

BỘ NÔNG NGHIỆP VÀ PHÁT TRIỂN NÔNG THÔN
VIỆN NGHIÊN CỨU NUÔI TRỒNG THỦY SẢN I
TRUNG TÂM QUỐC GIA GIỐNG THỦY SẢN NƯỚC NGỌT MIỀN BẮC

**ĐÁNH GIÁ CHI TIẾT NGUỒN GEN CÁ CHÉP TRẮNG (*Cyprinus carpio*
Linnaeus, 1758) BẰNG CHỈ THỊ PHÂN TỬ**

Thuộc nhiệm vụ:

Bảo tồn, lưu giữ nguồn gen và giống thủy sản khu vực miền Bắc năm 2021

Đơn vị chủ trì dự án: Viện Nghiên cứu nuôi trồng thủy sản 1

Chủ nhiệm nhiệm vụ: TS. Võ Văn Bình

Hải Dương, tháng 10 năm 2021

MỤC LỤC

I. THÔNG TIN CHUNG	3
1.1. Thông tin chu về mẫu phân tích.....	3
1.2. Một số chỉ tiêu phân tích.....	3
II. PHƯƠNG PHÁP	4
2.1. thu mẫu.....	4
2.2. Tách chiết DNA	4
2.3. Phản ứng PCR khuếch đại gen (PCR1).....	4
2.4. Tinh sạch sản phẩm PCR1	5
2.5. Phản ứng PCR giải trình tự (PCR2).....	5
2.6. Tinh sạch sản phẩm PCR2	5
2.7. Thực hiện giải trình tự trên máy GenomeLab GeXP.....	6
2.8. Xử lý số liệu.....	6
III. KẾT QUẢ.....	7
3.1. Kết quả về định danh loài bằng phương pháp sinh học phân tử.....	7
3.2. Đa dạng haplotype và đa dạng nucleotide	8
3.3. Đánh giá sai khác di truyền.....	9
3.4. Mối quan hệ di truyền	10
IV. KẾT LUẬN VÀ ĐỀ XUẤT.....	11
4.1. Kết luận	11
4.2. Đề xuất	11
TÀI LIỆU THAM KHẢO	12

I. THÔNG TIN CHUNG

1.1. Thông tin chung về mẫu phân tích

Địa điểm thu mẫu: Trung Tâm quốc gia giống thủy sản nước ngọt miền Bắc, địa chỉ: Phường Tân Dân – Thị xã Chí Linh - Thành phố Hải Dương.

Đặc điểm mẫu: Mẫu vây cá chép bảo quản trong ethanol 96%.

Số lượng mẫu: 30 mẫu

Chỉ thị phân tích: Giải trình tự gen đoạn gen COI (Cytochrome c oxidase I)

Ngày thu mẫu: 10/11/2021

Ngày phân tích mẫu: Từ 15/11/2021 đến 25/11/2021

Cán bộ phân tích mẫu: Vũ Thị Trang

Đơn vị phân tích: Trung tâm Công nghệ sinh học Thủy sản – Viện Nghiên cứu Nuôi trồng Thủy sản I.

1.2. Một số chỉ tiêu phân tích

Bảng 1. Các thông số mẫu thu và phân tích

STT	Chỉ tiêu phân tích	Số lượng mẫu đưa vào phân tích	Số lượng mẫu phân tích đạt yêu cầu
1	Tách chiết DNA	30	30
2	PCR khuếch đại gen COI	30	30
3	Điện di	60	60
4	Tinh sạch sản phẩm PCR	30	30
5	Giải trình tự	30	30

II. PHƯƠNG PHÁP

2.1. Thu mẫu

Tiến hành thu ngẫu nhiên 30 mẫu vây nguồn gen cá chép trắng Việt hiện đang lưu giữ, mẫu vây được bảo quản trong ethanol 96% ở 4°C cho đến khi phân tích. Định danh loài theo phương pháp sinh học phân tử theo tiêu chuẩn ISO/IEC 17025 chỉ tiêu HD.S.02 được thực hiện tại phòng thí nghiệm Sinh học phân tử, thuộc Trung tâm Công nghệ sinh học thủy sản – Viện NCNTNS1. Các bước tiến hành cụ thể như sau:

2.2. Tách chiết DNA

Trong nghiên cứu này DNA sẽ được tách chiết theo phương pháp kết tủa muối (Sambrook và Russell, 2001). Sản phẩm DNA tách chiết được kiểm tra sản phẩm trên gel agarose (nồng độ 0.8 - 1%). Nếu sau khi điện di, các vạch DNA tổng số sáng, gọn, không bị smear có nghĩa là DNA không bị đứt gãy, không bị nhiễm bản, đủ tiêu chuẩn cho các bước phân tích tiếp theo. Tiếp theo kiểm tra độ tinh sạch DNA (đo OD) bằng NanoDrop. Các chỉ số OD 260/280 nằm trong khoảng 1.8 - 2.0 và OD 260/230 nằm trong khoảng 1.8-2.2 là đạt tiêu chuẩn.

2.3. Phản ứng PCR khuếch đại gen (PCR1)

Nghiên cứu sử dụng đoạn mồi khuếch đại đoạn gen COI thuộc mtDNA của loài cá chép *Cyprinus carpio*. Đoạn mồi này được thiết kế dựa trên phần mềm Primer3Plus (<https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) trên cơ sở đoạn gen COI của loài cá này đã được công bố trên ngân hàng gen NCBI với mã số NC001606 (Chang và ctv., 1994). Thông tin chi tiết về chỉ thị được trình bày ở Bảng 2.

Phản ứng được thực hiện với tổng thể tích 50µl bao gồm: 25µl MyTaq™ Mix 2× (Bioline), 2µl mỗi loại mồi xuôi và mồi ngược (nồng độ mồi 10µM), 2µl DNA khuôn (~200 ng) và nước đê ion. Điều kiện phản ứng đối với chu kỳ nhiệt nhân đoạn gen COI là: giai đoạn khởi đầu ở 94°C trong 2 phút; tiếp đó là 35 chu kỳ bao gồm (giai đoạn biến tính ở 94°C trong 30 giây; giai đoạn gắn mồi ở 58°C trong 15 giây; giai đoạn kéo dài ở 72°C trong 30 giây); giai đoạn kết thúc kéo dài ở 72°C trong 5 phút và giữ ở 4°C. Sau đó sản

phẩm PCR được tinh sạch và giải trình tự. Sau khi phản ứng kết thúc lấy 3 µl sản phẩm điện di trên gel agarose 1.5% để kiểm tra kết quả sản phẩm PCR.

Bảng 2. Thông tin về chi thị sử dụng trong nghiên cứu

Tên môi	Trình tự môi	Nhiệt độ gắn môi (°C)	Kích thước sản phẩm PCR (bp)
COI-F. <i>C. Carp</i>	5'-CACGCAGGAGCATCAGTAGA-3'	58	819
COI-R. <i>C. Carp</i>	3'-TG TTCAGGTGCTGTGGAGAG-5'	58	

2.4. Tinh sạch sản phẩm PCR1

Sử dụng kit tinh sạch sản phẩm PCR, QIAquick PCR Purification Kit, Qiagen. Quy trình tinh sạch theo hướng dẫn của nhà sản xuất.

2.5. Phản ứng PCR giải trình tự (PCR2)

Thành phần của phản ứng PCR giải trình tự trong tổng thể tích 20 µl bao gồm: 8 µl DTCS Quick Start Kit, 0,5-10 µl sản phẩm PCR1 đã tinh sạch đảm bảo nồng độ 50 ng/100 fmol, 2 µl Primer (1,6 pmol/µl hoặc 1,6 µM). Sau đó cho nước cất lên tổng thể tích 20 µl. Chu trình nhiệt cho phản ứng PCR2 là 30 chu kỳ bao gồm: 96°C trong 20 giây, 50°C trong 20 giây và 60°C trong 4 phút. Cuối cùng giữ ở 4°C.

2.6. Tinh sạch sản phẩm PCR2

Tinh sạch bằng Ethanol (Merk) theo quy trình sau: Chuyển toàn bộ sản phẩm PCR2 (20 µl) sang ống eppendorf đã chứa 5 µl dung dịch ngừng phản ứng bao gồm: 2 µl 3M Natri Acetat pH 5,2; 2 µl EDTA 100 mM; pH 8,0, 1 µl Glycogen 20 mg/ml. Thêm 75 µl Ethanol 95% lạnh bảo quản trong -20°C, vortex, ly tâm 13.000 vòng/phút trong 20 phút tại 4°C. Bỏ dịch nổi phía trên, giữ lại DNA kết tủa ở đáy ống. Rửa kết tủa bằng 200 µl Ethanol 70% lạnh bảo quản trong -20°C, vortex, ly tâm 13.000 vòng/phút trong 3,5 phút tại 4°C. Bỏ dịch nổi phía trên, giữ lại DNA kết tủa ở đáy ống. Bước này thực hiện 2 lần. Để khô tự nhiên ở nhiệt độ phòng. Hòa tan DNA kết tủa bằng 40 µl SLS.

2.7. Thực hiện giải trình tự trên máy GenomeLab GeXP

Chuyển toàn bộ dung dịch DNA hòa tan trong SLS từ bước tinh sạch vào giếng tương ứng trên đĩa chạy mẫu. Nhỏ 1 giọt dầu (mineral oil - được cung cấp kèm theo DTCS Quick Start Kit) lên trên để tránh bay hơi mẫu. Đặt đĩa mẫu và đĩa đệm vào máy giải trình tự. Cài đặt và chạy chương trình.

2.8. Xử lý số liệu

Các trình tự gen được kiểm tra chất lượng bằng phần mềm Finch TV 1.4.0 (<http://www.geospiza.com>). Tiếp theo, chương trình ClustalW Multiple Alignment trong BioEdit 7.1.9 (Hall, 1999) được sử dụng nhằm so sánh, căn chỉnh trình tự, sau đó được kiểm tra và xác định mức độ tương đồng. Trình tự đoạn gen COI của các mẫu cá chép sẽ được so sánh với trình tự DNA của các mẫu cá chép ở Ngân hàng gen Quốc tế thông qua sử dụng công cụ BLAST trên NCBI hoặc BOLD để xác định loài tương đồng.

Các chỉ số đa dạng di truyền kiểu đơn - Haplotype diversity (h), đa dạng nucleotit - Nucleotide diversity (π) và xác định các kiểu haplotype được tính toán bằng cách sử dụng phần mềm DnaSP v6.10.01 (Rozas và ctv., 2017). Xây dựng mạng lưới haplotype bằng phần mềm Network v10.2.0.0 (Fluxus Technology Ltd, Clare. Available online at: fluxus-engineering.com). Phần mềm này sử dụng kết nối mạng dữ liệu đầu vào được tạo ra bởi phần mềm DnaSP v6.10.01 (Rozas và ctv., 2017) và sử dụng thuật toán Median Joining (chức năng calculate network) để tính. Chức năng draw network cho phép tự động vẽ ra mạng lưới giữa các haplotype được xem xét. Phần mềm Arlequin v.3.11 (Excoffier và ctv., 2005) được dùng để đánh giá sai khác di truyền, mức độ khác biệt bên trong và giữa các dòng qua phân tích F_{ST} (Population pairwise F_{ST}).

Cây phát sinh loài phân tử thể hiện mối quan hệ di truyền giữa các nhóm cá nghiên cứu được xây dựng dựa trên phương pháp UPGMA (Unweighted Pair Group Method with Arithmetic mean) (Sneath và Sokal, 1973), gồm 4 trình tự là trình tự phổ biến (consensus sequence) được xây dựng cho từng nhóm. Các đơn vị phân loại liên kết được nhóm lại với nhau trong thử nghiệm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh

(Felsenstein, 1985). Cây được vẽ theo tỷ lệ, với độ dài các nhánh bằng đơn vị của khoảng cách tiến hóa được sử dụng để suy ra cây phát sinh loài. Khoảng cách tiến hóa được tính bằng phương pháp Tajima-Nei (Tajima và Nei, 1984) và được tính bằng đơn vị của số lần thay thế bazơ trên mỗi vị trí. Các phân tích tiến hóa được thực hiện trong MEGA X (Kumar và ctv., 2018).

III. KẾT QUẢ

3.1. Kết quả về định danh loài bằng phương pháp sinh học phân tử

Bảng 3. Kết quả so sánh BLAST trên ngân hàng gen NCBI

STT	Kí hiệu mẫu	Loài tương đồng	Mã số trên NCBI	Mức độ tương đồng (%)	STT	Kí hiệu mẫu	Loài tương đồng	Mã số trên NCBI	Mức độ tương đồng (%)
1	1.1	<i>Cyprinus carpio</i>	MW450993.1	100	16	2.6	<i>Cyprinus carpio</i>	AP017363.1	99.88
2	1.2	<i>Cyprinus carpio</i>	MW451000.1	100	17	2.7	<i>Cyprinus carpio</i>	MW450993.1	100
3	1.3	<i>Cyprinus carpio</i>	MW450998.1	100	18	2.8	<i>Cyprinus carpio</i>	AP017363.1	99.88
4	1.4	<i>Cyprinus carpio</i>	MH202953.1	100	19	2.9	<i>Cyprinus carpio</i>	MK291479.1	100
5	1.5	<i>Cyprinus carpio</i>	MT780875.1	100	20	2.12	<i>Cyprinus carpio</i>	AP017363.1	99.87
6	1.9	<i>Cyprinus carpio</i>	MW450998.1	100	21	3.1	<i>Cyprinus carpio</i>	MK291479.1	100
7	1.11	<i>Cyprinus carpio</i>	MW450993.1	100	22	3.4	<i>Cyprinus carpio</i>	MK291479.1	100
8	1.13	<i>Cyprinus carpio</i>	MW450993.1	100	23	3.5	<i>Cyprinus carpio</i>	MK291479.1	100
9	1.14	<i>Cyprinus carpio</i>	MH202953.1	100	24	3.6	<i>Cyprinus carpio</i>	MK291479.1	100
10	1.15	<i>Cyprinus carpio</i>	MH202953.1	100	25	3.7	<i>Cyprinus carpio</i>	AP017363.1	99.87
11	2.1	<i>Cyprinus carpio</i>	MK291479	100	26	3.8	<i>Cyprinus carpio</i>	MK291479.1	100
12	2.2	<i>Cyprinus carpio</i>	AP017363.1	100	27	3.9	<i>Cyprinus carpio</i>	MK291479.1	100
13	2.3	<i>Cyprinus carpio</i>	AP017363.1	99.87	28	3.10	<i>Cyprinus carpio</i>	MW450993.1	100
14	2.4	<i>Cyprinus carpio</i>	AP017363.1	99.88	29	3.11	<i>Cyprinus carpio</i>	MK291479.1	100
15	2.5	<i>Cyprinus carpio</i>	MW450993.1	100	30	3.13	<i>Cyprinus carpio</i>	MK291479.1	100

Kết quả BLAST cho thấy độ tương đồng cao của các trình tự nucleotide trên đoạn gen nghiên cứu của 30 mẫu cá chép với các nghiên cứu khác đã được công bố trên ngân hàng gen NCBI (National Center for Biotechnology Information) và cơ sở dữ liệu mã vạch

BOLD (Barcode of Life Data System), tỷ lệ tương đồng từ 99.87 - 100% với trình tự nucleotide của loài có tên khoa học là *Cyprinus carpio* (Bảng 3).

3.2. Đa dạng haplotype và đa dạng nucleotide

Đa dạng haplotype hay còn được gọi là đa dạng gen/alen hoặc dị hợp tử dự kiến. Thông số này được tính toán là xác suất kết hợp giữa hai trình tự khác nhau một cách ngẫu nhiên. Đa dạng nucleotide là số lượng trung bình của sự khác biệt nucleotide trên mỗi vị trí giữa hai trình tự DNA trong tất cả các cặp có thể có trong quần thể mẫu (Nei, 1987).

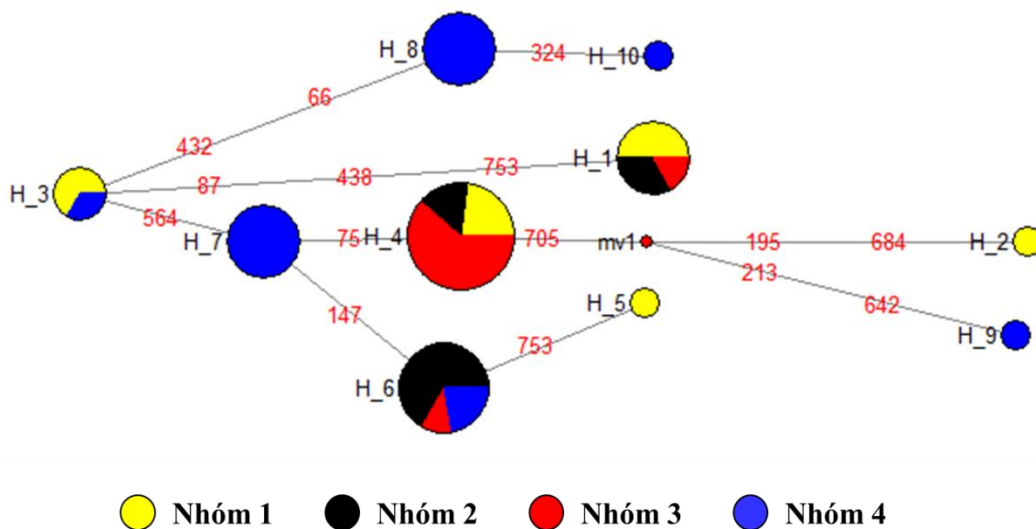
Theo Bảng 4, các nhóm cá có đa dạng haplotype cao là nhóm 1, 2 và 4, trong đó nhóm 1 là nhóm có đa dạng haplotype cao nhất ($Hd \pm SD = 0.844 \pm 0.080$), ngược lại nhóm 3 là nhóm có đa dạng haplotype thấp nhất ($Hd \pm SD = 0.378 \pm 0.181$). Về số lượng haplotype phát hiện được trong mỗi nhóm: nhóm 4 có số lượng haplotype cao nhất (6), sau đó là nhóm 1 (5). Điều này có thể lý giải do số lượng mẫu phân tích đầu vào của nhóm 4 là nhiều hơn so với nhóm 1 (17 so với 10) dẫn xác suất gặp haplotype tăng lên. Về haplotype đặc trưng (là haplotype chỉ xuất hiện ở nhóm này mà không xuất hiện ở nhóm khác): nhóm 4 có nhiều haplotype đặc trưng nhất (4), sau đó đến nhóm 1, cả hai nhóm 2 và 3 đều không có haplotype đặc trưng.

Bảng 4. Tổng hợp đa dạng haplotype và đa dạng nucleotide của các nhóm cá chép

Điểm thu mẫu	Số lượng mẫu có trình tự đạt tiêu chuẩn	Số lượng haplotype	Số lượng haplotype đặc trưng	Đa dạng haplotype ($Hd \pm SD$)	Đa dạng nucleotide ($\pi \pm SD$)
Nhóm 1	10	5	2	<i>0.844 ± 0.080</i>	<i>0.00440 ± 0.00073</i>
Nhóm 2	10	3	0	0.622 ± 0.138	0.00303 ± 0.00089
Nhóm 3	10	3	0	<i>0.378 ± 0.181</i>	<i>0.00178 ± 0.00099</i>
Nhóm 4 rd	17	6	4	0.772 ± 0.070	0.00311 ± 0.00056
Tổng/TB	47	10		<i>0.850 ± 0.026</i>	<i>0.00364 ± 0.00037</i>

Ghi chú: rd: reference data (dữ liệu tham khảo): là dữ liệu cá chép dòng VN, kết quả nghiên cứu thuộc ĐTCS. TB = trung bình

Về đa dạng nucleotide nhìn chung là thấp, dao động từ 0.00178 (nhóm 3) đến 0.00440 (nhóm 4). Đa dạng haplotype cao nhưng đa dạng nucleotide thấp cho thấy có sự khác biệt nhỏ giữa các kiểu haplotype được phát hiện trong từng nhóm nghiên cứu (Bảng 4).



Hình 1. Mạng lưới haplotype của các nhóm cá chép nghiên cứu

(Ghi chú: Kích thước của các vòng tròn tỷ lệ thuận với tần số haplotype, và độ dài của các đường kết nối tỷ lệ thuận với số bước đột biến giữa các haplotype. Thông tin bên ngoài được liên kết với các kiểu gen (nhóm 1, nhóm 2, nhóm 3, nhóm 4) được chú thích để các vòng tròn được thay thế bằng các màu hiển thị tần số của từng nhóm. Mạng haplotype tích hợp các màu cho biết sự phân bố tần xuất xuất hiện của mỗi dạng haplotype của mỗi nhóm. Mv - median vectors – vector trung vị: Mạng lưới haplotype bao gồm các nút và liên kết kết nối các nút. Các nút là trình tự từ tập dữ liệu hoặc vector trung vị. Các liên kết là sự khác biệt về ký tự.

3.3. Đánh giá sai khác di truyền

Trong nghiên cứu này, sự sai khác di truyền giữa các nhóm cá chép được ước lượng theo giá trị F_{ST} (Wei và Cockerham, 1984) và khoảng cách di truyền (Tamura và ctv, 2004; Kumar và ctv., 2016) được thể hiện ở bảng 2 và bảng 3. Theo Nei (1978), nếu $F_{ST} < 0.05$ thì được cho là sai khác nhỏ; nếu $0.05 < F_{ST} < 0.15$ là trung bình; nếu $F_{ST} > 0.15$ là lớn.

Theo Bảng 5 ta thấy, giữa các nhóm: nhóm 1 và nhóm 2 ($F_{ST} = 0.06181$), nhóm 1 và nhóm 3 ($F_{ST} = 0.10774$), nhóm 2 và nhóm 3 ($F_{ST} = 0.20290$), không có sự sai khác di truyền

do $P > 0.05$. Cả 3 nhóm cá (1, 2 và 3) đều có sự sai khác di truyền so với nhóm 4 (do $P > 0.05$).

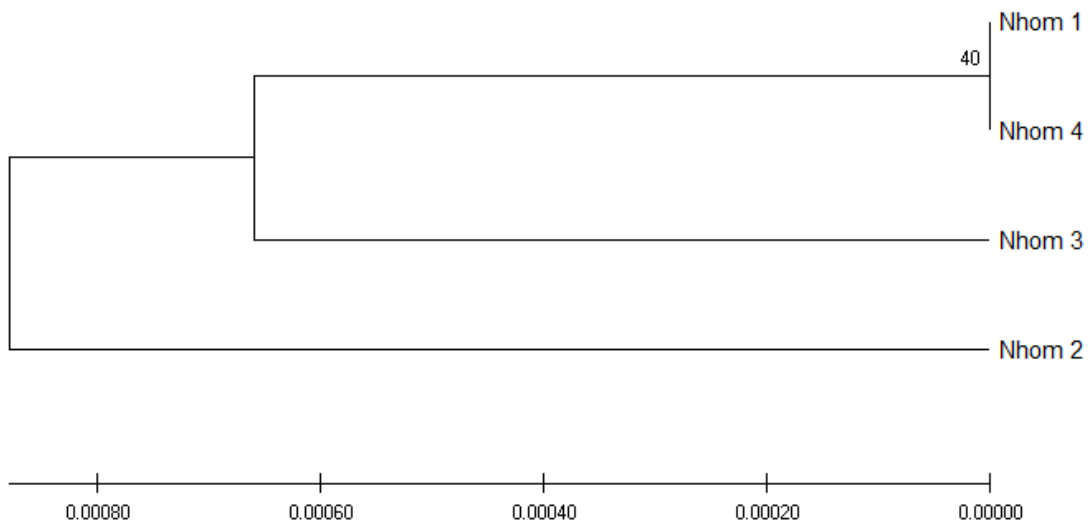
Bảng 5. Bảng thể hiện sự sai khác di truyền (F_{ST}) giữa các nhóm cá chép

	Nhóm 1	Nhóm 2	Nhóm 3	Nhóm 4
Nhóm 1				
Nhóm 2	0.06181			
Nhóm 3	0.10774	0.20290		
Nhóm 4	0.17049*	0.19019 *	0.28613*	

(*Sai khác có ý nghĩa, $P < 0.05$)

3.4. Mối quan hệ di truyền

Cây phát sinh loài phân tử thể hiện mối quan hệ di truyền giữa các nhóm cá nghiên cứu được xây dựng dựa trên phương pháp UPGMA (Unweighted Pair Group Method with Arithmetic mean) (Sneath, 1973), gồm 4 trình tự, đó là các trình tự phổ biến (consensus sequence) được xây dựng cho từng nhóm. Các đơn vị phân loại liên kết được nhóm lại với nhau trong thử nghiệm bootstrap (1000 lần lặp lại) được hiển thị bên cạnh các nhánh. Cây được vẽ theo tỷ lệ, với độ dài các nhánh bằng đơn vị của khoảng cách tiến hóa được sử dụng để suy ra cây phát sinh loài. Khoảng cách tiến hóa được tính bằng phương pháp Tajima-Nei (Tajima F. và Nei M, 1984) và được tính bằng đơn vị của số lần thay thế bazơ trên mỗi vị trí. Các phân tích tiến hóa được thực hiện trong MEGA X (Kumar và ctv., 2018). Hình 2 cho thấy nhóm 1 và 4 có mối quan hệ di truyền gần gũi với nhau hơn so với các nhóm còn lại.



Hình 2. Mối quan hệ di truyền của 4 nhóm cá chép

IV. KẾT LUẬN VÀ ĐỀ XUẤT

4.1. Kết luận

Trong 3 nhóm cá nghiên cứu, nhóm 1 có đa dạng haplotype và đa dạng nucleotide cao nhất, xuất hiện ở nhiều kiểu haplotype khác nhau và có haplotype đặc trưng cho nhóm.

Cả 3 nhóm đều không có sự sai khác di truyền khi so sánh từng cặp với nhau do phân tích không có ý nghĩa thống kê ($P > 0.05$).

Nhóm 1 có mối quan hệ di truyền xa so với 2 nhóm còn lại (nhóm 1, 2; nhóm 4 là dữ liệu tham khảo).

4.2. Đề xuất

Nên tăng số lượng mẫu phân tích cho từng nhóm để khẳng định các thông số đa dạng di truyền, bởi vì để phân tích đa dạng di truyền quần thể, số lượng mẫu nên đạt tối thiểu là 20 mẫu/nhóm hoặc 20 mẫu/quần thể.

TÀI LIỆU THAM KHẢO

- Chang Y.-S., Huang F.-L., Lo T.-B., The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome, *J. Mol. Evol.*, 1994, 38, 138-155.
- Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform.* 1, 47–50.
- Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549.
- Kumar, S., G. Stecher, and K. Tamura 2016 MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7):1870-4.
- Nei M (1987) *Molecular evolutionary genetics*. New York: Colombia University Press. 512 p.
- Nei. M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 89(3): 583-90.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., and Sánchez-Gracia, A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution*. 34(12): 3299-3302.
- Sambrook J and Russell D. 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
- Sneath P.H.A and Sokal R.R. 1973. *Numerical Taxonomy*. Freeman, San Francisco.
- Tajima F. and Nei M. 1984. Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* 1:269-285.
- Tamura K., Nei M. and Kumar, S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)* 101:11030-11035.

Weir B.S. and Cockerham C.C. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38(6):1358-1370.