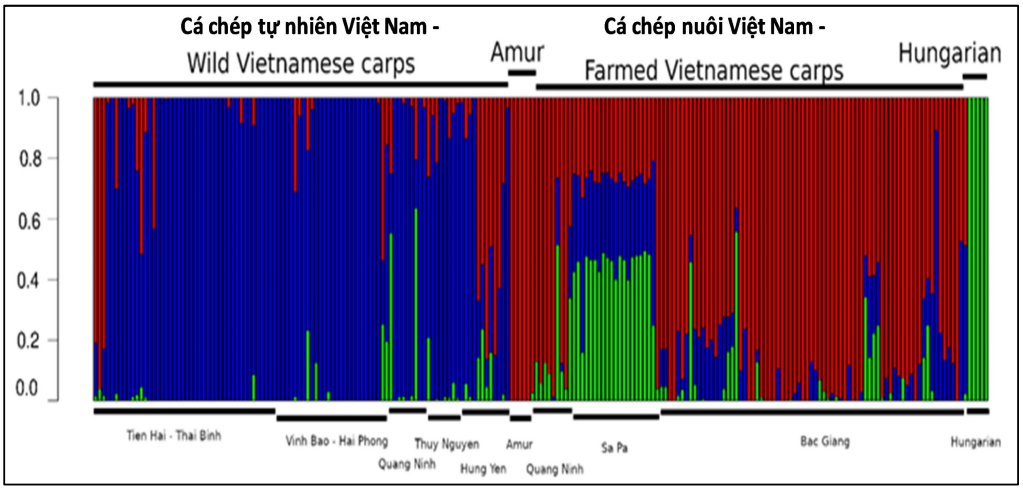
Khi dữ liệu về kiểu gen được lọc và chỉ giữ lại những locus mà hầu hết các mẫu có kiểu gen được xác định, và có 4515 locus đã được dùng cho các phân tích tiếp theo. Phân tích thành phần chính (PCA) đã được sử dụng nhằm xác định cấu trúc quần thể của các mẫu cá được nghiên cứu. Kết quả của phân tích PCA được thể hiện trên hình 2.



Hình 2. Phân tích thành phần chính (PCA) các mẫu cá chép nuôi (màu đỏ), cá chép tự nhiên (màu xanh nước biển) của Việt Nam, cá chép Hungary (màu xanh lá cây) và cá chép Amur (màu xanh dương)

Phân tích thành phần chính (PCA) (hình 2) cho thấy có ba nhóm phân loại riêng biệt, trong đó nhóm 1 là cá chép Tata Hungary, nhóm 2 là cá chép tự nhiên của Việt Nam và nhóm thứ 3 là các mẫu khác, bao gồm cả cá chép Amur tự nhiên (sông Amur, Nga), các mẫu cá chép từ 3 trại nuôi và một số mẫu cá chép tự nhiên của Việt Nam. Từ kết quả này gợi ý rằng, ba nhóm đại diện cho ba phân loài của cá chép là *C. carpio carpio, C. carpio haematopterus* và *C. carpio viridiviolaceus.* Giả thiết này cũng được thể hiện qua sự khác biệt đáng kể về hình thái của các phân loài cá chép ở Việt Nam (nó có thân hình ngắn hơn và kích thước nhỏ hơn hai loài còn lại) và về di truyền thì được thể hiện sự khác biệt trên hình 2. Do đó, có thể thấy rằng các phân loài cá chép nuôi tại Việt Nam có mức độ tương đồng cao về mặt di truyền với cá chép Amur tự nhiên (hoang dã).

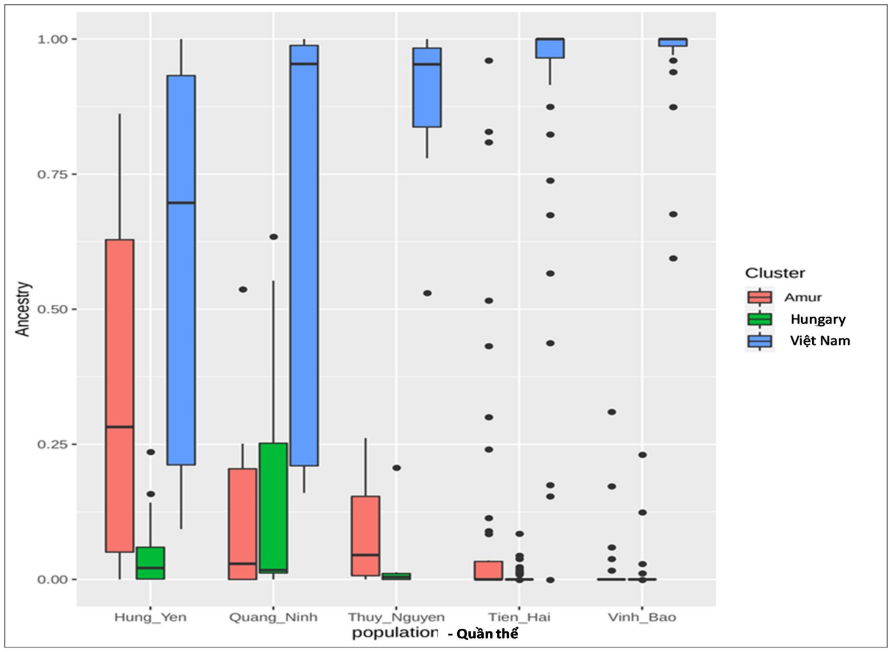
Mặt khác, để xác định các thành phần nguồn gốc / tổ tiên của ba phân loài cá trong các mẫu được nghiên cứu, phân tích tổng hợp đã được tiến hành bằng công cụ NGSadmix trong phần mềm thống kê R. Kết quả phân tích được thể hiện trên hình 3.



Hình 3. Phân tích NGSadmix các mẫu cá chép Việt Nam. Trong đó, phân loài cá chép Châu Âu (*C.c. carpio)* là màu xanh lá cây, phân loài cá chép Amur (*C.c. haem atopterus)* - màu đỏ và phân loài cá chép Việt Nam (*C.c. viridiviolaceus)* - màu xanh lam

Cũng như kết quả phân tích PCA (hình 2), kết quả trên hình 3 cho thấy, các mẫu cá chép nuôi và cá chép tự nhiên của Việt Nam có sự khác nhau về nguồn gốc/tổ tiên. Mỗi quần thể đều có thành phần và tỷ lệ đóng góp nguồn gen tổ tiên là khác nhau. Cụ thể, tất cả các mẫu cá chép nuôi thu từ khu vực Sa Pa đều có tỷ lệ đóng góp nguồn gen tổ tiên gần như nhau từ ba phân loài tổ tiên; các mẫu cá thu từ các trại nuôi ở Quảng Ninh chủ yếu có nguồn gốc tổ tiên là phân loài cá Amur cùng với một ít tổ tiên từ phân loài cá Hungary; Các mẫu cá thu từ các trại nuôi ở Bắc Giang cũng có nguồn gốc tổ tiên từ cả ba phân loài, nhưng độ phân tán giữa các mẫu tương đối cao.

Các quần thể cá chép tự nhiên là đối tượng nghiên cứu chính trong nghiên cứu này cũng cho thấy, có mức độ đóng góp về nguồn gốc di truyền từ cá chép Châu Âu và Amur phổ biến và rải rác ở các mẫu vật. Kết quả nghiên cứu sự biến động về đóng góp nguồn gốc tổ tiên của ba phân loài đối với mỗi quần thể cá chép tự nhiên được thể hiện trên hình 4.



Hình 4. Biến động sự đóng góp nguồn gốc gen trong 5 quần thể cá chép tự nhiên Việt Nam

Từ kết quả trên hình 4 cho thấy, quần thể cá chép tự nhiên ở khu vực Tiền Hải và Vĩnh Bảo gần như là có nguồn gốc thuần Việt, quần thể cá chép tự nhiên ở khu vực Hưng Yên có tỷ lệ nguồn gốc gen từ phân loài cá chép Amur lớn, và quần thể cá chép tự nhiên ở Quảng Ninh có sự đóng góp nguồn gốc gen vừa phải của cả cá phân loài chép Amur và cá chép Châu Âu. Cuối cùng, quần thể cá chép tự nhiên ở khu vực Thủy Nguyên có sự đóng góp nguồn gốc gen vừa phải từ cá chép Amur.

Các locus có thành phần alen khác nhau giữa phân loài cá chép tự nhiên của Việt Nam (*C. c. viridiviolaceus*) và phân loài cá chép tự nhiên Amur (*C. c. viridiviolaceus*) được sử dụng để phân tích sự đóng góp nguồn gốc gen của cá chép tự nhiên Việt Nam đối với ngành nuôi trồng cá chép của Việt Nam. Tổng số 3592 SNPs trong cá chép tự nhiên của Việt Nam có sự khác biệt với các phân loài cá chép tự nhiên Amur. Những locus này có giao nhau với 1100 gen trong hệ gen của cá chép đối chiếu, so sánh (mã GCF_000951615.1). Các gen có locus khác của cá chép tự nhiên Việt Nam có sự biến động từ 0 đến 80 trong các mẫu cá chép nuôi ở khu vực Quảng Ninh và Bắc Giang (trung bình là 41). Trong khi đó, 47 gen với locus bị chèn (giảm 5%) được tìm thấy trong hơn 71 mẫu cá chép nuôi của Việt Nam, trong khi 8 gen (giảm 1%) có trong hơn 76 mẫu cá chép nuôi (bảng 3).

Bảng 3. Tám gen đầu có alen cá chép tự nhiên Việt Nam được tìm thấy trong mẫu cá chép nuôi ở Quảng Ninh và Bắc Giang

Mã gen (ID)	Tên gen	Mô tả gen	Chức năng gen
LOC10909350	*ANKRD11*	Vùng lặp có chứa 11 protein	Mã hóa protein ức chế hoạt động chuyển hóa
LOC109049698	*PCK2*	Phosphoenol pyruvate carboxykinase 2	Liên quan tới con đường tạo glucose
LOC109049846	*TAX1BP1B*	Protein liên kết Tax1 homolog B 1	Có thể ức chế hoạt động apoptotic
LOC109091495	*SAT2*	Diamine acetyltransferase 2-	Liên quan tới hình thành mạch và acetyl hóa tinh trùng
LOC109045512	*UBOX5*	Giống protein vòng 37	Tham gia vào vận chuyển protein
LOC109045512	*CEACAM5*	5 phân tử kết dính tế bào, kháng nguyên Carcinoembryonic	glycoprotein trên bề mặt tế bào có vai trò kết dính
snx2	*SNX2*	Phân loại nexin 2	Tham gia vào một số giai đoạn nội bào
LOC109045682	*CAPN15*	Calpain-15	Canxi, cysteine và hoạt động endopeptidase

3.2. Bàn luận các kết quả nghiên cứu

Sự tương tác trong các quần thể cá chép Việt Nam đã cung cấp cơ sở khoa học cho việc nghiên cứu sự tác động di truyền của các loài cá chép tự nhiên bởi dòng cá chép nuôi, vì các phân loài cá chép thuần bản địa của Việt Nam (*C. c. viridiviolaceus*) nói chung hiếm khi được nuôi do chúng có tốc độ tăng trưởng thấp. Đồng thời, hai phân loài phụ khác của cá chép (đặc biệt là Cá chép Amur) được nuôi rộng rãi trong các trại nuôi cá ở Việt Nam, và chúng có sự khác biệt rõ rệt với các phân loài cá chép Việt Nam về các chỉ thị di truyền (hình 3). Theo kết quả của nghiên cứu này, loài cá được nuôi nhiều nhất ở Việt Nam có nguồn gốc gen bị lai tạp với cá chép Amur và nó có đóng góp nhiều hơn từ hai phân loài cá còn lại. Cá chép Amur dường như là đại diện cho các dòng cá chép nội địa Trung Quốc đã từng được Việt Nam nhập nuôi. Chúng thuộc phân loài Amur nhưng không phải là cá chép vàng Indonesia, chúng có màu cam tươi mặc dù thuộc phân loài Amur. Tuy nhiên, 12 mẫu được thu thập từ trại nuôi cá ở Bắc Giang có màu sắc kiểu Indonesia và dường như nó cũng có nguồn gốc tổ tiên từ cá chép vàng Indonesia, nhưng hầu hết các loài cá chép nuôi khác đều có màu hung (nâu vàng).

Nguồn gốc gen của các phân loài Châu Âu (hoặc Ponto-Caspian) *C. c. carpio* (chủng nội địa Hungary) cũng thấy hiện diện trong cá chép nuôi của Việt Nam, nhưng giá trị của sự hiện diện có sự khác nhau giữa các trại nuôi. Ví dụ, ở trại nuôi tại Sa Pa thì hầu hết các mẫu cá đều có tỷ lệ nguồn gốc gen bằng nhau từ cả ba phân loài. Tuy nhiên, đặc trưng về cấu trúc di truyền của phân loài cá chép Amur vẫn chiếm ưu thế trong cá chép từ các trại nuôi cá được thu mẫu trong nghiên cứu này. Sự khác biệt về di truyền của các phân loài cá chép giữa môi trường tự nhiên hoang dã và môi trường nuôi rất hữu ích để xác định được tổ tiên từ cá nuôi hoặc di nhập vào trong các quần thể cá tự nhiên. Theo kết quả trong nghiên cứu này cho thấy, tỷ lệ nguồn gốc gen từ cá chép tự nhiên Việt Nam là rất lớn so với tỷ lệ đóng góp nguồn gốc gen từ cá chép Amur cũng như cá chép Châu Âu (Hungary). Tuy nhiên, nguồn gốc gen từ cá chép Amur đều hiện diện ở trên hầu hết các mẫu vật thu được, các mẫu này có thể là các con lai pha trộn nhiều dòng. Thực tế cho thấy, các mẫu cá chép được cho là có nguồn gốc hoàn toàn từ Việt Nam lại rất hiếm khi bắt gặp trong trong môi trường tự nhiên.

Chúng tôi cho rằng, phát hiện này có thể là kết quả của việc các cá thể cá nuôi thoát khỏi trại cá cũng như kết quả của sự du nhập cá chép Hungary và Indonesia vào những năm 1970. Tuy nhiên, vẫn cần thêm các kết quả nghiên cứu tiếp theo để làm sáng tỏ mức độ ảnh hưởng giữa cá thoát khỏi trại nuôi và cá du nhập. Mặc dù vậy, theo cách gián tiếp, có thể giả định rằng cá nuôi thoát ra ngoài có tác động đến cấu trúc quần thể cá tự nhiên do nguồn gốc gen của cá chép Amur chiếm ưu thế trong cá tự nhiên Việt Nam so với gen tổ tiên từ cá chép Hungary; trong các cơ sở nuôi cá, nguồn gốc gen từ cá chép Amur cũng chiếm ưu thế. Hơn nữa, tổ tiên cá chép Amur được di nhập từ giống cá chép vàng Indonesia, có màu sắc cam sáng đặc trưng và không bắt gặp cá có màu sắc hoặc giống như vậy trong số các mẫu đã thu từ tự nhiên.

Kết quả trên hình 4 cũng đã chỉ rõ ràng rằng, ba trong số năm quần thể cá tự nhiên của Việt Nam có có nguồn gốc gen từ cá nuôi hoặc cá di nhập với giá trị từ khoảng 5% đối với quần thể cá ở Thủy Nguyên đến gần 34% đối với quần thể cá ở Hưng Yên. Đồng thời, số mẫu của các nhóm cá nuôi chiếm tới 87%, có thể do cá chép được nuôi rộng rãi ở ruộng lúa, dẫn đến sự cách ly hoàn toàn với môi trường tự nhiên. Tuy nhiên, chúng tôi thấy có sự khác biệt rõ ràng giữa các mẫu cá chép tự nhiên và cá chép nuôi về mặt phân loài tổ tiên di truyền, rõ ràng có sự liên quan đến việc cá nuôi bị giảm sự thích nghi trong môi trường tự nhiên cũng như mức độ giao phối thấp giữa các phân loài cá chép khác nhau.

Cá chép nuôi của Việt Nam rõ ràng có nguồn gốc chủ yếu từ các dòng cá chép của Trung Quốc, thuộc phân loài Amur; tuy nhiên, một nửa trong số chúng chứa 11% hoặc nhiều hơn mức xáo trộn từ phân loài cá chép Việt Nam. Được biết, việc lai giữa cá chép nuôi với cá chép tự nhiên của Việt Nam thường được sử dụng để làm tăng khả năng chống trọi với điều kiện môi trường của cá nuôi [5, 6]. Trong nghiên cứu này, chúng tôi đã xác định được các gen thường chứa alen "cá tự nhiên Việt Nam" trong các cá thể cá chép nuôi ở Việt Nam, những thông tin này có vai trò rất quan trọng đối với việc lựa chọn chỉ thị di truyền trong các chương trình nhân giống cá chép trong thời gian tới. Các gen đầu tiên có tần số xuất hiện tối đa được trình bày trong bảng 3.

Kết quả gợi ý rằng, những khác biệt về gen giữa các quần thể cá chép tự nhiên và cá chép nuôi có thể được sử dụng để phát triển một thử nghiệm SNP mật độ cao nhằm phân biệt chính xác giống cá chép và sự biến đổi của các giống cá chép nuôi đã thoát khỏi các trang trại nuôi trồng ở Việt Nam, và sử dụng các chỉ thị phân tử trong chọn lọc cá chép nhằm tạo ra các dòng cá chép nội địa mới.

4. KẾT LUẬN VÀ KHUYẾN NGHỊ

4.1. Kết luận

- Hệ gen của quần thể cá chép tự nhiên ở Việt Nam có tỷ lệ nguồn gốc gen từ phân loài cá chép Việt Nam là chủ yếu so với nguồn gen từ phân loài cá chép Amur và cá chép Châu Âu (Hungary), và chúng cũng đã bị tác động mạnh mẽ từ nguồn gen của cá nuôi.

- Sự khác biệt về hệ gen giữa các quần thể cá chép tự nhiên và cá chép nuôi trong nghiên cứu này có thể được sử dụng để xác định nguồn giống cá chép, sự biến đổi của các giống cá chép nuôi đã thoát khỏi các trại nuôi trồng ở Việt Nam.

- Một số gen chức năng về sự thích nghi với các điều kiện nuôi trồng và đặc tính kinh tế đã được xác định trong cá chép ở Việt Nam có thể lựa chọn làm chỉ thị di truyền trong lai tạo và chọn giống cá chép trong thời gian tới.

4.2. Khuyến nghị

Cần có chương trình bảo tồn nguồn gen của phân loài cá chép tự nhiên Việt Nam (*C. carpio viridiviolaceus)* nói riêng và các nguồn gen sinh vật bản địa nói chung, cần tiếp tục ứng dụng phương pháp ddRAD và các chỉ thị di truyền, gen chức năng để mở rộng quy mô nghiên cứu về xác định thực trạng nguồn gốc di truyền, tác động di truyền và chọn lọc giống mới đối với cá chép nói riêng và các loài sinh vật khác nói chung trong thời gian tới.

Dữ liệu đính kèm: Tất cả các trình tự DNA đã được tải lên Cơ sở dữ liệu Genbank (NCBI) với ID của dự án là PRJNA573845 và ID của mẫu được thể hiện trong bảng 1.

Lời cảm ơn: *Nghiên cứu này được thực hiện với sự hỗ trợ kinh phí từ Viện Hàn lâm Khoa học và Công nghệ Việt Nam (QTRU01.07/19-20), Quỹ Nghiên cứu Cơ bản của Nga (19-54-54004) và Bộ Khoa học và Giáo dục Đại học Liên bang Nga (075-15-2019-1659).*

TÀI LIỆU THAM KHẢO

1. Kirpitchnikov V. S., *Homologous hereditary variation and evolution of wild common carp Cyprinus carpio* L., Genetika, 1967, **8**:65-72.

2. Kohlmann K., Kersten P., Flajšhans M., *Microsatellite-based genetic variability and differentiation of domesticated, wild and feral common carp (Cyprinus carpio L.) populations*, Aquaculture, 2005, **247**:253-266.

3. Memiş D., Kohlmann K., *Genetic characterization of wild common carp (Cyprinus carpio* L*.) from Turkey*, Aquaculture, 2006, **258**:257-262.

4. Fisheries Global Information System (FAO-FIGIS), *Fish. Glob. Inf. Syst. FIGIS*, 2020, URL http://www.fao.org/fishery/

5. Penman D. J., *Carp genetic resources for aquaculture in Asia,* WorldFish Center, 2005, P.O. Box 500 GPO, 10670 Penang, Malaysia.

6. Thai B. T., Burridge C. P., Pham T. A., Austin C. M., *Using mitochondrial nucleotide sequences to investigate diversity and genealogical relationships within common carp (Cyprinus carpio L.)*, Anim. Genet, 2005, **36**:23-28.

7. Thai B. T., Pham T. A., Austin C. M., *Genetic diversity of common carp in Vietnam using direct sequencing and SSCP analysis of the mitochondrial DNA control region*, Aquaculture, 2006, **258**:228-240.

8. Thai B. T., Burridge C. P., Austin C. M., *Genetic diversity of common carp (Cyprinus carpio L.) in Vietnam using four microsatellite loci*, Aquaculture, 2007, **269**:174-186.

9. Hohenlohe P. A., Bassham S., Etter P. D., Stiffler N., Johnson E. A., Cresko W. A., *Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags,* PLoS Genet., 2010, **6**, e1000862.

10. Franchini P., Monné Parera D., Kautt A. F., Meyer A., *QuaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage*, Mol. Ecol., 2017, **26**:2783-2795.

11. Sambrook J., Fritsch E. R., & Maniatis T., *Molecular cloning: a laboratory manual,* Cold Spring Harbor, 1989.

12. Xu P., Zhang X., Wang X., Li J., Liu G., Kuang Y., Xu J., Zheng X., Ren L., Wang G., Zhang Y., Huo, et al., *Genome sequence and genetic diversity of the common carp, Cyprinus carpio*, Nat. Genet., 2014, **46**:1212-1219.

13. Langmead B., Wilks C., Antonescu V., Charles R., *Scaling read aligners to hundreds of threads on general-purpose processors*, 2017.

14. Li H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data*, Bioinformatics, 2011, **27**:2987-2993.

15. Knaus B. J., Grünwald N. J., *Vcfr: a package to manipulate and visualize variant call format data in R*, Mol. Ecol. Resour., 2017, **17**:44-53.

16. Jombart T., Ahmed I., *Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data*, Bioinforma. Oxf. Engl., 2011, **27**:3070-1.

17. Skotte L., Korneliussen T. S., Albrechtsen A., *Estimating individual admixture proportions from next generation sequencing data*, Genetics, 2013, **195**:693-702.

18. Nedoluzhko A. V., et al., *A new strain group of common carp: The genetic differences and admixture events between Cyprinus carpio breeds*, Ecol Evol 2020, **10**(12):5431-5439.

19. Quinlan A. R. and Hall I. M., *BEDTools: a flexible suite of utilities for comparing genomic features*, Bioinformatics, 2010, **26**(6):841-842.

SUMMARY

USING THE GENETIC ddRAD SEQUENCEING TECHNIQUES TO DETERMINE THE GENOME ORIGINATION AMONG DOMESTIC EUROPEAN COMMON CARP (*Cyprinus carpio carpio*) AND AMUR CARP (*Cyprinus carpio haematopterus*) TO THE WILD VIETNAMESE POPULATION

There are four subspecies of the common carp of the world: *C. carpio carpio*, *C. carpio haematopterus* and *C. carpio viridiviolaceus* and *C. carpio aralensis*. These subspecies habit natively in Ponto-Caspian, Far-Eastern regions, Northern Vietnam and Central Asia accordingly, but they were introduced into almost all regions with suitable environmental conditions; therefore, it is one of the most widely cultured freshwater fish species in the world. In Northern Vietnam, common carp fisheries and aquacultural production have great economic importance. Investigation of the subspecies distribution among aquacultured and wild common carps is important for the understanding of the impact of human economic activity, such as aquaculture production, on the wild environment. ddRAD sequencing was used to estimate the genetic impact of aquacultured carps to wild common carp populations of Northern Vietnam. It was shown that a part of wild carp populations has significant contribution of farmed fish ancestry. For some populations, this contribution exceeds 25% of population ancestry, but the ancestry value distributed across the most specimens of population, whereas pure aquacultured specimens are quite uncommon for wild habitat.

Keywords: *Cyprinus carpio, Vietnamese carp, Amur carp, European carp, ddRAD.*

Nhận bài ngày 07 tháng 11 năm 2021

Phản biện xong ngày 18 tháng 11 năm 2021

Hoàn thiện ngày 23 tháng 02 năm 2022

(1) Viện Tài nguyên và Môi trường biển, Viện Hàn lâm KH&CN Việt Nam

(2) Học viện Khoa học và Công nghệ

(3) Khoa Khoa học Sinh học và Nuôi trồng Thủy sản, Đại học Nord, Na Uy

(4) Trung tâm Nghiên cứu Quốc gia "Viện Kurchatov"

(5) Trung tâm Nhiệt đới Việt - Nga