

XÁC ĐỊNH ĐA HÌNH NUCLEOTIDE ĐƠN KHÔNG ĐỒNG NGHĨA (NON-SYNONYMOUS SNP) TIỀM NĂNG LIÊN QUAN ĐẾN TÍNH TRẠNG TĂNG TRƯỞNG TRÊN CÁ CHÉP LAI BẰNG GIẢI TRÌNH TỰ RNA-SEQUENCING

Lưu Thị Hà Giang^{1, *}, Vũ Văn In², Kim Thị Phương Oanh³

¹ Viện Nghiên cứu Nuôi trồng Thủy sản I

² Trường Đại học Việt Nhật, Đại học Quốc gia Hà Nội

³ Viện Hàn lâm Khoa học và Công nghệ Việt Nam

*Email: hagiang@ria1.org

TÓM TẮT

Cá chép (*Cyprinus carpio*) là loài cá nuôi nước ngọt truyền thống quan trọng trong nuôi trồng thủy sản ở nhiều quốc gia. Chọn giống truyền thống kết hợp di truyền phân tử đóng vai trò quan trọng trong cải thiện tốc độ sinh trưởng, nâng cao năng suất và hiệu quả kinh tế. Công nghệ RNA-sequencing được áp dụng trong nghiên cứu này để xác định các SNP không đồng nghĩa (nsSNP) tiềm năng của 2 nhóm cá chép lai (sinh trưởng nhanh và sinh trưởng chậm) giữa cá chép thuần Hungary và cá chép thuần Việt Nam. Sau khi giải trình tự, tổng số reads sau sàng lọc đạt ~24,6 triệu ở nhóm sinh trưởng nhanh (FG) và ~24,8 triệu ở nhóm sinh trưởng chậm (SG), với tỷ lệ mapping tương ứng là 87,64% và 84,79%. Phân tích xác định 333.328 SNP ở FG, 453.729 SNP ở SG, trong đó có 138.837 SNP chung. Kết quả kiểm định SNP cho thấy, 1.984 SNP có ý nghĩa thống kê liên quan đến khác biệt tăng trưởng, phân bố trên 47 nhiễm sắc thể. Phân tích chức năng cho thấy, các nsSNP chủ yếu nằm ở các gen liên quan đến chuyển hóa các - bon và năng lượng. Đặc biệt, 37 nsSNP tiềm năng được phát hiện trong các gen như: Myogenic Factor 6, Myozenin, Myosin, có vai trò quan trọng trong phát triển cơ và tăng trưởng. Kết quả nghiên cứu là cơ sở khoa học giúp cho các chương trình chọn giống cá chép sinh trưởng nhanh ở Việt Nam đạt hiệu cao hơn.

Từ khóa: RNA-Sequencing, SNP không đồng nghĩa (nsSNP), cá chép lai, tính trạng sinh trưởng.

1. ĐẶT VẤN ĐỀ

Cá chép (*Cyprinus carpio*) là một trong những loài cá nước ngọt quan trọng nhất trong nuôi trồng thủy sản của thế giới, đặc biệt tại khu vực châu Á. Theo FAO (2022), tổng sản lượng cá chép năm 2022 đạt 4,7 triệu tấn, với giá trị ước tính trên 8 tỷ USD [1]. Việc cải thiện tốc độ tăng trưởng của cá chép thông qua chọn giống có ý nghĩa quan trọng trong nâng cao năng suất và hiệu quả kinh tế. Phương pháp chọn giống kết hợp giữa phương pháp truyền thống (dựa trên

thông tin về kiểu hình và phả hệ) và phương pháp di truyền phân tử (thông tin về sự sai khác trong trình tự ADN - chỉ thị phân tử) đã và đang được ứng dụng cho kết quả chính xác và rút ngắn quá trình chọn lọc. Gần đây, với sự phát triển của công nghệ giải trình tự thế hệ mới bao gồm RNA-sequencing, việc nghiên cứu các đa hình đơn nucleotide (SNP) đơn có liên quan đến gen/protein quy định các tính trạng kinh tế quan trọng trong chọn giống được thực hiện một cách hiệu quả và nhanh chóng hơn, giúp tối ưu hóa quá trình chọn lọc/lai tạo nhằm cải thiện

năng suất ứng dụng trong lĩnh vực di truyền chọn giống [2]. Trong đó, các SNP không đồng nghĩa (nsSNP - Non synonymous SNP) là những đột biến đơn nucleotide xảy ra trong vùng mã hóa của gen, dẫn đến sự thay đổi axit amin trong chuỗi protein, có khả năng ảnh hưởng cấu trúc protein, tương tác protein và tác động đến chức năng sinh học của gen. Bằng cách phân tích nsSNPs, các nhà nghiên cứu có thể đánh giá mối liên hệ giữa biến đổi di truyền và biểu hiện kiểu hình, từ đó dự đoán chính xác giá trị di truyền của các cá thể trong quần thể. Ngoài ra, nsSNPs cung cấp thông tin quan trọng trong nghiên cứu chức năng gen và cơ chế điều hòa sinh học, với khoảng 20 - 30% nsSNPs được ước tính có ảnh hưởng đến chức năng protein, việc tập trung nghiên cứu vào các SNP này mang lại tiềm năng lớn trong việc cải thiện các giống cá có năng suất cao và khả năng chống chịu tốt [3].

Công nghệ RNA-sequencing đã được áp dụng trong nhiều nghiên cứu, tạo ra được bộ dữ liệu lớn, nâng cao sự hiểu biết về dinh dưỡng, tăng trưởng, sinh sản, chức năng miễn dịch và khả năng thích ứng với môi trường của một số loài nuôi thủy sản [4 - 6]. Trên cá chép, công nghệ RNA-sequencing RNA-seq đã được áp dụng vào nghiên cứu hệ gen biểu hiện (transcriptome) cho các tính trạng về sinh sản và miễn dịch [7]; công bố cơ sở dữ liệu (tổng số 712.042 SNPs) cho các loài cá chép ở Trung Quốc [8]. Tuy nhiên, các cơ sở dữ liệu SNP từ transcriptome này đều chưa chỉ ra mối liên quan với các tính trạng quan trọng (như tính trạng sinh trưởng) trong chọn giống cũng như nuôi trồng thủy sản của loài cá chép *Cyprinus carpio*.

Nghiên cứu này ứng dụng công nghệ giải trình tự toàn bộ hệ gen RNA (RNA-sequencing) kết hợp với phương pháp tin sinh học nhằm xác định các nsSNP tiềm năng, là các đa hình đơn nucleotide xảy ra trong vùng mã hóa của gen và dẫn đến thay đổi axit amin trong protein, từ ảnh hưởng trực tiếp đến tính trạng sinh trưởng của cá chép lai. Kết quả nghiên cứu sẽ cung cấp cơ

sở khoa học cho việc ứng dụng các chỉ thị phân tử trong chọn giống cá chép có tốc độ tăng trưởng cao, góp phần nâng cao hiệu quả sản xuất trong ngành nuôi trồng thủy sản.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

Cá chép lai (*Cyprinus carpio*) (dòng cá chép thuần Hungary x dòng cá chép thuần Việt Nam) được sản xuất (6 gia đình full-sib/phép lai) và nuôi đánh giá sinh trưởng tại Viện Nghiên cứu Nuôi trồng Thủy sản 1. Cụ thể, 40 cá thể cá chép lai được lựa chọn căn cứ vào các chỉ tiêu đánh giá tốc độ sinh trưởng và cân nặng của cá chép lai, số mẫu thu trong mỗi nhóm đều có đại diện của mỗi gia đình trong phép lai xuôi và lai ngược. Tổng cộng 40 mẫu mô cơ thịt từ hai nhóm cá chép lai đã được kiểm định có sự sai khác sinh trưởng (P value < 0,05) bao gồm 20 mẫu từ những cá thể thuộc nhóm cá chép lai sinh trưởng nhanh (FG) ($608,5 \pm 45,44$ g) và 20 mẫu từ những cá thể thuộc nhóm cá chép lai sinh trưởng chậm (SG) ($424,6 \pm 33,45$ g), được thu thập để giải trình tự hệ gen RNA và phân tích tin sinh học xác định các biến thể. Cắt ít nhất 50 mg mẫu mô cơ, bảo quản trong RNAlater ở 4°C cho đến khi tách chiết.

2.2. Phương pháp nghiên cứu

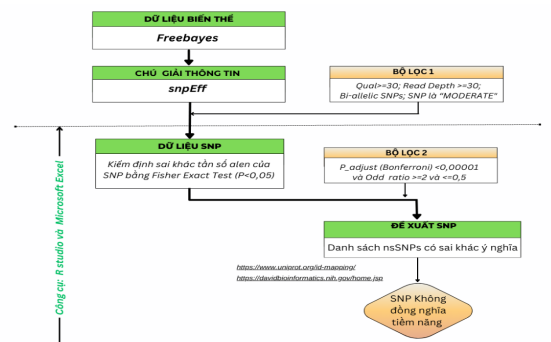
RNA tổng số được tách chiết theo hướng dẫn của TRIsure Plus RNA Purification Kit, được áp dụng để tách chiết RNA tổng số từ các mẫu mô cá chép đã thu. Để giảm thiểu các biến thể trong quá trình chuẩn bị và giải trình tự thư viện, đảm bảo sự khác biệt mức độ biểu hiện gen giữa các nhóm cá đều là do yếu tố sinh học chứ không phải do yếu tố kỹ thuật. Các mẫu RNA (cùng nồng độ) sẽ được gộp mẫu (3 - 4 mẫu cho 1 nhóm thư viện), trong nghiên cứu này 3 mẫu gộp sẽ được xây dựng cho 2 nhóm kiểu hình nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm, tổng số là 6 thư viện cDNA. Thư viện cDNA được thiết lập bằng cách sử dụng Illumina Truseq Stranded mRNA Sample Preparation Kit.

Các bước tiến hành được thực hiện theo hướng dẫn của nhà sản xuất. Kiểm tra chất lượng thư viện bằng Bioanalyzer (Using DNA High Sensitive Kit) và qPCR (Using KAPA Library Qualification Kit). Mẫu cDNA được đưa lên hệ thống giải trình tự gen thế hệ mới Illumina MiSeq và thực hiện theo hướng dẫn của nhà sản xuất. Quá trình giải trình tự Novaseq 100PE được thực hiện bởi NOVOGENE (<https://www.novogene.com/amea-en/>), mẫu được giải trình tự theo hai chiều (paired-end) với độ phủ 20 triệu reads. Chất lượng của dữ liệu giải trình tự được đánh giá bằng công cụ FastQC v.0.11.2 [9]. Quá trình lọc chất lượng và loại bỏ các đoạn Adapter dư thừa được thực hiện trên các cặp reads bằng công cụ Trimmomatic v.0.32 [10] với các tham số cơ bản như sau: Điểm chất lượng tối thiểu là 30; chiều dài tối thiểu 50 bp. Dữ liệu giải trình tự hệ gen RNA sau khi sàng lọc sẽ được xuất ra ở định dạng FASTQ, được sử dụng làm tệp tin đầu vào để phát hiện các SNP. Dữ liệu sau khi được đóng hàng bằng công cụ căn chỉnh HISAT2 [11] với hệ gen tham chiếu cá chép GCF_018340385.1 (NCBI), sẽ được xác định các biến thể bằng công cụ freebayes [12]. Kết quả là các đa hình có thể có trong mẫu được lưu trữ dưới dạng variant call file (VCF). Sau đó, các SNP này được chú giải và dự đoán ảnh hưởng lên gen và protein bằng công cụ snpEff [13].

2.3. Phương pháp xác định và sàng lọc SNP

Sau khi thu được các biến thể từ dữ liệu RNA-sequencing dưới định dạng VCF, quá trình xác định và sàng lọc SNP được thực hiện theo hình 1 để chọn các SNP có độ tin cậy cao và có tiềm năng sinh học. Tất cả các biến thể là đột biến chèn/xóa (InDels) và biến thể cấu trúc lớn (CNVs, inversion, translocation) sẽ được loại bỏ, chỉ giữ lại các đột biến điểm (SNP). Các SNP được sàng lọc chất lượng dựa trên chỉ số QUAL, đại diện cho xác suất Phred-scaled về sự tồn tại của biến thể tại một vị trí cụ thể. Chỉ các biến thể có $QUAL \geq 30$ được giữ lại, tương ứng với độ chính xác $\geq 99,9\%$ [14]. Độ sâu đọc (DP) thể hiện tổng số lượt đọc bao phủ vị trí biến thể. Để

đảm bảo độ phủ đầy đủ, các SNP có $DP < 30$ bị loại bỏ khỏi phân tích tiếp theo [14]. Để xác định SNP liên quan đến sự khác biệt sinh trưởng giữa hai nhóm cá chép, kiểm định Fisher's exact test (2 chiều) được thực hiện, kết hợp tính toán Odds Ratio (OR) và hiệu chỉnh p-value bằng phương pháp Bonferroni. Chỉ các SNP có p-value $< 0,05$, $p_{\text{adjust}} < 0,00001$ và $OR \geq 2$ hoặc $\leq 0,5$ được xem là có ý nghĩa thống kê cao. Đồng thời, tần số alen tối thiểu (MAF) cũng được tính để loại bỏ các biến thể hiếm.



Hình 1. Sơ đồ phương pháp xác định và sàng lọc SNP không đồng nghĩa tiềm năng

Tất cả phân tích thống kê được thực hiện bằng phần mềm SNP tiếp tục được phân loại dựa trên tác động lên chức năng protein, các SNP có tác động MODERATE được giữ lại, đây là chính là các SNP không đồng nghĩa. Danh sách các SNP tiềm năng sẽ được đưa vào nghiên cứu chú giải chức năng bằng công cụ UniProt Retrieve/ID mapping và nền tảng DAVID (Database for Annotation, Visualization, and Integrated Discovery) [15]. Các thuật ngữ Gene Ontology (GO) liên quan đến quá trình sinh học, chức năng phân tử và thành phần tế bào được xác định cho các gen chứa nsSNP tiềm năng.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Kết quả giải trình tự hệ gen RNA và đóng hàng hệ gen tham chiếu

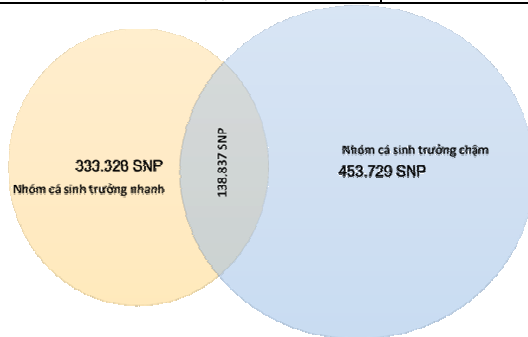
Kết quả giải trình tự và đóng hàng với hệ gen tham chiếu của cá chép được trình bày trong bảng 1. Hệ gen RNA của hai nhóm cá chép lai đã được giải trình tự thành công từ mẫu mô cơ thịt. FG có trung bình 52,76 triệu reads và 7,9 Gb dữ

liệu, cao hơn so với SG (48,15 triệu reads; 7,23 Gb). Hiệu quả giải trình tự cao ở cả hai nhóm (98,5%), chỉ số Q30 > 93%, tỷ lệ GC tương đương (~48,8%). Số reads được đóng hàng đạt khoảng 24,6 triệu reads mỗi nhóm, với tỷ lệ mapping cao hơn ở nhóm sinh trưởng nhanh (87,64% so với 84,79%). Kết quả trong nghiên cứu này tương đồng với các nghiên cứu gần đây trên cá chép. Nghiên cứu trên ba dòng cá chép, giải trình tự từ mô cơ và gan nhằm xác định các gen và con đường sinh hóa liên quan đến dinh dưỡng đã thu

được từ 38 - 51 triệu reads với điểm chất lượng Q30 đạt trên 95% [16]. Tương tự, nghiên cứu trên cá chép (*Cyprinus carpio rubrofusculus*) ở hai nhóm cá lớn nhanh và lớn chậm cũng thu được 20 - 23 triệu reads, với tỷ lệ GC dao động từ 46,83 - 50,07% và Q30 trên 95,2% [17]. Nhìn chung, dữ liệu trình tự RNA-seq có chất lượng cao, tỷ lệ mapping hiệu quả, đạt yêu cầu cho các phân tích tiếp theo để xác định các biến thể ở hai nhóm sinh trưởng trưởng của cá chép lai.

Bảng 1. Thống kê kết quả giải trình tự và đóng hàng với hệ gen tham chiếu của cá chép

Nhóm cá chép	Nhóm sinh trưởng nhanh	Nhóm sinh trưởng chậm
Số lượng Reads*	52.761.616 ± 3.864.102	48.152.444 ± 4.633.784
Dung lượng trình tự thô (Gb)	7,9 ± 0,6	7,233 ± 0,74
Hiệu quả (%)	98,54 ± 0,68	98,48 ± 0,23
Độ tin cậy Q30(%)	93,6 ± 0,69	93,73 ± 0,95
Tỷ lệ GC (%)	48,91 ± 0,85	48,78 ± 0,39
Tổng số reads đưa vào đóng hàng (reads)	24.598.256 ± 849.080	24.861.129 ± 1.082.078
Tỷ lệ mapping với hệ gen tham chiếu (%)	87,64 ± 4,7	84,79 ± 4,42



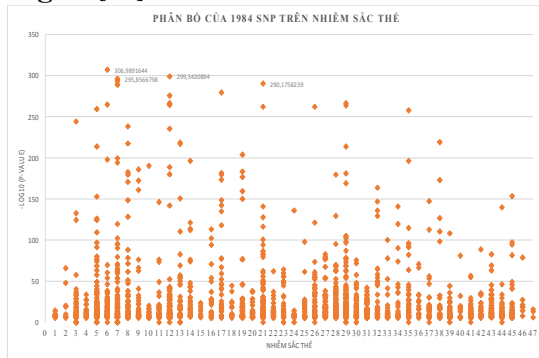
Hình 2. Số lượng SNP phát hiện trên hai nhóm cá chép

* Ghi chú: Số lượng reads = tổng số trình tự của 2 lần đọc; dung lượng trình tự thô (Gigabite -Gb) = (Số lượng đoạn trình tự) x (độ dài chuỗi), tính bằng Gigabite. Đối với giải trình tự 2 chiều, độ dài chuỗi bằng 150. Hiệu quả = (Số lượng đoạn trình tự sạch/Số lượng đoạn trình tự) * 100%; Q30: (Số lượng base của giá trị Phred > 30)/(Tổng số base). Tỷ lệ GC: (Số lượng base G & C)/(Tổng số lượng base).

3.2. Kết quả xác định SNP

Dựa trên hệ gen tham chiếu của cá chép *GCF_018340385.1* với tổng chiều dài 1.680.118.328 bp và 50 nhiễm sắc thể. Kết quả sau khi gọi biến thể đã phát hiện được 333.328 SNP cho nhóm cá sinh trưởng nhanh và 453.729 SNP cho nhóm cá sinh trưởng chậm. Cả hai nhóm chia sẻ 138.837 SNP (Hình 2) cho thấy, sự tương đồng di truyền nhưng cũng phản ánh sự khác biệt về số lượng biến dị riêng biệt của từng nhóm. Kết quả sau khi kiểm định các SNP có liên quan đến sự khác biệt trong sinh trưởng giữa hai nhóm cá chép sinh trưởng nhanh và sinh trưởng chậm sử dụng Fisher's exact test (2-tail test), kết hợp với p-adjust bằng phương pháp Bonferroni, đã xác định được 1984 SNP có ý nghĩa thống kê (p-value < 0,05 và (p_adjust) < 0,00001; OR ≥ 2 hoặc ≤ 0,5), trên 47 nhiễm sắc thể của cá chép. Kết quả xác định biến thể SNP dựa trên RNA-sequencing trong

nguyên cứu này cho tổng số SNP gọi ra thấp hơn so với phương pháp giải trình tự hệ gen ADN trên cá bơn (*Cynoglossus semilaevis*), đã gọi ra 6.545.735 SNP. Tuy nhiên, sau khi kiểm sai khác tần số alen giữa 2 nhóm các bơn lớn nhanh và lớn chậm bằng Fisher exact test lại cho kết quả ít hơn nguyên cứu này với 1.206 SNP có ý nghĩa thống kê [18].



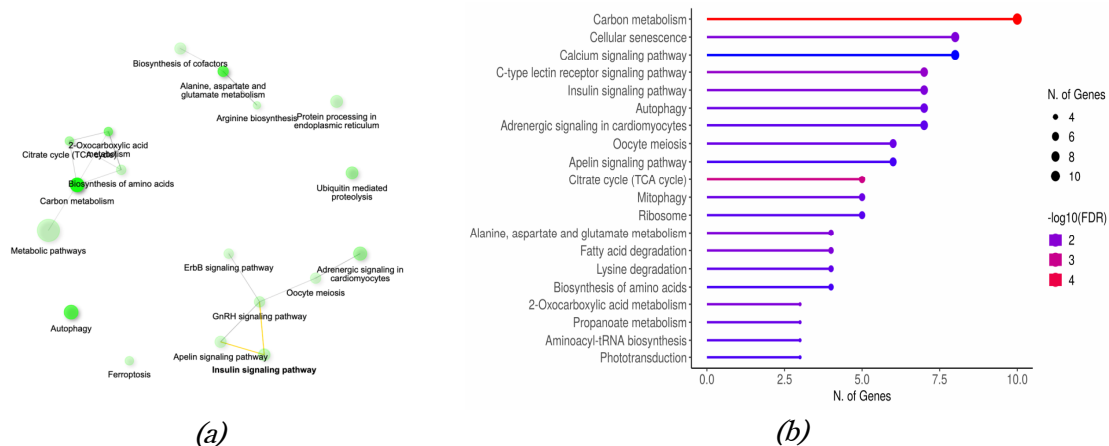
Hình 3. Biểu đồ phân bố của 1.984 SNP trên các nhiễm sắc thể

Kết quả ở hình 3 cho thấy, SNP phân bố trên các nhiễm sắc thể với mật độ và mức độ ý nghĩa thống kê có sự khác biệt rõ rệt. Đặc biệt, các nhiễm sắc thể số 6, 10, 14, 17 và 25 có nhiều SNP với giá trị $-\log_{10}(P\text{-value})$ cao, gợi ý các

vùng này có thể chứa những gen hoặc yếu tố di truyền quan trọng liên quan đến đặc điểm sinh trưởng hoặc thích nghi của cá chép. Ngược lại, một số nhiễm sắc thể khác có các SNP với mức độ ý nghĩa thống kê thấp, cho thấy mối liên hệ yếu hơn hoặc không đáng kể (nhiễm sắc thể số 19, 21 và 42). Các SNP có giá trị P - value cao nên được ưu tiên trong các nghiên cứu tiếp theo để xác định các gen tiềm năng liên quan đến tính trạng mong muốn, đồng thời làm cơ sở cho các ứng dụng trong chọn giống và cải thiện giống cá chép.

3.3. Kết quả thông tin gen và chú thích chức năng các gen có SNP

Tổng số 1984 SNP có ý nghĩa thống kê trong nghiên cứu sai khác giữa nhóm cá sinh trưởng nhanh và nhóm cá sinh trưởng chậm được phát hiện trên 881 gen. Kết quả phân tích chức năng gen sử dụng cơ sở dữ liệu UniProt và DAVID [15] làm rõ hơn vai trò của các gen mang những SNP tiềm năng này, là cơ sở đề xuất những SNP được ưu tiên trong nghiên cứu tính trạng sinh trưởng trên cá chép được thể hiện trong hình 4.



Hình 4. Biểu đồ chú giải chức năng sinh học các gen chứa SNP. (a) Biểu đồ mạng liên kết giữa các gen chứa SNP. Hai con đường được kết nối với nhau nếu chúng chia sẻ 20% số gen trở lên. (b) Biểu đồ số lượng gen tham gia trong mỗi con đường KEGG

Kết quả ở hình 4 cho thấy, các SNP chủ yếu nằm trong nhóm gen liên quan đến con đường chuyển hóa cac - bon (10 gen), có chức năng chuyển hoá, tổng hợp và sản xuất năng lượng,

duy trì cân bằng nội môi [19]. Điều này cho thấy, vai trò trung tâm của quá trình chuyển hóa năng lượng trong điều hòa sinh trưởng. Đáng chú ý, con đường Insulin signaling pathway (7 gen) có

vai trò trung tâm được liên kết với nhiều con đường tín hiệu khác như: Apelin signaling pathway, GnRH signaling pathway và ErbB signaling pathway. Insulin signaling pathway, được xác định là một hệ thống tín hiệu quan trọng, linh hoạt và có nhiều chức năng đa dạng. Ngoài vai trò chính trong duy trì cân bằng đường huyết, insulin còn điều hòa các quá trình quan trọng như: Tăng trưởng, tuổi thọ, trao đổi chất trung gian và sinh sản [20]. Nghiên cứu tác động của insulin đã được thực hiện cho một số loài cá, đã chỉ ra có sự khác biệt tùy theo loài và mô, một số gen liên quan đến con đường này như: MAPK/ERK, Akt/PKB và TOR [21]. Ngoài ra, các con đường liên quan đến chuyển hóa như Citrate cycle (TCA cycle), C-type lectin receptor signaling và 2-Oxocarboxylic axit metabolism cũng thấy xuất hiện từ 6 -7 gen, phản ánh sự điều chỉnh mạnh mẽ của các quá trình chuyển hóa năng lượng. Việc làm giàu các con đường Alanine, aspartate and glutamate metabolism và Arginine biosynthesis cho thấy, sự tham gia tích cực của các axit amin trong cân bằng nội môi và sản xuất năng lượng. Nghiên cứu trên cá chình châu Âu cũng cho thấy, các SNP chủ yếu nằm trên các gen liên quan đến các con đường tín hiệu bao gồm: Các con đường tín hiệu MAPK/ERK, tín hiệu canxi và tín hiệu GnRH, có liên quan đến tăng trưởng và phản ánh sự thích nghi với các điều kiện sống khác nhau của cá chình [22]. Nghiên cứu trên 107 SNP bằng công nghệ GWAS cũng chỉ ra các con đường liên quan đến sinh trưởng trên cá chim vây vàng có liên quan đến tổ chức bộ khung tế bào và quá trình trao đổi chất [23]. Kết quả cho thấy, các gen chứa SNP không chỉ phân bố ngẫu nhiên mà còn tập trung trong một số con đường sinh học cụ thể, đặc biệt là các con đường liên quan đến chuyển hóa và tín hiệu sinh trưởng. Dựa trên tài liệu tham khảo và kết quả phân tích Gene Ontology [24]; các gen liên quan đến sinh trưởng, trao đổi chất và phát triển cơ bắp được ưu tiên cho phân tích SNP tiềm năng. Đặc biệt chú ý đến các gen tham gia vào con đường phát

triển cấu trúc cơ và tổng hợp protein liên quan đến điều hòa trao đổi chất trên cá chép.

Tăng trưởng là một tính trạng đa gen, chịu ảnh hưởng bởi các gen nằm trong nhiều con đường sinh lý khác nhau, chẳng hạn như các con đường điều hòa trao đổi chất, phát triển cơ, hình thành xương và kiểm soát sự thèm ăn, việc chọn giống MAS dựa trên một vài chỉ thị phân tử đơn lẻ sẽ không hiệu quả trong việc cải thiện các đa gen này [25]. Phương pháp chọn giống MAS dựa trên nhiều gen liên quan hoặc chọn giống hệ gen (Genome selection), cho phép đánh giá toàn diện tiềm năng di truyền bằng cách tính đến tác động kết hợp của nhiều locus trên toàn hệ gen, từ đó giúp cải thiện độ chính xác trong dự đoán và nâng cao hiệu quả chọn giống đối với các tính trạng liên quan đến tăng trưởng trong thủy sản [26].

3.4. Các nsSNP tiềm năng liên quan đến tính trạng sinh trưởng

Kết quả phân tích SNP căn cứ vào kết quả Gene Ontology, lựa chọn trên các gen và nhóm gen quan trọng có liên quan đến tính trạng sinh trưởng, nghiên cứu đề xuất tổng số 37 nsSNP tương ứng với 14 gen có tiềm năng cho nghiên cứu di truyền liên quan đến sự khác biệt về tốc độ tăng trưởng giữa hai nhóm cá. Thông tin chi tiết của các nsSNP được mô tả ở bảng 2.

Các nsSNP đáng chú ý trong nghiên cứu này bao gồm biến thể nsSNP- c.465T>G (làm thay đổi axit amin p.His155Gln) trong gen *Myogenic Factor 6* (MYF6), có vai trò quan trọng liên quan đến tính trạng tăng trưởng ở nhiều loài động vật, như nghiên cứu trên cá hilsa (*Tenualosa ilisha*) bằng phương pháp giải mã transcriptome mô cơ đã chỉ ra Myf6 có liên kết chức năng với các gen quan trọng khác trong con đường phát triển cơ, qua đó xác định được 3 SNP làm thay đổi axit amin (p.Glu32Gly; p.Glu48Gly và p.Val67Ala) giữa các nhóm cá tăng trưởng chậm và tăng trưởng nhanh có tiềm năng sử dụng làm chỉ thị phân tử để phân biệt giữa các loài cá tăng trưởng chậm và nhanh [27]. Các SNP nằm trong vùng mã hóa của gen *Myf6*

của cá rô phi vẫn được báo cáo có liên quan đến các đặc điểm về chiều dài và cân nặng cá, là những chỉ thị quan trọng trong chọn giống cá rô phi sinh trưởng nhanh [28]. Myf6: nsSNP-c.465T>G có tần số alen thiểu số khác biệt giữa hai nhóm cá (MAF = 0,45 trên nhóm cá sinh trưởng chậm và MAF = 0,19 trên nhóm cá sinh trưởng nhanh); cho thấy đột biến này mặc dù được quan sát thấy trên cả hai nhóm cá, tuy nhiên xuất hiện phổ biến hơn ở nhóm cá sinh trưởng chậm.

Các SNP cũng được báo cáo cho *MEF* trong nghiên cứu chọn lọc nhân tạo cá ngựa vằn dựa trên kích thước cá thể [29]. Biến thể này lại được quan sát xuất hiện phổ biến hơn ở nhóm cá sinh trưởng nhanh với MAF = 0,39 so với tần số alen tối thiểu rất thấp MAF = 0,02 ở nhóm cá sinh trưởng chậm. Hai gen *MYF6* và *MEF2* đều có vai trò quan trọng trong quá trình sinh học phát triển và biệt hóa cơ xương (Myogenesis), ảnh hưởng trực tiếp lên sinh trưởng và phát triển của cá.

Bảng 2. Thông tin các nsSNP tiềm năng cho nghiên cứu sinh trưởng trên cá chép lai

STT	Vị trí SNP trên NST	Tên gen	Kiểu thay thế nucleotide	Kiểu thay thế Axit amin	Nhóm cá sinh trưởng nhanh		Nhóm cá sinh trưởng chậm	
					Kiểu gen	MAF	Kiểu gen	MAF
1	NC_056579.1-10532428	<i>GLS2A</i>	c.76A>G	p.Ile26Val	T/C	0,39	C/C	0,11
2	NC_056581.1-16456014	<i>MEF2CA</i>	c.1340C>A	p.Thr447Asn	G/T	0,39	T/T	0,02
3	NC_056575.1-14637504	<i>MYF6</i>	c.465T>G	p.His155Gln	G/G	0,19	T/G	0,45
4	NC_056608.1-14758333	<i>MYOZ1B</i>	c.433A>G	p.Asn145Asp	T/C	0,43	C/C	0,11
5	NC_056619.1-7112923	<i>MYBPHA</i>	c.95T>C	p.Val32Ala	C/C	0	T/C	0,37
6	NC_056602.1-21802374	<i>MYH</i>	c.3457T>G	p.Ser1153Ala	A/C	0,49	C/C	0
7	NC_056581.1-16741859	<i>MYO18B</i>	c.5386A>G	p.Arg1796Gly	G/G	0	A/G	0,46
8	NC_056581.1-16710640		c.454T>A	p.Ser152Thr	A/A	0	T/A	0,42
9	NC_056602.1-21738773	<i>PERM1</i>	c.2274G>C	p.Leu758Phe	C/G	0,31	G/G	0
10	NC_056602.1-21738904		c.2143A>G	p.Thr715Ala	T/C	0,5	C/C	0
11	NC_056602.1-21740534		c.513G>T	p.Trp171Cys	C/A	0,32	A/A	0
12	NC_056602.1-21740292		c.755C>A	p.Ser252Tyr	G/T	0,36	T/T	0,02
13	NC_056602.1-21740713		c.334G>T	p.Gly112Trp	C/A	0,32	A/A	0
14	NC_056602.1-21740322		c.725C>T	p.Thr242Met	G/A	0,38	A/A	0,02
15	NC_056602.1-21739572		c.1475A>T	p.Glu492Val	T/A	0,45	A/A	0
16	NC_056602.1-21739146		c.1901C>A	p.Thr634Asn	G/T	0,45	T/T	0
17	NC_056602.1-21739263		c.1784A>T	p.Lys595Ile	T/A	0,49	A/A	0
18	NC_056588.1-933928	<i>ADSS1</i>	c.1110A>T	p.Glu370Asp	A/T	0,31	A/T	0,43
19	NC_056588.1-933971		c.1153A>G	p.Ile385Val	A/G	0,31	A/G	0,43
20	NC_056578.1-39568680	<i>SYNPO2A</i>	c.3437G>C	p.Gly1146Ala	C/G	0,39	G/G	0
21	NC_056578.1-39570023		c.2094C>G	p.Asp698Glu	G/C	0,5	C/C	0
22	NC_056578.1-39570446		c.1671T>A	p.His557Gln	A/T	0,29	A/T	0,3
23	NC_056578.1-39569215		c.2902G>C	p.Glu968Gln	C/G	0,34	C/G	0,3
24	NC_056578.1-39569161		c.2956G>A	p.Glu986Lys	C/T	0,38	C/T	0,3
25	NC_056578.1-39569152		c.2965T>G	p.Ser989Ala	A/C	0,38	A/C	0,32
26	NC_056603.1-41164083	<i>SYNPO2L</i>	c.2422G>A	p.Ala808Thr	C/T	0,46	C/T	0,17
27	NC_056603.1-41162964		c.3541G>A	p.Val1181Ile	C/T	0,42	C/T	0,22
28	NC_056603.1-41161880		c.4625C>G	p.Ala1542Gly	G/C	0,45	G/C	0,25
29	NC_056603.1-41163621		c.2884C>A	p.His962Asn	G/T	0,45	G/T	0,25

30	NC_056620.1-20326784	<i>OBSL1</i>	c.20111A>G	p.His6704Arg	A/G	0,48	G/G	0
31	NC_056620.1-20313832		c.15564A>T	p.Glu5188Asp	A/T	0,46	T/T	0
32	NC_056620.1-20313909		c.15641T>A	p.Phe5214Tyr	T/A	0,38	A/A	0
33	NC_056620.1-20312997		c.14729A>T	p.Glu4910Val	A/T	0,43	T/T	0
34	NC_056598.1/2173700	<i>MYLK3L</i>	c.861C>G	p.Asp287Glu	C/C	0,07	G/C	0,35
35	NC_056598.1/2173536		c.1025C>A	p.Thr342Lys	T/T	0,09	G/T	0,39
36	NC_056598.1/2173516		c.1045C>A	p.His349Asn	T/T	0,09	G/T	0,37
37	NC_056578.1-6639735	<i>ACTN3</i>	c.285A>G	p.Phe876Leu	A/G	0,26	A/G	0,3

Ghi chú: GLS2A (Glutaminase 2, mitochondrial-like A), MEF2CA (Myocyte Enhancer Factor 2C A-like), MYF6 (Myogenic Factor 6), MYOZ1B (Myozenin 1b), MYBPHA (Myosin Binding Protein H A-like), MYH (Myosin Heavy Chain), MYO18B (Myosin XVIII B), PERM1 (PPARGC1 và ESRR-induced Regulator, Muscle 1), ADSS1 (Adenylosuccinate Synthase 1), SYNPO2A (Synaptopodin 2A), SYNPO2L (Synaptopodin 2-like), OBSL1 (Obscurin-like 1), MYLK3L (Myosin Light Chain Kinase 3-like), ACTN3 (Alpha-actinin-3).

Số lượng nhiều các nsSNP có ý nghĩa thống kê phát hiện được trên gen *PERM1* (9 nsSNP) và *SYNPO2* (10 nsSNP) với tần số alen thiếu số lớn hơn ở nhóm cá sinh trưởng trưởng nhanh (MAF của cá sinh trưởng nhanh = 0,29 - 0,5 và MAF của cá sinh trưởng chậm = 0 - 0,32), cho thấy các chỉ thị SNP này có thể được sử dụng để đại diện cho nhóm cá chép sinh trưởng nhanh. Về chức năng sinh học, hai gen *PERM1* và *SYNPO2* đều tham gia vào quá trình myogenesis; *PERM1* là một protein đặc hiệu của cơ, được kích hoạt bởi PGC-1 α và ERRs, có vai trò quan trọng trong điều hòa gen liên quan đến chuyển hóa năng lượng, sinh tổng hợp ty thể và chức năng cơ cơ. Sự suy giảm *PERM1* làm giảm khả năng hô hấp của tế bào cơ và ảnh hưởng đến quá trình sinh tổng hợp ty thể [30]. *SYNPO2* là một protein tham gia vào việc duy trì cấu trúc tế bào và hỗ trợ phát triển cơ bắp của Myogenesis, đặc biệt trong các tế bào cơ. Nó điều chỉnh sự phân chia tế bào và phát triển mô, ảnh hưởng đến tốc độ tăng trưởng của cá. *SYNPO2* cũng có vai trò trong việc ứng phó với stress sinh lý, duy trì sự ổn định của tế bào trong môi trường khắc nghiệt [31, 2].

Tổng số 08 nsSNP cũng được đề xuất cho nhóm gen (*MYOZ1B*: Myozenin 1B; *MYBPH*: Myosin Binding Protein H; *MYH*: Myosin Heavy Chain và *MYO18B*: Myosin XVIII B) liên quan đến chức năng phát triển mô cơ, nằm trong con

đường sinh học tế bào muscle cytoskeleton, bao gồm các biến thể: *MYOZ1B*: c.433A>G (p.Asn145Asp); *MYBPH*: c.95T>C (p.Val32Ala); *MYH*: c.3457T>G (p.Ser1153Ala); hai SNP trên *MYO18B*: c.5386A>G (p.Arg1796Gly) và c.454T>A (p.Ser152Thr); ba SNP trên gen *MYLK3L*: c.861C>G (p.Asp287Glu), c.1025C>A (p.Thr342Lys), c.1045C>A (p.His349Asn). Tần số alen thiếu số cũng cho thấy có sự khác biệt đáng kể giữa nhóm cá chép sinh trưởng nhanh và nhóm cá chép sinh trưởng chậm cho các nsSNP này, các biến thể trên *MYBPH*, *MYO18B*, *MYLK3L* (MAF = 0,35 - 0,46) có phổ biến hơn cho nhóm cá sinh trưởng chậm, trong khi các biến thể trên gen *MYOZ1B* (MAF = 0,43) và *MYH* (MAF = 0,49) là phổ biến hơn trên nhóm cá sinh trưởng nhanh. Họ gen Myosin đã được nghiên cứu chức năng trong quá trình phát triển và mối liên kết với các tính trạng như: Sinh trưởng nhanh, chất lượng mô cơ, thích nghi với nhiệt độ trên nhiều loài cá bao gồm cá chép [32]; cá hồi [33]; cá quế (*Siniperca chuatsi*) [34]... Các nghiên cứu về SNP trên các gen họ Myosin cũng được báo cáo liên quan đến tính trạng sinh trưởng trên các loài cá, như chỉ thị SNP gây thay đổi cấu trúc protein (Tr40Ser) trên gen myosin heavy fast skeletal nằm trên LG6 của các bon (*Scophthalmus maximus*) được báo cáo có sai khác tần số alen giữa nhóm cá sinh trưởng nhanh và chậm [35].

Một số gen có xác định được các nsSNP có ý nghĩa trong nghiên cứu này là *ADSS* (*Adenylosuccinate Synthase*), *GLS2A* (*Glutaminase 2A*) và *OBSL* (*Obscurin-like*) cũng có vai trò quan trọng liên quan đến sinh trưởng của cá; tuy nhiên chưa có nhiều nghiên cứu báo cáo cho các chỉ thị phân tử cho các gen này. Gen *Obscurin-like* đặc trưng ở nhóm cá sinh trưởng nhanh báo cáo trên cá ngừ vằn cũng mô tả đây là các gen có kích thước lớn, có liên quan đến cấu trúc và các yếu tố phát triển mô cơ. Gen *ADSS1* có chức năng liên quan đến điều hòa phát triển cơ đã được báo cáo ở động vật có vú và cá ngừ vằn (Park và cs (2016); Li và cs (2007); trên cá chép, gen *ADSS1* nằm trên QTL cho khối lượng cơ thể tại linked- group 41 đã phát hiện được một SNP tiềm năng, tuy nhiên vẫn cần xác nhận thêm về quan hệ của gen này với các tính trạng tăng trưởng [36]. *GLS2a* là một gen mã hóa enzyme glutaminase 2 trong cá và nhiều loài động vật khác, đóng vai trò quan trọng trong việc vận chuyển nitơ, các - bon và chuyển hóa. Chức năng và biểu hiện của gen này đã được báo cáo trên cá chép [37]; nghiên cứu trên cá diếc tam bội (*Carassius auratus*) cho thấy dinh dưỡng có ảnh hưởng lên hoạt động của gen *GLS* [38]; trong khi nghiên cứu trên cá quế (*Siniperca chuatsi*) chỉ ra chức năng phân tử của *GLS* trong miễn dịch [39]; cho thấy vai trò của *GLS* trong các con đường chuyển hóa năng lượng từ đó liên quan đến chức năng phân tử của các con đường sinh học khác, các SNP trong gen này *GLS* chưa được báo cáo cho phân tích tính trạng sinh trưởng cho các loài cá.

Phân tích chức năng gen cho thấy, tiềm năng của các nsSNP này trong việc nghiên cứu cơ chế di truyền của tính trạng sinh trưởng trên cá chép lai ở Việt Nam, qua đó có thể xây dựng được các chỉ thị hoặc bộ chỉ thị cho tính trạng sinh trưởng, ứng dụng quan trọng trong chọn giống, cho phép chọn lọc các cá thể có tốc độ tăng trưởng cao hơn và cải thiện năng suất nuôi trồng thủy sản.

4. KẾT LUẬN VÀ ĐỀ NGHỊ

Dữ liệu toàn bộ hệ gen RNA từ hai nhóm cá chép lai sinh trưởng nhanh và chậm đã được giải trình tự và đóng hàng thành công với hệ gen tham chiếu, tỷ lệ mapping đạt mức cao, đảm bảo chất lượng cơ sở dữ liệu phục vụ cho các phân tích các SNP. Kết quả đã xác định được tổng số 333.328 SNP cho FG và 453.729 SNP cho SG và chia sẻ 138.837 SNP chung cho cả hai nhóm mẫu. Quá trình sàng lọc và kiểm định SNP đã xác định được tổng số 1.984 SNP có ý nghĩa thống kê và liên quan đến sự khác biệt trong sinh trưởng giữa hai nhóm cá chép sinh trưởng nhanh và chậm, phân bố trên 47 nhiễm sắc thể. Phân tích chức năng gen ontology cho thấy, nsSNP được phát hiện trên nhóm các gen liên quan đến con đường chuyển hóa các - bon, có chức năng chuyển hoá, tổng hợp và sản xuất năng lượng, duy trì cân bằng nội môi. Tổng số 37 nsSNP tiềm năng quan trọng được xác định trên 14 gen có vai trò trong quá trình phát triển cơ bắp và chức năng chuyển hóa năng lượng, tác động lên quá trình sinh trưởng, phát triển của cá chép. Đây là những biến dị di truyền quan trọng có tiềm năng để nghiên cứu ứng dụng trong chọn giống.

Nghiên cứu tương lai cần tiếp tục kiểm định các nsSNP trên các nhóm mẫu lớn hơn, nghiên cứu ảnh hưởng đến biểu hiện kiểu hình sinh trưởng, qua đó xây dựng chỉ thị/bộ chỉ thị phân tử hữu ích cho chương trình chọn giống cá chép lai sinh trưởng nhanh trong tương lai.

TÀI LIỆU THAM KHẢO

1. FAO (2022). *The State of World Fisheries and Aquaculture 2022: Towards Blue Transformation*, FAO, Rome, Italy.
2. Koopae H. K. and Koshkoiyeh A. E. (2014). SNPs genotyping technologies and their applications in farm animals breeding programs: review. *Braz arch biol technol*, 57, 87 - 95.
3. Lee M. A., Keane O. M., Glass B. C. et al. (2006). Establishment of a pipeline to analyse non-synonymous SNPs in *Bos taurus*. *BMC Genomics*, 7(1), 298.

4. Escalante-Rojas M., Peña E., Hernández C. *et al.* (2018). De novo transcriptome assembly for the rose spotted snapper *Lutjanus guttatus* and expression analysis of growth/atrophy - related genes. *Aquaculture Research*, 49(4), 1709 - 1722.
5. Leduc A., Zatylny-Gaudin C., Robert M. *et al.* (2018). Dietary aquaculture by-product hydrolysates: Impact on the transcriptomic response of the intestinal mucosa of European seabass (*Dicentrarchus labrax*) fed low fish meal diets. *BMC Genomics*, 19(1), 396.
6. Yang Y., Han T., Xiao J. *et al.* (2018). Transcriptome analysis reveals carbohydrate-mediated liver immune responses in *Epinephelus akaara*. *Sci Rep*, 8(1), 639.
7. Cai S., Li Y., Xu R. *et al.* (2023). Transcriptome analysis provides new insights into ovarian follicle growth in common carp (*Cyprinus carpio*). *Aquaculture Reports*, 33, 101864.
8. Xu P., Zhang X., Wang X. *et al.* (2014). Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat Genet*, 46(11), 1212 - 1219.
9. Andrews, Simon (2017). Babraham Bioinformatics - FastQC A quality control tool for high throughput Sequence data. <<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>>, accessed: 04/20/2024.
10. Bolger Anthony M., Marc Lohse and Bjoern Usadel (2014). Trimmomatic: a flexible trimmer for Illumina sequence data | Bioinformatics | Oxford Academic. *Bioinformatics*, 30(15), 2114 - 2120.
11. Kim D., Langmead B. and Salzberg S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, 12(4), 357 - 360.
12. Garrison E. and Marth G. (2012). Haplotype-based variant detection from short-read sequencing. <<http://arxiv.org/abs/1207.3907>>, accessed: 02/07/2024.
13. Cingolani P. (2022). Variant annotation and functional prediction: SnpEff. *Variant Calling: Methods and Protocols*. Springer US, New York, NY, 289 - 314.
14. Lin Y. -L., Chang P. - C., Hsu C. *et al.* (2022). Comparison of GATK and deep Variant by trio sequencing. *Sci Rep*, 12(1), 1809.
15. Dennis G., Sherman B. T., Hosack D. A. *et al.* (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*, 4(9), R60.
16. Wu Y., Xiao P., Sha H. *et al.* (2024). Transcriptome analysis reveals the potential key genes in nutritional deposition in the common carp (*Cyprinus carpio*). *Animals*, 14(13), 1939.
17. Zhong Z., Fan J., Tian Y. *et al.* (2024). Comparative transcriptome, ultrastructure and histology analyses provide insights into the potential mechanism of growth arrest in south China carp (*Cyprinus carpio rubrofasciatus*). *BMC Genomics*, 25, 1164.
18. Song W., Zhu H., Wang Y. *et al.* (2021). Large-scale screening of growth-related variants in Chinese tongue sole (*Cynoglossus semilaevis*). *J Ocean Univ China*, 20(3), 669 - 680.
19. Weber J.-M. and Haman F. (1996). Pathways for metabolic fuels and oxygen in high performance fish. *Comparative Biochemistry and Physiology Part A: Physiology*, 113(1), 33 - 38.
20. Paloma Álvarez-Rendón J., Manuel Murillo-Maldonado J. and Rafael Riesgo - Escovar J. (2023). The insulin signaling pathway a century after its discovery: Sexual dimorphism in insulin signaling. *General and Comparative Endocrinology*, 330, 114146.
21. Caruso M. A. and Sheridan M. A. (2011). New insights into the signaling system and function of insulin in fish. *General and Comparative Endocrinology*, 173(2), 227 - 247.
22. Pujolar J. M., Jacobsen M. W., Bekkevold D. *et al.* (2015). Signatures of natural selection between life cycle stages separated by

- metamorphosis in European eel. *BMC Genomics*, 16(1), 600.
23. Zhu F., Sun H., Jiang L. *et al.* (2023). Genome - wide association study for growth - related traits in golden pompano (*Trachinotus ovatus*). *Aquaculture*, 572, 739549.
24. Ashburner M., Ball C. A., Blake J. A. *et al.* (2000). Gene Ontology: Tool for the unification of biology. *Nat Genet*, 25(1), 25 - 29.
25. Garcia A., Tsuruta S., Gao G. *et al.* (2023). Genomic selection models substantially improve the accuracy of genetic merit predictions for fillet yield and body weight in rainbow trout using a multi-trait model and multi-generation progeny testing. *Genet Sel Evol*, 55(1), 11.
26. Robledo D., Matika O., Hamilton A. *et al.* (2018). Genome - wide association and genomic selection for resistance to amoebic gill disease in atlantic salmon. *G3 (Bethesda)*, 8(4), 1195 - 1203.
27. Divya B. K., Yadav P., Masih P. *et al.* (2017). *In silico* characterization of myogenic factor 6 transcript of Hilsa, *Tenualosa ilisha* and putative role of its SNPs with differential growth. *Meta Gene*, 13, 140 - 148.
28. Zou G., Zhu Y., Liang H. *et al.* (2015). Association of pituitary adenylate cyclase-activating polypeptide and myogenic factor 6 genes with growth traits in Nile tilapia (*Oreochromis niloticus*). *Aquacult Int*, 23(5), 1217 - 1225.
29. Harish P., Mareco E. A. and Garcia de la serrana D. (2019). A pilot study to elucidate effects of artificial selection by size on the zebrafish (*Danio rerio*) fast skeletal muscle transcriptome. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 233, 65 - 73.
30. Cho Y., Hazen B. C., Russell A. P. *et al.* (2013). Peroxisome proliferator-activated receptor γ Coactivator 1 (PGC-1)- and Estrogen-related receptor (ERR) - induced regulator in muscle 1 (PERM1) is a Tissue-specific regulator of oxidative capacity in skeletal muscle cells*. *Journal of Biological Chemistry*, 288(35), 25207 - 25218.
31. Jin J. J., Lv W., Xia P. *et al.* (2018). Long noncoding RNA SYISL regulates myogenesis by interacting with polycomb repressive complex 2. *Proceedings of the National Academy of Sciences*, 115(42), E9802-E9811.
32. Ikeda D., Nihei Y., Ono Y. *et al.* (2010). Three embryonic myosin heavy chain genes encoding different motor domain structures from common carp show distinct expression patterns in cranial muscles. *Marine Genomics*, 3(1), 1 - 9.
33. Hevrøy E. M., Jordal A.-E. O., Hordvik I. *et al.* (2006). Myosin heavy chain mRNA expression correlates higher with muscle protein accretion than growth in *Atlantic salmon, Salmo salar*. *Aquaculture*, 252(2), 453 - 461.
34. Chen L., Pan Y., Cheng J. *et al.* (2023). Characterization of myosin heavy chain (MYH) genes and their differential expression in white and red muscles of Chinese perch, *Siniperca chuatsi*. *International Journal of Biological Macromolecules*, 250, 125907.
35. Robledo D., Rubiolo J. A., Cabaleiro S. *et al.* (2017). Differential gene expression and SNP association between fast and slow-growing turbot (*Scophthalmus maximus*). *Sci Rep*, 7(1), 12105.
36. Hu X., Zhang X., Ge Y. *et al.* (2023). Mapping of quantitative trait loci for growth traits around the first overwintering period in Songpu mirror carp (*Cyprinus carpio* L.) cultured in Northeast China. *Aquaculture Reports*, 31, 101686.
37. Jiang J., Feng L., Liu Y. *et al.* (2014). Molecular cloning and expression of kidney-type glutaminase from common carp (*Yprinus carpio*) and its up-regulation by glutamine in primary culture enterocyte. *Aquaculture Nutrition*, 20(6), 731 - 740.
38. Xiao Y., Huang R., Cao S. *et al.* (2022). Molecular characterization and dietary

regulation of glutaminase 1 (gls1) in triploid crucian carp (*Carassius auratus*). *Fishes*, 7(6), 377.

39. Liu S., Li N., Lin Q. *et al.* (2020). Glutaminase 1 in mandarin fish *Siniperca*

chuatsi. Molecular characterization, expression pattern and function involving in virus replication. *Aquaculture*, 519, 734924.

IDENTIFICATION OF POTENTIAL NON-SYNONYMOUS SNPS ASSOCIATED WITH GROWTH TRAITS IN HYBRID COMMON CARP USING RNA-SEQ

Luu Thi Ha Giang¹, Vu Van In², Kim Thi Phuong Oanh³

¹ *Research Institute for Aquaculture No. 1*

² *Vietnam Japan University, Vietnam National University, Hanoi*

³ *Vietnam Academy of Science and Technology*

Abstract

Common carp (*Cyprinus carpio*) is a traditionally important freshwater aquaculture species in many countries. Conventional selective breeding, when combined with molecular genetics, plays a crucial role in improving growth rate, productivity, and economic efficiency. This study applied RNA-sequencing technology to identify potential non-synonymous single-nucleotide polymorphisms (nsSNPs) in two hybrid carp groups (fast-growing and slow-growing), derived from crosses between pure Hungarian and Vietnamese carp lines. After sequencing, the total number of filtered reads was approximately 24.6 million in the fast-growing group (FG) and 24.8 million in the slow-growing group (SG), with mapping rates of 87.64% and 84.79%, respectively. A total of 333,328 SNPs were identified in FG and 453,729 in SG, with 138,837 SNPs shared between the two groups. Statistical analysis revealed 1,984 SNPs significantly associated with growth differences, distributed across 47 chromosomes. Functional annotation indicated that most nsSNPs were located in genes related to carbon and energy metabolism. Notably, 37 potential nsSNPs were detected in genes such as Myogenic Factor 6, Myozenin, and Myosin, which are known to be involved in muscle development and growth regulation. These findings provide valuable insights and a scientific foundation for improving the efficiency of fast-growing common carp breeding programs in Vietnam.

Keywords: *Hybrid common carp, growth traits, nonsynonymous SNP (nsSNP), RNA-Seq.*

Ngày nhận bài: 12/3/2025

Ngày chuyển phản biện: 13/3/2025

Ngày thông qua phản biện: 13/4/2025

Ngày duyệt đăng: 28/4/2025